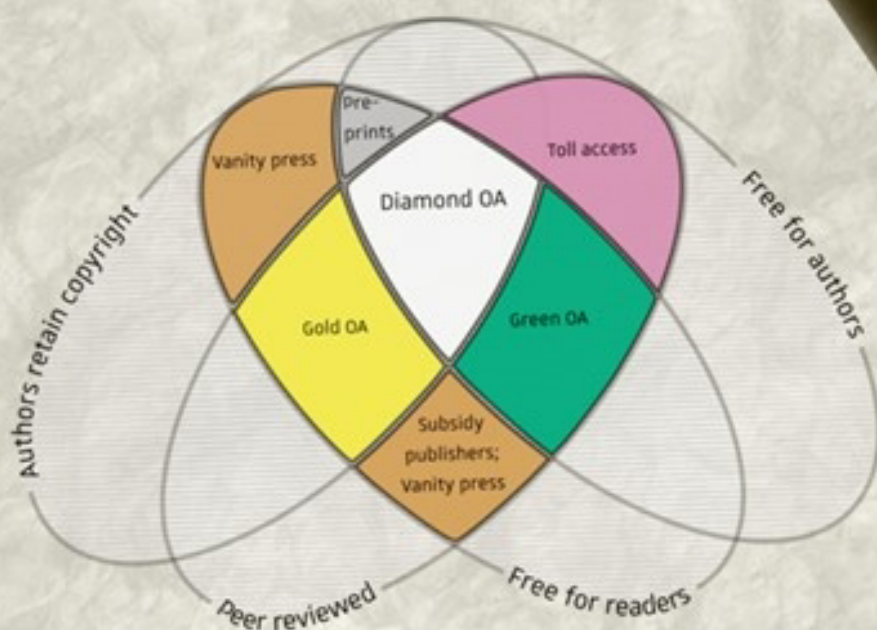


ЦИФРОВАЯ ЭКОНОМИКА



Редакционный совет электронного журнала «Цифровая экономика»

- Агеев Александр Иванович – д.э.н., генеральный директор Института экономических стратегий, заведующий кафедрой НИЯУ «МИФИ», профессор, академик РАН.
- Афанасьев Михаил Юрьевич – д.э.н. Заведующий лабораторией прикладной эконометрики ЦЭМИ РАН
- Бабаян Евгений Борисович – Генеральный директор НП «Агентство научных и деловых коммуникаций»
- Бахтизин Альберт Рауфович – член-корреспондент РАН, д.э.н., профессор РАН, директор ЦЭМИ РАН
- Войниканис Елена Анатольевна – д.ю.н. Ведущий научный сотрудник Института права и развития ВШЭ — Сколково.
- Волынкина Марина Владимировна – д.ю.н. Ректор НОЧУ ВПО «Институт гуманитарного образования и информационных технологий».
- Гурдус Александр Оскарович – д.э.н., к.т.н., президент группы компаний «21Company».
- Димитров Илия Димитрович – исполнительный директор НКО «Ассоциации Электронных Торговых Площадок».
- Ерешко Феликс Иванович – д.т.н. профессор, заведующий отделом информационно-вычислительных систем (ИВС) ВЦ РАН.
- Засурский Иван Иванович – к.ф.н. президент Ассоциации интернет-издателей, заведующий кафедрой новых медиа и теории коммуникации факультета журналистики МГУ имени М.В. Ломоносова
- Калятин Виталий Олегович – к.ю.н., главный юрист по интеллектуальной собственности ООО «Управляющая компания «РОСНАНО»
- Китова О.В. – д.э.н., к.ф.-м.н. зав. кафедрой Информатики РЭУ им. Г.В. Плеханова.
- Козырь Юрий Васильевич – д.э.н., ведущий научный сотрудник ЦЭМИ РАН
- Ливадный Евгений Александрович – к.т.н., к.ю.н., Руководитель проектов по интеллектуальной собственности Государственной корпорации «Ростех».
- Макаров Валерий Леонидович – академик РАН, научный руководитель ЦЭМИ РАН
- Паринов Сергей Иванович – д.т.н., главный научный сотрудник ЦЭМИ РАН.
- Райков Александр Николаевич – д.т.н., профессор, ведущий научный сотрудник Института проблем управления РАН, Генеральный директор ООО «Агентство новых стратегий»
- Семячкин Дмитрий Александрович – к.ф.-м.н., директор Ассоциации «Открытая наука»
- Серго Антон Геннадьевич – д.ю.н., Профессор кафедры авторского права, смежных прав и частоправовых дисциплин Российской государственной академии интеллектуальной собственности (РГАИС)
- Соловьев Владимир Игоревич – д.э.н. руководитель департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве РФ
- Фролов Владимир Николаевич, – д.э.н., профессор, научный руководитель проекта «Copernicus Gold».
- Хохлов Юрий Евгеньевич – к.ф.-м.н., доцент, председатель Совета директоров Института развития информационного общества, академик Российской инженерной академии
- Терелянский Павел Васильевич, – д.э.н., профессор, главный научный сотрудник Научно-исследовательского института "Управления цифровой трансформацией экономики", ФГБОУ ВО "Государственный университет управления".

Миссия журнала

Миссия журнала — поддерживать высокий научный уровень дискуссии о цифровой экономике, методах ее изучения и развития, вовлекая в этот процесс наиболее квалифицированных экспертов — исследователей и практиков; доносить научное знание о самых сложных ее аспектах до тех, кто реально принимает решения, и тех, кто их исполняет. Одновременно журнал направлен на обеспечение возможности для обмена мнениями между профессиональными исследователями.

Название и формат издания

Название «Цифровая экономика» подчеркивает междисциплинарный характер журнала, а также ориентацию на новые методы исследования и новые формы подачи материала, возникшие вместе с цифровой экономикой. В современном ее понимании цифровая экономика — не только новый сектор экономики, но и новые методы сбора информации на основе цифровых технологий, психометрия и компьютерное моделирование, а также иные методы экспериментальной экономики.

Тематика научных и научно-популярных статей

Основную тематику журнала представляют научные и научно-популярные статьи, находящиеся в предметной области цифровой экономики, информационной экономики, экономики знаний. Основное направление журнала — это статьи, освещающие применение подходов и методов естественных наук, математических моделей, теории игр и информационных технологий, а также использующие результаты и методы естественных наук, в том числе, биологии, антропологии, социологии, психологии.

В журнале также публикуются статьи о цифровой экономике и на связанные с ней темы, в том числе, доступные для понимания людей, не изучающих предметную область и применяемые методы исследования на профессиональном уровне. Основная тема — создание и развитие единого экономического пространства России и стран АТР. Сюда можно отнести статьи по обсуждаемым вопросам оптимизации использования ресурсов и государственному регулированию, по стандартам в цифровой экономике. Сегодня или очень скоро это стандарты — умный город, умный дом, умный транспорт, интернет вещей, цифровые платформы, BIM-технологии, умные рынки, умные контракты, краудсорсинг и краудфандинг и многие другие.

Журнал «Цифровая экономика», № 22(1) (2023)

Выпуск № 1 2023 год

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций

Свидетельство о регистрации № ЭЛ № ФС77-70455 от 20 июля 2017 г.

Редакционная коллегия

Козырев А. Н. – главный редактор, д.э.н., к.ф.-м.н., руководитель научного направления – математическое моделирование, г.н.с. ЦЭМИ РАН

Ведута Е. Н. – д.э.н., профессор, зав. кафедрой стратегического планирования и экономической политики факультета государственного управления имени М. В. Ломоносова

Гатауллин Т.М. – д.э.н., к.ф.-м.н., зам. директора Центра цифровой экономики Государственного университета управления

Китов Владимир Анатольевич, к.т.н., зам. Зав. кафедрой Информатики по научной работе РЭУ им. Г.В. Плеханова

Лебедев В. В. – д.э.н., к.ф.-м.н., профессор кафедры высшей математики Государственного университета управления

Лугачев М.И. – д.э.н., заведующий кафедрой Экономической информатики Экономического факультета МГУ имени М.В. Ломоносова.

Макаров С.В. – к.э.н., ведущий научный сотрудник ЦЭМИ РАН.

Неволин И.В. – к.э.н., ведущий научный сотрудник ЦЭМИ РАН

Ноак Н.В. – к.п.н., ведущий научный сотрудник ЦЭМИ РАН

Скрипкин К.Г. – к.э.н., доцент кафедры Экономической информатики Экономического факультета МГУ имени М.В. Ломоносова.

Тевелева О.В. – к.э.н., старший научный сотрудник ЦЭМИ РАН

Писарева О.М. – к.э.н., заведующий кафедрой математических методов в экономике и управлении, Директор Института информационных систем ФГБОУ ВО "Государственный университет управления" (ГУУ)

Чесноков А.Н. – руководитель проекта АН2

Все работы опубликованы в авторской редакции.

Композиция на обложке составлена Елизаветой Вершининой.

Подписано к опубликованию в Интернете 28.03.2023, Авт. печ.л. 9,7

Сайт размещения публикаций: <http://digital-economy.ru/>

Адрес редакции: 117418 Москва, Нахимовский проспект, 47, комн. 520

При использовании материалов ссылка на журнал «Цифровая экономика» и на автора статьи обязательна (на условиях creative commons).

© Журнал «Цифровая экономика», 2023

I S S N 2 6 8 6 - 9 5 6 X



9 772686 956001 >

СОДЕРЖАНИЕ

Слово редактора.....	4
1. Научные статьи.....	5
1.1. Афанасьев М. Ю., Гусев А. А. Об оценке экономической сложности регионов	5
1.2. Малиновский В. К. Модульный анализ сложных систем и сложные составные суммы	16
1.3. Егунов В. А., Филимонов О. Ю. Реализация геометрического преобразования отражения для прототипа высокопроизводительного геометрического ядра.....	20
1.4. Шатравин В. В, Шашев Д. В. Применение вычислительных сред для ускорения рекуррентных нейронных сетей.....	27
1.5. Варшавский Л. Е. Исследование структурных изменений в сегменте высокопроизводительных суперкомпьютеров.....	36
1.6. Гурин А. А., Жуков Т. А. Автоматическое определение сарказма в текстах на русском языке	44
1.7. Козырев А. Н. Оптимальные двухкомпонентные цены в экономиках с возрастающей отдачей.....	54
1.8. Попов Д. В., Ральникова К. В. Архив как фактор цифровой трансформации организации.....	65
2. Мнения.....	72
2.1. Апенько С. Н., Лукаш А. В. Индоктринация экономических нарративов для молодежи как цифровой поколенческой общности посредством ФГОС ВО 3++	72
2.2. Родионов А. А. Блокчейн и большие данные в финансовой сфере	84
2.3. Калинина Е.С, Манохина Т. В. Ступаков С. А. Реляционные базы данных и их особенности.....	89
2.4. Калинина Е.С, Манохина Т. В. Ступаков С. А. Визуализация реляционных баз данных..	93

Слово редактора

Дорогие читатели, перед вами первый в этом году выпуск журнала «Цифровая экономика». Его основная тема – сложность в различных ее проявлениях и технические приемы, позволяющие успешно с ней справиться, решая сложные задачи математики, экономики и техники. Как обычно, выпуск состоит из двух разделов. В первый раздел вошли научные статьи, удовлетворяющие всем требованиям, предъявляемым редколлекцией журнала к научным статьям, включая формальные требования по объему и оформлению материала. Во второй раздел – «Мнения» – вошли материалы, удовлетворяющие этим требованиям в целом и поданные как научные статьи, но с некоторыми нарушениями либо в части оформления ссылок, либо в соотношении объема основного текста и вспомогательной информации. В дальнейшем мы планируем подходить к оформлению материалов более строго, возлагая основную работу по подготовке материалов к печати на самих авторов.

Необходимость таких мер связана с тем, что поток статей, поступающих к нам в журнал, увеличился, а принципы работы журнала остались прежними. В нем, как и раньше, нет штатных сотрудников, вся работа, включая корректуру и верстку, выполняется на общественных началах. При этом мы сохраняем приверженность принципам Diamond Open Access – бесплатная публикация для авторов, бесплатный доступ для читателей и сохранение авторских прав за авторами. Вместе с тем мы не можем допустить снижения качества подготовки и оформления материалов, проверяем их на оригинальность, грамотность и выполнение других стандартных требований, в том числе. А теперь еще предлагаем авторам самим готовить научные статьи в том формате, как они потом будут представлены в журнале. Мы будем приветствовать подачу материалов, прошедших проверку на оригинальность и грамотность. Им будет предоставляться приоритет, хотя основное остается прежним – это научный уровень и актуальность. Статьи должны быть интересны читателю, а не только авторам.

Если говорить о содержании данного выпуска, то можно отметить очевидный технический уклон. Техническим вопросам посвящены две трети материалов, вошедших в этот выпуск, тогда как экономические мотивы можно проследить лишь в четырех из двенадцати статей. О них мы поговорим чуть подробнее, оставляя читателю возможность догадываться об отношении редакции к остальным статьям.

Открывающая выпуск статья д.э.н. М. Ю. Афансьева и А. А. Гусева посвящена главной теме этого выпуска – сложности, точнее, *оценке экономической сложности регионов России по видам экономической деятельности*. Проведён сравнительный анализ оценок экономической сложности по 82 секторам и 24 видам экономической деятельности за 2019 г. Отмечено, что оценка экономической сложности регионов по видам экономической деятельности может быть полезна при решении задач управления, направленных на диверсификацию региональной экономики. Основным инструментом исследования – методы математической статистики. Ту же линию по изучению сложности математическими методами, но уже без привязки к экономике продолжает в своей статье д.ф.-м.н. В. К. Малиновский.

Следующие четыре статьи посвящены техническим вопросам, из них хотелось бы особо отметить статью об автоматическом определении сарказма в текстах на русском языке, причем не потому, что другие три статьи хуже, а в силу необычности ее тематики. Статья подготовлена двумя молодыми авторами А. А. Гуриным и Т.А. Жуковым сразу в том виде, как представлена в журнале.

Далее идет статья, которую можно назвать редакционной. Тематически она очень тесно связана с цифровой экономикой, хотя из названия этого не видно. Речь в ней идет о ценообразовании в условиях возрастающей отдачи при расширении производства, что характерно для всех наукоемких отраслей именно в силу больших затрат на исследования. Но именно в цифровой экономике, или, точнее, в экономике цифровых продуктов и услуг, эта тенденция достигает своего пика. А потому именно здесь принципы оптимальности функционирования бизнеса сталкиваются с антимонопольными ограничениями, тормозящими технический прогресс, построенными на устаревших представлениях и по факту препятствующими эффективности. В первую очередь речь идет о запрете ценовых алгоритмов и ценовой дискриминации (дифференциации цен) в различных ее видах. То и другое принято относить к злоупотреблениям монопольным положением, но без них многие проекты останутся нерентабельными и потому нереализованными, а потребители останутся без продуктов и услуг, которые могли бы появиться, но не появятся благодаря вмешательству регуляторов. Парадокс заключается в том, что само существование антимонопольных служб оправдывается необходимостью обеспечить рыночную конкуренцию, которая, в свою очередь, обеспечит эффективность. Но тут мы имеем случай, когда вмешательство в экономику основано на ложных представлениях и несет много больше вреда, чем пользы.

Экономические вопросы также рассматриваются и в следующих трех статьях. Но рассматриваемые в них вопросы не нуждаются в пояснениях, а связь с цифровизацией очевидна. А завершают выпуск статьи, где снова речь о технических вопросах и методах решения. Такова тенденция и, судя по всему, она надолго.

Всем потенциальным читателям желаю, как всегда, увлекательного и не всегда легкого чтения.

Главный редактор журнала

д.э.н. А. Н. Козырев

1. НАУЧНЫЕ СТАТЬИ

1.1. ОБ ОЦЕНКЕ ЭКОНОМИЧЕСКОЙ СЛОЖНОСТИ РЕГИОНОВ

Афанасьев М. Ю.¹, Гусев А. А.²,
¹²ЦЭМИ РАН, Москва, Россия

Предложен подход к оценке экономической сложности регионов России по видам экономической деятельности. Подход основан на стандартном методе оценки экономической сложности. Рассмотрен вопрос выбора порогового значения показателя выявленных сравнительных преимуществ RCA. Построены 0-1 матрицы для секторов и видов экономической деятельности при разных порогах. Их структуры правильно отражают идею, заложенную в индексе экономической сложности. В результате корреляционного анализа показано, что при пороге 1 индекс экономической сложности по секторам и индекс экономической сложности по видам экономической деятельности обладают большей устойчивостью к изменению порога, чем при других пороговых значениях. Проведён сравнительный анализ оценок экономической сложности по 82 секторам и 24 видам экономической деятельности за 2019 г. Оценка экономической сложности регионов по видам экономической деятельности может быть полезна при решении задач управления, направленных на диверсификацию региональной экономики и повышение ее экономической сложности.

1. Введение

В работе [Афанасьев, Гусев, 2022] был предложен подход к формированию рекомендаций по развитию секторов с целью диверсификации региональной экономики, ориентированный на повышение ее экономической сложности. Оценки экономической сложности регионов получены в этой работе на основе данных о налоговых поступлениях по секторам экономики, что позволяет характеризовать структуры региональных экономик, включающих сектора, ориентированные как на внешний, так и на внутренний рынки. Однако, для оценки экономической сложности регионов могут использоваться альтернативные подходы.

В предложенной далее модификации стандартного подхода для оценки экономической сложности регионов используются показатели промышленного производства: объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по четырём укрупнённым видам экономической деятельности, а также структуры объёмов отгруженной продукции по каждому укрупнённому виду экономической деятельности. В результате формируются оценки объёмов отгруженной продукции по 24 видам экономической деятельности (ВЭД). На этой основе построена матрица выявленных сравнительных преимуществ, описывающая структуру региональных экономик по ВЭД, и получены оценки экономической сложности регионов в соответствии со стандартным подходом.

Рассмотрен один из актуальных вопросов применения индекса экономической сложности, который связан с выбором порогового значения RCA на этапах формирования и сравнительного анализа индексов, построенных на основе различных данных.

Проведено сравнение оценок экономической сложности регионов, полученных на основе данных по 24 ВЭД, с ранее полученными оценками экономической сложности на основе данных об объемах производства по 82 секторам экономики. Высокий уровень корреляции этих оценок указывает на их устойчивость по отношению к используемым данным и уровню детализации описания структуры региональной экономики.

2. Методология

Описание структуры региональной экономики. Для описания структуры региональной экономики использованы данные об объемах отгруженной продукции по ВЭД. Сначала определим показатель RCA_{cp} выявленных сравнительных преимуществ:

$$RCA_{cp} = (y_{cp} / \sum_p y_{cp}) / (\sum_c y_{cp} / \sum_{cp} y_{cp}), \quad (1)$$

где y_{cp} — объем производства по ВЭД p экономики региона c ; RCA_{cp} — отношение доли производства по ВЭД p в общем объеме производства по всем ВЭД экономики региона c к доле производства по ВЭД p всех регионов в объеме производства по всем ВЭД экономик всех регионов. В соответствии с работой [Hausmann, Klinger, 2006], для выявления сравнительных преимуществ в экономиках используется показатель RCA_{cp} для которого проверяется условие типа ограничения снизу. А именно, если значение RCA_{cp} превышает единицу, то считается, что экономика региона c обладает выявленными сравнительными преимуществами в выпуске продукции по ВЭД p ; в противном случае — выявленных сравнительных преимуществ не существует:

$$a_{c,p} = \begin{cases} 1, & \text{если } RCA_{cp} \geq 1; \\ 0, & \text{если } RCA_{cp} < 1. \end{cases}$$

Матрица $A = (a_{c,p})$ содержит данные о ВЭД, которые в разных регионах развиты на уровне выявленных сравнительных преимуществ, определенных при помощи выражения (1). Строки этой матрицы соответствуют регионам, столбцы — ВЭД. Вектор $(a_{c,p_1}, \dots, a_{c,p_m})$ будем называть *структурой сильных ВЭД* экономики региона c .

Экономическая сложность. Понятие «*экономическая сложность региона*» рассматривается как характеристика, отражающая уровень его технологического развития, который определяется сильными ВЭД в структуре его экономики. Аналогично экономическая сложность ВЭД зависит от уровня технологического развития тех регионов, в структуре которых этот ВЭД присутствует в качестве сильного. Экономическая сложность является латентной характеристикой региона ECl_c или ВЭД ECl_p . Оценки экономической сложности обладают следующими свойствами: экономическая сложность региона пропорциональна среднему уровню экономической сложности сильных ВЭД в структуре его экономики:

$$ECl_c = a_1 \sum_p r_{c,p} ECl_p, \quad r_{c,p} = a_{c,p}/k_{c,0}, \quad k_{c,0} = \sum_p a_{c,p}, \quad (2)$$

где a_1 — положительная константа.

Экономическая сложность ВЭД пропорциональна среднему уровню экономической сложности регионов, в структуре экономик которых этот ВЭД является сильным:

$$ECl_p = a_2 \sum_c r_{p,c}^* ECl_c, \quad r_{p,c}^* = a_{c,p}/k_{p,0}, \quad k_{p,0} = \sum_c a_{c,p}, \quad (3)$$

где a_2 — положительная константа. Показатель $k_{c,0}$, равный числу сильных ВЭД в регионе c , будем называть *диверсификацией структуры экономики* региона c по ВЭД. Пусть $\mathbf{c} = (ECl_{c_1}, ECl_{c_2}, \dots)^T$ — вектор-столбец значений экономической сложности для регионов; $\mathbf{p} = (ECl_{p_1}, ECl_{p_2}, \dots)^T$ — вектор-столбец значений экономической сложности для ВЭД; $\mathbf{R}_1 = (r_{c,p})$, $\mathbf{R}_2 = (r_{p,c}^*)$ — матрицы весов. Из соотношений (2) и (3) следует, что $\mathbf{c} = a_1 a_2 \mathbf{R}_1 \mathbf{R}_2 \mathbf{c}$, $\mathbf{p} = a_1 a_2 \mathbf{R}_2 \mathbf{R}_1 \mathbf{p}$. Таким образом, экономическая сложность региона определяется как собственный вектор матрицы $\mathbf{R}_1 \mathbf{R}_2$, а экономическая сложность ВЭД — как собственный вектор матрицы $\mathbf{R}_2 \mathbf{R}_1$. Матрицы $\mathbf{R}_1 \mathbf{R}_2$ и $\mathbf{R}_2 \mathbf{R}_1$ являются стохастическими: их элементы неотрицательны, а их сумма по строкам равна 1. В силу стохастичности матрица $\mathbf{R}_1 \mathbf{R}_2$ имеет собственное значение, равное 1, и отвечающий ему собственный вектор, который состоит из одинаковых координат. В работах [Hausmann, Rodrik, 2003; Sciarra et al., 2020] в качестве значений оценок экономической сложности регионов и ВЭД предлагается использовать значение собственного вектора матриц $\mathbf{R}_1 \mathbf{R}_2$, которые соответствуют второму максимальному собственному значению.

Описание данных. В предложенной далее модификации оценки экономической сложности регионов по ВЭД используются показатели промышленного производства по четырём укрупнённым видам экономической деятельности, а также структуры объёмов отгруженной продукции по 24 ВЭД. В результате по данным Федеральной службы государственной статистики¹ получены объёмы отгруженных товаров, выполненных работ и услуг за 2019 год по 24 ВЭД.

С помощью объёмов отгруженной продукции, выполненных работ и услуг промышленного производства по четырём основным укрупнённым видам экономической деятельности $V_c^{(r)}$ по регионам за 2019 год (раздел 13.01), а также структур распределения этих объёмов в процентах $x_{cp}^{(r)}$ для каждого ВЭД внутри соответствующего основного укрупнённого вида экономической деятельности в промышленности $r = \{1,2,3,4\}$ «Добыча полезных ископаемых», «Обрабатывающие производства», «Обеспечение электрической энергией, газом и паром; кондиционирование воздуха» и «Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений» (соответственно разделы: 13.04, 13.05, 13.06 и 13.07) можно оценить объёмы отгруженной продукции по ВЭД для большинства регионов. Для заполнения пропусков, появившихся вследствие отсутствия нужной информации по отдельным регионам и ВЭД, было установлено соответствие между 24 ВЭД по промышленному производству и 82 секторами по налоговым поступлениям². Пропорции между ними позволили получить оценки объёмов отгруженной продукции по всем ВЭД. Умножение долей ВЭД в объёме соответствующего укрупнённого вида экономической деятельности на общий объём отгруженной продукции по этому виду даёт значения объёмов отгруженной продукции за указанный год для каждого ВЭД по всем регионам:

$$y_{cp} = \frac{x_{cp}^{(r)}}{100} \cdot V_c^{(r)},$$

где c — регион; p — ВЭД. В результате получены оценки объёмов отгруженной продукции по 24 ВЭД для 79 регионов России за 2019 год.

¹ Регионы России. Социально-экономические показатели – 2020 г. URL: https://gks.ru/bgd/regl/b20_14p/Main.htm (дата обращения 15.01.2022)

² Данные о налоговых поступлениях по секторам экономики. URL: https://www.nalog.ru/m77/related_activities/statistics_and_analytics/forms/8826515/ (дата обращения 15.01.2022)

3. Результаты моделирования

Экономическая сложность по ВЭД. Для оценки экономической сложности по ВЭД к данным отгруженной продукции промышленного производства за 2019 г. был применён стандартный подход [Hartmann, 2017; Hausmann, Rodrik 2003; Hausmann, Hwang, Rodrik, 2006; Hidalgo, Hausmann, 2009]. В результате получены оценки экономической сложности регионов по 24 ВЭД (столбец (3) табл. П1 приложения) и оценки экономической сложности ВЭД (столбец (3) табл. П2 приложения).

На рис.1 можно наблюдать нелинейную зависимость экономической сложности регионов от числа сильных ВЭД. У регионов с малым числом сильных ВЭД оценки экономической сложности относительно низкие. С возрастанием числа сильных ВЭД наблюдается тенденция роста экономической сложности регионов.

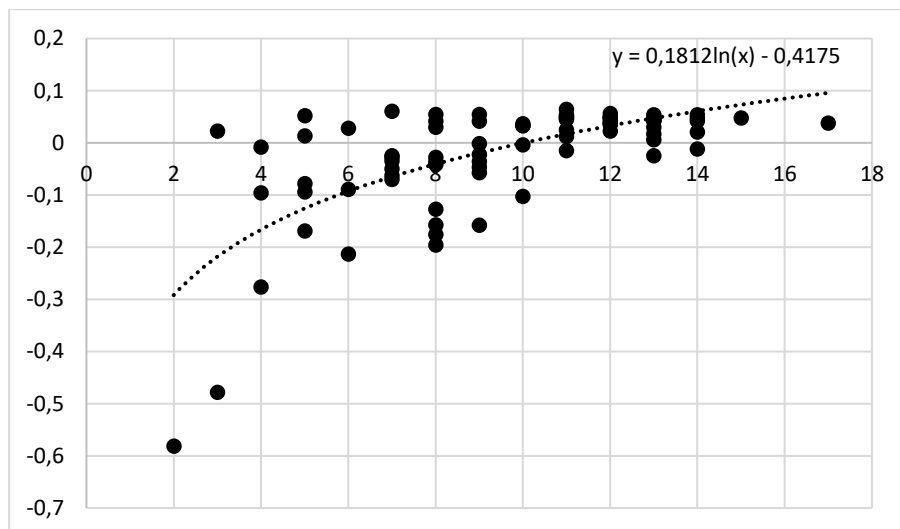


Рис. 1. Зависимость оценок экономической сложности регионов (ось ординат) от числа сильных ВЭД (ось абсцисс).

По результатам расчётов оценок экономической сложности регионов (табл. П1 приложения, столбец (3)) наиболее высокие значения у Тамбовской области (11 сильных ВЭД; экономическая сложность региона 0,06441), Калужской области (7; 0,06103), Ульяновской области (12; 0,05662), Республики Марий Эл (9; 0,05488), Республики Мордовия (8; 0,05437), Пензенской области (13; 0,05418), Чувашской Республики (14; 0,054) и т.д. Эти регионы отличаются высокой диверсификацией экономики по ВЭД. С учетом видов специализации региональных экономик, описанных в работе [Айвазян, Афанасьев, Кудров, 2016], эти регионы можно отнести к смешанному типу. Они специализируются на обрабатывающей промышленности и сельском хозяйстве.

Относительно низкие оценки экономической сложности у Тюменской области (2; -0,58096), Сахалинской области (3; -0,47772), Астраханской области (4; -0,2764); Оренбургской области (6; -0,21304), Республики Саха (Якутия) (8; -0,19579), Республики Коми (8; -0,17566). Это преимущественно регионы со специализацией в добывающей промышленности. Таким образом, относительно высокие оценки экономической сложности у развитых «обрабатывающих» и «сельскохозяйственных регионов», а низкие – у «добывающих» регионов.

Если расположить по возрастанию оценки экономической сложности ВЭД (табл. П2 приложения, столбец (3)), то сначала идут ВЭД из укрупнённого вида экономической деятельности «Добыча полезных ископаемых». В том числе, «предоставление услуг в области добычи полезных ископаемых» Д5, который является сильным ВЭД у 13 регионов и имеет экономическую сложность -0,643959; «добыча нефти и природного газа» Д2 (15; -0,617883); «добыча угля» Д1 (14; -0,294573); «добыча металлических руд» Д3 (22; -0,165385); «добыча прочих полезных ископаемых» Д4 (25; -0,073529). Затем идёт «производство металлургическое; производство готовых металлических изделий, кроме машин и оборудования» ОП8 (21; -0,067646); «ремонт и монтаж машин и оборудования» ОП12 (33; -0,053073); «производство, передача и распределение пара и горячей воды <...>» ОЭ3 (52; -0,019888); «обработка древесины и производство изделий из дерева <...>» ОП3 (31; -0,013296); «предоставление услуг в области ликвидации последствий загрязнений и удалением отходов <...>» В4 (30; -0,001626); «производство, передача и распределение электроэнергии» ОЭ1 (43; 0,01941); «производство бумаги и бумажных изделий; деятельность полиграфическая и копирование носителей информации» ОП4 (27; 0,02714); «производство и распределение газообразного топлива» ОЭ2 (48; 0,02902); «производство кокса и нефтепродуктов <...>» ОП5 (14; 0,02962); «сбор и обработка сточных вод» В2 (43; 0,03215); «забор, очистка и распределение воды» В1 (52; 0,03792); «производство химических веществ и химических продуктов; производство лекарственных средств и материалов, применяемых в медицинских целях» ОП6 (24; 0,06375); «производство машин и оборудования, не включенных в другие группировки»

<...>» ОП10 (30; 0,07254); «сбор, обработка и утилизация отходов; обработка вторичного сырья» ВЗ (37; 0,08353).

В конце с самыми высокими оценками экономической сложности находятся обрабатывающие производства: «производство мебели; производство прочих готовых изделий» ОП11 (31; 0,08514); «производство прочей неметаллической минеральной продукции» ОП7 (38; 0,08907); «производство компьютеров, электронных и оптических изделий; производство электрического оборудования» ОП9 (35; 0,10366); «производство пищевых продуктов; производство напитков; производство табачных изделий» ОП1; (41; 0,10943); «производство текстильных изделий; производство одежды; производство кожи и изделий из кожи» ОП2 (33; 0,12856). Таким образом, добыча полезных ископаемых относится к наименее экономически сложным ВЭД, а обрабатывающие производства – к наиболее экономически сложным.

К вопросу о выборе порога RCA. В работе [Афанасьев, Гусев, 2022] были получены оценки экономической сложности регионов по секторам экономики ECI_c^s . Они представлены в столбце (5) табл. П1 приложения. Индекс экономической сложности (ЕС) измеряет сложность производственной структуры региона путем объединения информации о разнообразии экономики (количество сильных секторов, или ВЭД) и распространенности сильных секторов (количество регионов, в которых сектор или ВЭД является сильным, то есть производит продукцию на уровне выявленных сравнительных преимуществ). Идея, лежащая в основе ЕС, заключается в том, что развитые региональные экономики разнообразны (диверсифицированы) и производят продукцию сильных секторов или ВЭД, которые в среднем имеют низкую распространенность, потому что только несколько региональных экономик развили сектор или ВЭД до уровня сильного. Характеристики сложности производственных структур регионов можно рассматривать как показатель уровня человеческого и социального капитала региональной экономики, поскольку способность региона производить продукцию сильных секторов с высокими оценками сложности зависит от накопленных знаний и способности людей формировать социальные и профессиональные сети для того, чтобы собирать, накапливать и использовать в производстве новые знания [Hidalgo, 2015]. Один из актуальных вопросов практического применения индекса экономической сложности связан с выбором порогового значения RCA на этапах формирования и сравнительного анализа индексов, построенных на основе различных данных.

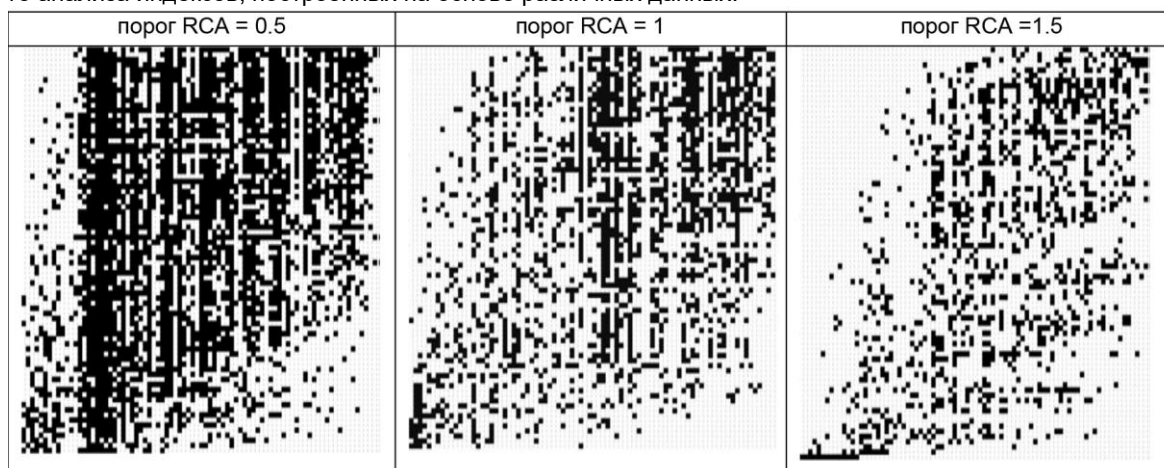


Рис. 2. Матрица 0–1 регион-сектор со строками, упорядоченными по ECI_c^s и столбцами, упорядоченными по ECI_p^s , для различных порогов RCA

На рисунке 2 представлены 0–1 матрицы, описывающие структуру региональных экономик и построенные для различных значений порога RCA. Строки матриц соответствуют регионам, столбцы – секторам экономики. Темная ячейка матрицы означает, что соответствующий элемент матрицы равен 1. То есть сектор является сильным в экономике региона. В противном случае элемент матрицы равен нулю и сектор сильным не является. Строки каждой матрицы упорядочены снизу вверх по возрастанию оценок экономической сложности регионов ECI_c^s . Столбцы упорядочены слева направо по возрастанию оценок экономической сложности секторов ECI_p^s . В прикладных исследованиях по умолчанию используется пороговое значение 1, так как допускает простую интерпретацию. В контексте регионов и секторов при пороге RCA, равном 1, доля производства сильного сектора в экономике региона выше, чем доля этого сектора в национальной экономике.

Рассмотрим структуру матрицы при пороге RCA, равном 1 (на рис. 2 в центре). Верхние строки матрицы содержат существенно больше единиц, чем нижние строки. Соответственно, регионы с более высокими оценками экономической сложности более диверсифицированы, чем регионы с низкими оценками. Сектора с относительно высокими оценками сложности ECI_p^s являются сильными преимущественно в регионах с относительно высокими оценками экономической сложности ECI_c^s . Соответственно, правый нижний угол матрицы слабо заполнен единицами. Сектора с относительно низкими

оценками сложности ECI_p^s являются сильными в регионах с относительно низкими оценками экономической сложности ECI_c^s . Соответственно, левый верхний угол матрицы слабо заполнен единицами.

В матрице, построенной для порога RCA, равного 0.5 (на рис. 2 слева), больше элементов, равных 1. Следует отметить, что при пороге 0.5 некоторые сектора являются сильными практически во всех регионах. Это несколько снижает информативность матрицы. Но эта матрица имеет примерно ту же структуру, что матрица при пороге RCA, равном 1. В матрице, построенной для порога RCA, равного 1.5 (на рис. 2 справа), относительно мало элементов, равных 1. Поэтому она также может быть недостаточно информативна для расчета индекса экономической сложности. Структура каждой из трех матриц, построенных для близких пороговых значений RCA, правильно отражает идею, заложенную в индексе экономической сложности. Поэтому нет серьезных оснований для того, чтобы использовать для расчета RCA порог, отличный от 1. Однако, желательно проверить, что индекс экономической сложности ECI_c^s , рассчитанный при пороге 1, обладает устойчивостью. То есть не меняется сильно при относительно небольшом изменении порога RCA. Для этого целесообразно рассмотреть корреляционную матрицу индексов экономической сложности (табл. 1), построенных для разных пороговых значений.

Таблица 1. Корреляция ECI_c^s по 82 секторам для разных порогов RCA

ECI по секторам. Порог	0.5	0.75	1	1.125	1.25	1.5
0.5	1	0.9136	0.8360	0.8114	0.7299	0.5659
0.75	0.8824	1	0.9361	0.9050	0.8070	0.5755
1	0.7776	0.9565	1	0.9508	0.8479	0.5919
1.125	0.7610	0.9362	0.9699	1	0.9506	0.7198
1.25	0.6854	0.8377	0.8632	0.9434	1	0.8418
1.5	0.1678	0.1919	0.2199	0.5810	0.5810	1

Таблица 1 отражает корреляционную взаимосвязь шести индексов экономической сложности регионов, построенных для пороговых значений в интервале от 0.5 до 1.5. Под главной диагональю таблицы приведены коэффициенты корреляции Пирсона. Над главной диагональю – коэффициенты ранговой корреляции Спирмена. Использование меньших или больших пороговых значений не целесообразно, так как при таких пороговых значениях трудно интерпретировать понятие «сильный сектор» и соответствующий индекс экономической сложности. Кроме того, оценка взаимосвязи такого индекса с индексом, построенным для порога 1, становится незначимой и утрачивается возможность их сопоставления. Наблюдается высокая устойчивость индекса, построенного для порога 1 при изменении порогового значения в диапазоне от 0.75 до 1.125. Устойчивость индекса, построенного при других значениях, ниже. Эти выводы не противоречат рекомендациям использовать в прикладных исследованиях индекс экономической сложности, построенный для порога RCA, равного 1.

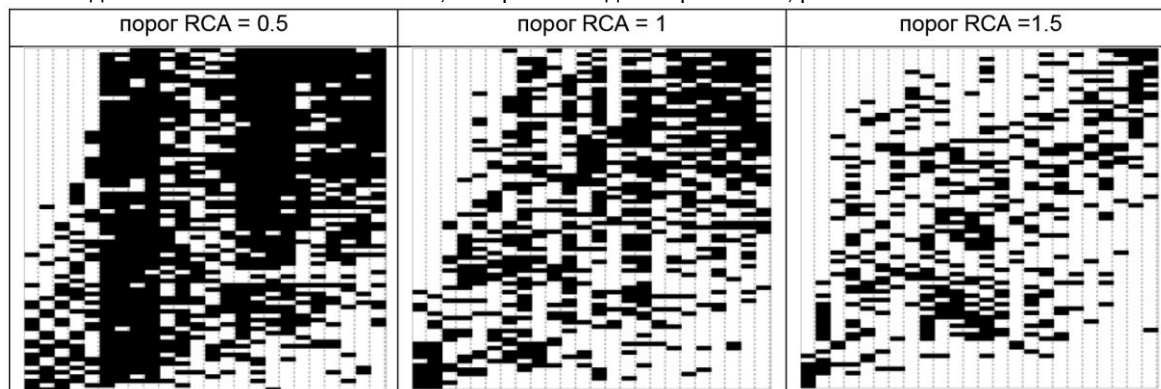


Рис. 3. Матрица 0–1 регион-ВЭД со строками, упорядоченными по ECI_c и столбцами, упорядоченными по ECI_p , для различных порогов RCA

На рисунке 3 представлены 0–1 матрицы, описывающие структуру региональных экономик на основе ВЭД для различных значений порога RCA. Темная ячейка матрицы означает, что соответствующий элемент матрицы равен 1. То есть продукция данного ВЭД производится регионом на уровне выявленных сравнительных преимуществ. Строки каждой матрицы упорядочены снизу вверх по возрастанию оценок ECI_c экономической сложности регионов. Столбцы упорядочены слева направо по возрастанию оценок ECI_p экономической сложности ВЭД. Структура матриц на рисунке 3 отражает те же особенности, которые мы наблюдаем на рисунке 2. Поэтому для сравнительного анализа мы можем

использовать порог RCA 1, убедившись в устойчивости соответствующего индекса экономической сложности.

Таблица 2. Корреляция ECI_c по 24 ВЭД для разных порогов RCA

ECI по ВЭД. Порог	0.5	0.75	1	1.125	1.25	1.5
0.5	1	0.9157	0.8957	0.7361	0.7224	0.2669
0.75	0.9815	1	0.9400	0.7967	0.8027	0.3246
1	0.9385	0.9539	1	0.8222	0.8254	0.3021
1.125	0.8357	0.8513	0.8977	1	0.7560	0.2165
1.25	0.8362	0.8596	0.9351	0.8704	1	0.4746
1.5	0.0689	0.0499	0.0842	0.0147	0.1215	1

В таблице 2 под главной диагональю приведены коэффициенты корреляции Пирсона индексов экономической сложности регионов, построенных по ВЭД. Над главной диагональю – коэффициенты корреляции Спирмена. Устойчивость индекса, построенного по ВЭД при пороге 0,75 несколько выше, чем при пороге 1. Но высокие коэффициенты корреляции Пирсона 0.95 и Спирмена 0.94 этих индексов позволяют нам использовать любой из них. Как видно из таблицы 3, коэффициент корреляции Пирсона индекса по секторам при пороговом значении 1 выше с индексом по ВЭД при пороговом значении 1, чем с индексом по ВЭД при пороговом значении 0,75. Различие в индексах корреляции Спирмена незначительно. Из этих соображений в приложениях имеет смысл использовать индекс экономической сложности по секторам с порогом 1 и индекс экономической сложности по ВЭД с порогом 1.

Таблица 3. Корреляция Пирсона и Спирмена ECI_c и ECI_c^s для порога 1 и 0,75

	ECI по ВЭД. Порог 0,75	ECI по ВЭД. Порог 1
ECI по секторам при пороге 1. Корреляция Пирсона	0,696	0,771
ECI по секторам при пороге 1. Корреляция Спирмена	0,769	0,758

Сравнительный анализ оценок экономической сложности. Сравним оценки экономической сложности 79 регионов, полученные на основе данных о налоговых поступлениях по 82 секторам (столбец (5) таблицы П1 приложения) и на основе данных по отгруженной продукции по 24 ВЭД за 2019 г. (столбец (3) таблицы П1 приложения). Коэффициент корреляции Пирсона индекса экономической сложности регионов по 82 секторам и индекса экономической сложности по 24 ВЭД составляет 0,771. Коэффициент ранговой корреляции Спирмена составляет 0,758.

Также прослеживается относительная близость оценок по наиболее экономически сложным ВЭД и секторам. Это обрабатывающая промышленность ОП2, которая соответствует сектору 1125 – «производство кожи и изделий из кожи». ВЭД и сектора с низкими оценками экономической сложности относятся к добыче полезных ископаемых. У ВЭД – это Д5 «предоставление услуг в области добычи полезных ископаемых», а у секторов – 1060 «добыча природного газа и газового конденсата».

При визуализации зависимости оценок экономической сложности регионов по 82 секторам (рис. 4) можно наблюдать те же тенденции, что были упомянуты ранее для 24 ВЭД (рис. 1).

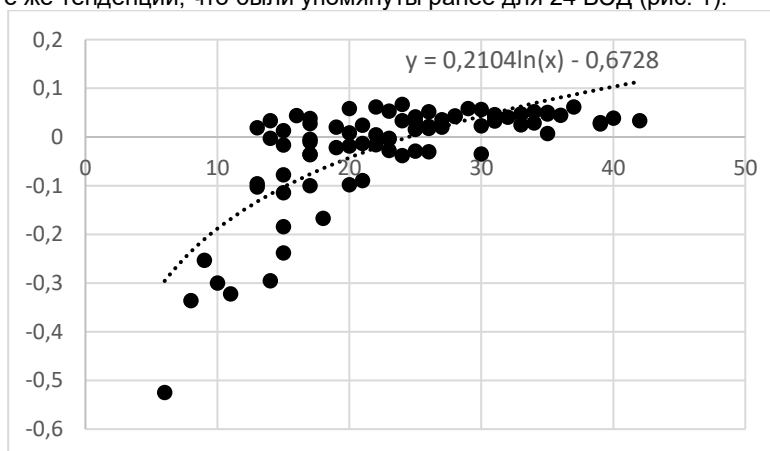


Рис. 4. Зависимость оценок экономической сложности регионов (ось ординат) от числа сильных секторов (ось абсцисс).

На рис. 5 точка характеризует регион. Ее координата по оси абсцисс – оценка экономической сложности регионов по секторам. Координата по оси ординат – оценка по ВЭД. В нижней части рисунка – точки, соответствующие трем добывающим регионам (Оренбургская, Тюменская, Сахалинская области). Эти регионы имеют самые большие объемы производства в секторе 1055 (добыча сырой нефти). В верхнем правом углу регионы с развитыми секторами обрабатывающей и сельскохозяйственной продукции. Регионы, расположенные близко к диагонали, имеют близкие оценки экономической сложности по секторам и ВЭД.

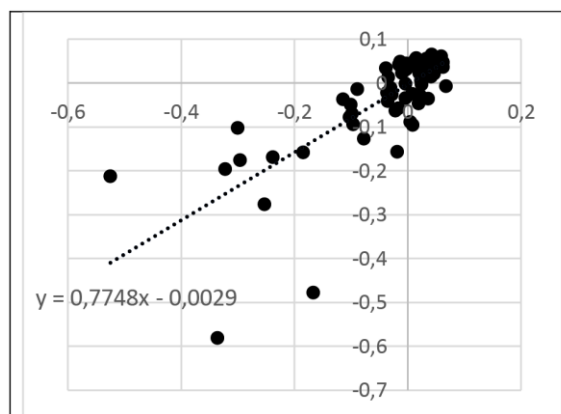


Рис. 5. Оценки экономической сложности регионов по секторам (ось абсцисс) и ВЭД (ось ординат) для 79 регионов

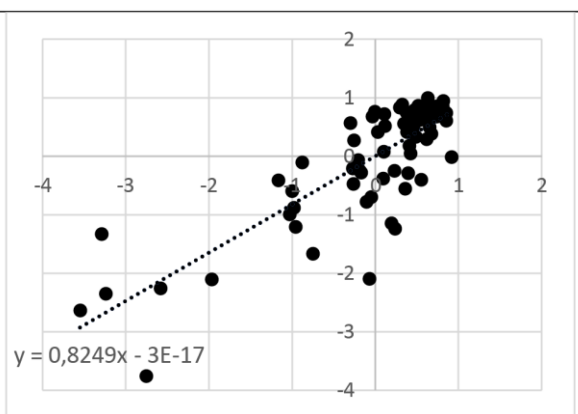


Рис. 6. Оценки экономической сложности регионов по секторам (ось абсцисс) и ВЭД (ось ординат) без трёх добывающих регионов

Отдельный интерес представляет рассмотрение ситуации без трех указанных выше добывающих регионов (рис. 6). При отсутствии трёх добывающих регионов (Оренбургская, Тюменская и Сахалинская области) корреляция оценок возрастает до 0,825. Это говорит о том, что в некоторых случаях могут быть использованы оценки экономической сложности регионов по 24 ВЭД. Оценки экономической сложности регионов имеют высокую устойчивость при переходе от данных по налоговым поступлениям к данным по объемам производства и от данных по секторам к данным по ВЭД.

6. Заключение

Представлен подход к оценке экономической сложности регионов на основе данных об объемах производства по 24 видам экономической деятельности (ВЭД). Расчет оценок экономической сложности выполнен на основе стандартного подхода [Hartmann, 2017; Hausmann, Rodrik 2003; Hausmann, Hwang, Rodrik, 2006; Hidalgo, Hausmann, 2009]. Показано, что регионы с относительно высокими оценками экономической сложности специализируются в обрабатывающих и сельскохозяйственных производствах. Регионы с относительно низкими оценками – в добывающей промышленности.

Рассмотрен вопрос выбора порогового значения RCA. Структура 0–1 матриц для секторов и ВЭД, построенных для близких пороговых значений RCA, правильно отражает идею, заложенную в индексе экономической сложности. Построена корреляционная матрица индексов экономической сложности регионов для разных пороговых значений. Показано, что при пороге 1 индекс экономической сложности по секторам и индекс экономической сложности по ВЭД обладают большей устойчивостью к изменению порога, чем при других пороговых значениях.

Проведён сравнительный анализ оценок экономической сложности 79 регионов по 24 ВЭД и представленных в работе [Афанасьев, Гусев, 2022] оценок экономической сложности по 82 секторам на данных за 2019 г. Их корреляция составила 0,771. Без трех добывающих регионов (Оренбургская, Тюменская и Сахалинская области) корреляция оценок экономической сложности на основе данных о налоговых поступлениях по секторам и объемах отгруженной продукции по ВЭД повышается до 0,825. Таким образом, оценки экономической сложности имеют высокую устойчивость при переходе от данных по налоговым поступлениям к данным по объемам производства и от данных по секторам к данным по ВЭД. Оценка экономической сложности регионов по 24 ВЭД может быть полезна при решении задач управления, направленных на повышение экономической сложности региона.

Литература

1. Айвазян, С. А. Метод кластеризации регионов РФ с учетом отраслевой структуры ВРП / С. А. Айвазян, М. Ю. Афанасьев, А. В. Кудров // Прикладная эконометрика. – 2016. – № 1(41). – С. 24-46.

2. Афанасьев, М. Ю. Аппроксимация оценок экономической сложности при выборе приоритетных направлений диверсификации / М. Ю. Афанасьев, А. А. Гусев // Цифровая экономика. – 2022. – № 1(17). – С. 52–59. – DOI 10.34706/DE-2022-01-05.
3. Hartmann D. (2017). Linking economic complexity, institutions, and income inequality. *World Development*, 93, 75–93.
4. Hausmann R., Hwang J., Rodrik D. (2006). What you export matters. *Journal of Economic Growth*, 12 (1), 1–25.
5. Hausmann R., Klinger B. (2006). Structural transformation and patterns of comparative advantage in the product space. CID Working Paper No. 128.
6. Hausmann R., Rodrik D. (2003). Economic development as selfdiscovery. *Journal of Development Economics*, 72 (2), 603–633.
7. Hidalgo C. (2015). *Why information grows: The evolution of order, from atoms to economies*. New York: Penguin Press.
8. Hidalgo C.A., Hausmann R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106 (26), 10570–10575.
9. Sciarra C., Chiarotti G., Ridolfi L. et al. (2020). Reconciling contrasting views on economic complexity. *Nat Commun*, 11, 3352. DOI: 10.1038/s41467-020-16992-1.

References in Cyrillics

1. Ajvazjan, S. A. Metod klasterizatsii regionov RF s uchetom otraslevoj struktury VRP / S. A. Ajvazjan, M. Ju. Afanas'ev, A. V. Kudrov // *Prikladnaja `ekonometrika*. – 2016. – № 1(41). – С. 24-46.
2. Afanas'ev, M. Ju. Approksimatsija otsenok `ekonomicheskoj slozhnosti pri vybore prioritnykh napravlenij diversifikatsii / M. Ju. Afanas'ev, A. A. Gusev // *Tsifrovaja `ekonomika*. – 2022. – № 1(17). – С. 52-59. – DOI 10.34706/DE-2022-01-05.
3. Hartmann D. (2017). Linking economic complexity, institutions, and income inequality. *World Development*, 93, 75–93.
4. Hausmann R., Hwang J., Rodrik D. (2006). What you export matters. *Journal of Economic Growth*, 12 (1), 1–25.
5. Hausmann R., Klinger B. (2006). Structural transformation and patterns of comparative advantage in the product space. CID Working Paper No. 128.
6. Hausmann R., Rodrik D. (2003). Economic development as selfdiscovery. *Journal of Development Economics*, 72 (2), 603–633.
7. Hidalgo C. (2015). *Why information grows: The evolution of order, from atoms to economies*. New York: Penguin Press.
8. Hidalgo C.A., Hausmann R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106 (26), 10570–10575.
9. Sciarra C., Chiarotti G., Ridolfi L. et al. (2020). Reconciling contrasting views on economic complexity. *Nat Commun*, 11, 3352. DOI: 10.1038/s41467-020-16992-1.

*Афанасьев Михаил Юрьевич, д.э.н.,
заведующий лабораторией прикладной эконометрики ЦЭМИ РАН, Москва,
mi.afan@yandex.ru, ORCID: <https://orcid.org/0000-0001-6928-8821>
Гусев Алексей Александрович, ЦЭМИ РАН, Москва,
gusevalexyal@yandex.ru, ORCID: <https://orcid.org/0000-0002-0551-6001>*

Ключевые слова

Региональная экономика, эконометрика, экономическая сложность.

ABOUT ASSESSING THE ECONOMIC COMPLEXITY OF THE REGIONS

Mikhail Yu. AFANAS'EV, Aleksei A. GUSEV

Mikhail Yu. AFANAS'EV

Central Economics and Mathematics Institute, RAS, (CEMI RAS) Moscow, Russian Federation

mi.afan@yandex.ru, ORCID: <https://orcid.org/0000-0001-6928-8821> Corresponding author

Aleksei A. GUSEV

Central Economics and Mathematics Institute, RAS (CEMI RAS), Moscow, Russian Federation

gusevalexyal@yandex.ru, ORCID: <https://orcid.org/0000-0002-0551-6001>

Keywords

Regional economy, econometrics, economic complexity.

DOI: 10.34706/DE-2023-01-01

JEL classification C02 – Математические методы; C53 – Методы прогнозирования и моделирования

Abstract

An approach to assessing the economic complexity of Russian regions by types of economic activity is proposed. The approach is based on the standard method of assessing economic complexity. The question of choosing the threshold value of the indicator of the identified comparative advantages is considered. Matrices describing the structures of regional economies for sectors and types of economic activity at different thresholds are constructed. Their structures correctly reflect the idea embedded in the economic complexity index. As a result of the correlation analysis, it is shown that at threshold 1, the index of economic complexity by sector and the index of economic complexity by type of economic activity are more resistant to changes in the threshold than at other threshold values. A comparative analysis of estimates of economic complexity for 82 sectors and 24 types of economic activity for 2019 was carried out. Assessment of the economic complexity of regions by types of economic activity can be useful in solving management tasks aimed at diversifying the regional economy and increasing its economic complexity.

Приложение**Таблица П1.** Оценки экономической сложности регионов по ВЭД и секторам

Структура таблицы по столбцам:

- (1) наименование региона;
- (2) число сильных ВЭД в регионе;
- (3) оценки экономической сложности регионов по ВЭД;
- (4) число сильных секторов в регионе (из работы [Афанасьев, Гусев, 2022]);
- (5) оценки экономической сложности регионов по секторам (из работы [Афанасьев, Гусев, 2022]).

(1)	(2)	(3)	(4)	(5)
Белгородская область	4	-0.00795	24	0.0670
Брянская область	11	0.04989	31	0.0402
Владимирская область	13	0.04643	37	0.0617
Воронежская область	12	0.05004	34	0.0285
Ивановская область	12	0.03532	28	0.0430
Калужская область	7	0.06103	29	0.0584
Костромская область	11	0.0253	33	0.0246
Курская область	10	0.0366	22	0.0618
Липецкая область	6	0.02831	36	0.0443
Московская область	17	0.03826	39	0.0266
Орловская область	12	0.04852	30	0.0561
Рязанская область	12	0.04535	16	0.0435
Смоленская область	11	0.0508	31	0.0455
Тамбовская область	11	0.06441	28	0.0419
Тверская область	13	0.04177	42	0.0335
Тульская область	9	0.04189	34	0.0514
Ярославская область	14	0.04206	25	0.0318
г. Москва	10	0.03332	24	-0.0386
Республика Карелия	8	-0.0277	26	0.0215
Республика Коми	8	-0.17566	14	-0.2957
Архангельская область	8	-0.15681	20	-0.0186
Вологодская область	5	0.01381	25	0.0413
Калининградская область	5	0.05247	15	0.0126
Ленинградская область	12	0.03752	14	0.0332
Мурманская область	9	-0.02193	17	-0.0367
Новгородская область	12	0.03768	32	0.0404
Псковская область	13	0.04384	35	0.0503
г. Санкт-Петербург	13	0.04498	23	-0.0031

Республика Адыгея	8	0.04157	22	-0.0160
Краснодарский край	9	-0.03585	27	0.0354
Астраханская область	4	-0.2764	9	-0.2532
Волгоградская область	8	0.03034	17	0.0274
Ростовская область	14	0.02063	33	0.0461
Республика Дагестан	11	0.04681	19	0.0198
Республика Ингушетия	9	-0.05716	15	-0.0168
Кабардино-Балкарская Республика	11	0.05357	17	0.0381
Карачаево-Черкесская Республика	12	0.02282	27	0.0202
Республика Северная Осетия – Алания	13	0.03022	14	-0.0029
Чеченская Республика	9	-0.04686	13	0.0186
Ставропольский край	12	0.04797	23	0.0527
Республика Башкортостан	7	-0.04997	17	-0.1003
Республика Марий Эл	9	0.05488	31	0.0325
Республика Мордовия	8	0.05437	20	0.0582
Республика Татарстан	5	-0.07789	13	-0.1025
Удмуртская Республика	8	-0.12676	15	-0.0780
Чувашская Республика	14	0.054	40	0.0383
Пермский край	7	-0.07007	20	-0.0983
Кировская область	14	0.0477	35	0.0473
Нижегородская область	12	0.04005	24	0.0330
Оренбургская область	6	-0.21304	6	-0.5249
Пензенская область	13	0.05418	26	0.0513
Самарская область	8	-0.0367	15	-0.1144
Саратовская область	15	0.04786	21	-0.0135
Ульяновская область	12	0.05662	25	0.0152
Курганская область	10	0.03317	26	0.0175
Свердловская область	13	0.00622	30	0.0229
Тюменская область	2	-0.58096	8	-0.3363
Челябинская область	7	-0.02482	35	0.0071
Республика Алтай	11	0.01246	30	-0.0352
Республика Тыва	7	-0.03444	25	-0.0047
Республика Хакасия	6	-0.08919	17	0.0044
Алтайский край	11	0.04551	22	0.0473
Красноярский край	5	-0.16893	33	-0.2383
Иркутская область	9	-0.15761	19	-0.1845
Кемеровская область	4	-0.09598	15	0.0080
Новосибирская область	13	0.0172	15	0.0295
Омская область	3	0.02255	20	-0.0104
Томская область	10	-0.10242	39	-0.3000
Республика Бурятия	13	-0.0243	17	-0.0295
Республика Саха (Якутия)	8	-0.19579	10	-0.3225
Забайкальский край	7	-0.06277	11	-0.0221
Камчатский край	9	-0.00123	23	-0.0039
Приморский край	14	-0.01156	26	-0.0307
Хабаровский край	11	-0.01456	21	-0.0896
Амурская область	8	-0.0408	17	-0.0358
Магаданская область	7	-0.02697	23	-0.0278
Сахалинская область	3	-0.47772	18	-0.1673
Еврейская автономная область	10	-0.00344	21	0.0240
Чукотский автономный округ	5	-0.09384	13	-0.0964

Таблица П2. Оценки экономической сложности ВЭД

Структура таблицы по столбцам:

- (1) наименование ВЭД (по возрастанию экономической сложности ВЭД);
- (2) число регионов, в которых ВЭД является сильным;
- (3) оценка экономической сложности ВЭД.

(1)	(2)	(3)
Д5	13	-0.643959
Д2	15	-0.617883
Д1	14	-0.294573
Д3	22	-0.165385
Д4	25	-0.073529
ОП8	21	-0.067646
ОП12	33	-0.053073
ОЭ3	52	-0.019888
ОП3	31	-0.013296
В4	30	-0.001626
ОЭ1	43	0.01941
ОП4	27	0.02714
ОЭ2	48	0.02902
ОП5	14	0.02962
В2	43	0.03215
В1	52	0.03792
ОП6	24	0.06375
ОП10	30	0.07254
В3	37	0.08353
ОП11	31	0.08514
ОП7	38	0.08907
ОП9	35	0.10366
ОП1	41	0.10943
ОП2	33	0.12856

1.2. МОДУЛЬНЫЙ АНАЛИЗ СЛОЖНЫХ СИСТЕМ И СЛОЖНЫЕ СОСТАВНЫЕ СУММЫ

Малиновский В.К., ЦЭМИ РАН, Москва, Россия

Статья посвящена модульному анализу сложных составных сумм случайных величин, возникающих при исследовании различных сложных систем в экономике, технике, и проч. Составные суммы представляют значительный теоретический интерес, например, как вспомогательный инструмент при анализе марковских зависимостей. Они представляют значительный практический интерес, поскольку к таким суммам сводится ряд известных динамических моделей, начиная со стандартных моделей теории восстановления. Сложные составные суммы лежат в основе ряда многопериодических, с несколькими периодами, обычно от трех до семи, динамических моделей. В таких моделях, представляющих интерес для анализа различных количественных методов регулирования, модели, относящиеся к одному периоду, обычно статичны и многокомпонентны.

1. Модульность: способ управления сложностью.

К анализу природных, социальных, экономических и технических сложных систем (см., например, [Anscombe, 1952], [Callebaut et al., 2005], [Holling, 1973], [Ito, 2008], [Langlois, 1999]) применяется модульный подход. Под модульностью (см. [Baldwin et al., 2000]) обычно понимается то, что компоненты сложной системы сперва разделяются на модули, а потом объединяются специальным образом для достижения управляемости системой или для повышения гибкости управления системой.

Модульный подход имеет очевидное практическое значение. Примером является (см. [Brooks, 1975]) разработка операционной системы OS/360 для линейки компьютеров IBM 360, которая сначала была организована неразложимым образом. Изначально считалось, что каждый программист должен быть знаком со всей текущей работой, т.е. должен иметь в своем распоряжении копию общей рабочей документации. Но это быстро привело к нарастанию проблемам в управлении проектом, в частности из-за того, что даже ведение такой документации стало отнимать значительное время из-за постоянного роста ее объема.

Возник вопрос, нужно ли программистам, разрабатывающим один конкретный модуль, знать о деталях всего проекта. Утверждалось, что, особенно в крупных проектах, ответ отрицательный: руководители должны стремиться к минимизации несущественных взаимозависимостей (см. [Parnas, 1972]). Информация, относящаяся только к отдельному модулю, если она не может повлиять на другие составляющие проекта, не должна передаваться прочим исполнителям. Эта идея была сформулирована как «сокрытие информации», ключевое понятие в современной объектно-ориентированном подходе к программированию.

3. Многокомпонентные модели экологического регулирования.

Модульный подход, обычно приводящий к появлению составных сумм, применим ко многим техническим и экономическим проблемам. Приведем два примера многокомпонентных моделей экологического контроля, когда промышленность действует в рамках государственного регулирования.

Пример 1 (Простая модель). Пусть отрасль состоит из ряда производственных цепочек, т. е. групп предприятий, находящихся в зависимости друг от друга. Эти производственные цепочки называются единицами или модулями. Если зависимость между цепочками намного меньше, чем внутри цепочек, то при использовании модульного подхода можно делать предположения о независимости между модулями. Пусть t – общий объем промышленных отходов, например, сточных вод или оксидов углерода, выбрасываемых в атмосферу, который регулирующий орган считает допустимым. Пусть T_i – отходы, выбрасываемые i -м модулем, а X_i – общий продукт, произведенный i – m модулем. В некоторых случаях естественно предположить, что это случайные величины. Тогда $N_t = \max\{k \geq 1: \sum_{i=1}^k T_i \leq t\}$, или 0, если $T_1 > t$ – это количество групп предприятий, или модулей, которые исчерпывают лимит промышленных отходов, а составная сумма $S_{N_t} = \sum_{i=1}^{N_t} X_i$ – это выход, или общий продукт, выпускаемый этими группами.

С одной стороны, эта модель излишне проста. Так, в ней не учитывается динамика по годам, изменение ситуации от года к году, зависимость между группами предприятий, которые считаются независимыми модулями. Однако эта модель лежит в основе построения многошаговой модели по нескольким годам, включающей все перечисленное. Такая модель аналогична многошаговой модели долгосрочного страхового регулирования (см. [Малиновский, 2020]) и будет рассматриваться отдельно.

Модель примера 1 может способствовать выработке гибких и обоснованных требований к производствам, являющимся источником загрязнений. Не запрещая такие производства, органы экологического контроля предъявляют количественные требования на основе анализа изменяющейся по годам (обычно, с горизонтом планирования не выше пяти-семи лет) ситуации с экологической безопасностью. С точки зрения постановки и методов исследования, к модели примера 1 близки модели, возникающие при рассмотрении мер риска в страховании (см. [Malinovskii, 2021a] и [Malinovskii, 2021b]).

Заметим, что ни приведенная выше модель, ни ее многошаговое расширение, не являются динамическими моделями в смысле теории восстановления, широко известными в теории массового обслуживания и теории надежности технических восстанавливаемых систем.

Пример 2 (Усложненная модель). Предположим, что в модели примера 1 каждая производственная цепочка снабжена очистным сооружением, которое частично, полностью или даже избыточно перерабатывает сточные воды или оксиды углерода, т. е. совершает действие, противоположное загрязнению: очистку, рекультивацию или мелиорацию окружающей среды. Тогда T_i могут принимать отрицательные значения. В этом случае общий промышленный выход моделируется сложной составной суммой S_{N_t} , базис которой образован независимыми случайными векторами $(T_i, X_i), i = 1, 2, \dots$, где T_i принимают не только положительные, но и отрицательные значения.

С точки зрения приложений модель примера 1 мало отличается от модели примера 2, но с позиций математического исследования разница существенная: если $T_i, i = 1, 2, \dots$, не положительны, то суммы $V_n := \sum_{i=1}^n T_i$ не возрастают монотонно с ростом n и множество

$$\{N_t = n\} = \{V_1 \leq t, \dots, V_n \leq t < V_{n+1}\}$$

не совпадает с множеством

$$\{N_t = n\} = \{V_n \leq t < V_{n+1}\}.$$

3. Имитационное моделирование и аналитические методы.

Для широкого круга пользователей имитационное моделирование характеристик составной суммы $S_{N_t} := \sum_{i=1}^{N_t} X_i$ представляются естественным и часто единственным

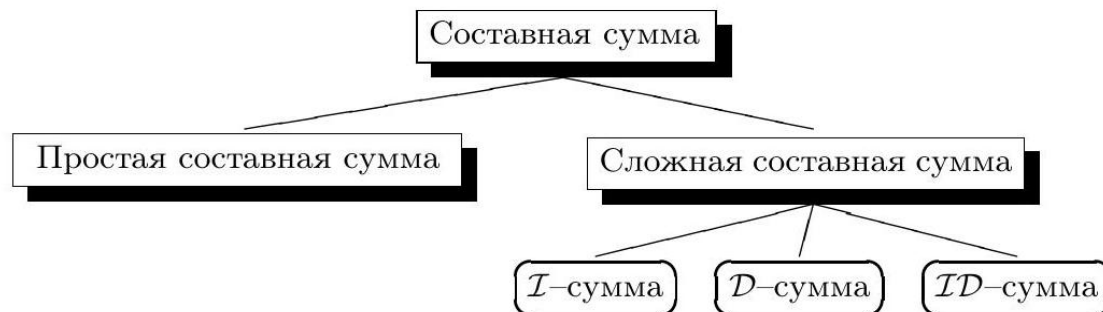


Рис. 1. Классификация составных сумм.

методом реализации модульного анализа. К сожалению, имитационное моделирование имеет ряд врожденных недостатков. Одним из них является относительно низкая точность при заданном объеме вычислений. По существу, имитационное моделирование похоже на определение вероятности выпадения орла и решки с помощью многократного подбрасывания монеты. Известно, что для определения этой вероятности с заданной точностью необходимо произвести большое количество бросков; количественный ответ на вопрос, сколько именно их нужно, дает теорема Муавра-Лапласа. Другой недостаток заключается в том, что имитационное моделирование дает только частные, численные результаты, т. е. охватывает только ситуацию, относящуюся к выбранному набору параметров модели.

Наоборот, если имеется аналитическое решение, то оно дает общее видение для ситуаций, относящихся ко всему набору параметров модели, хотя и в тех условиях, при которых такое решение было получено. Поэтому аналитический подход, в отличие от имитационного моделирования, лучше подходит для выработки общих рекомендаций. Аналитические приближения, даже если не всегда подходят для практических вычислений, много способствуют как качеству, так и эффективности процедур имитационного моделирования.

Простые и сложные составные суммы. В моделях примеров 1 и 2 центральным объектом является составная сумма

$$S_{N_t} := \sum_{i=1}^{N_t} X_i,$$

базисом которой являются двумерные случайные векторы $(T_i, X_i), i = 1, 2, \dots$

По-видимому, впервые составная сумма S_{N_t} впервые появилась (см. [Dœblin, 1938]) в 1938 г. при доказательстве центральной предельной теоремы для дискретной цепи Маркова, где случайные индексы $0 < Y_0 < Y_1 < \dots$ являются моментами последовательных возвращений в некоторое фиксированное состояние. Эти случайные индексы задают модульную структуру, где независимые и одинаково распределенные (н.о.р.) блочные слагаемые $X_i := \sum_{j=Y_{i-1}+1}^{Y_i} \zeta_j$ складываются друг с другом до тех пор, пока соответствующая сумма блочных слагаемых $T_i := Y_i - Y_{i-1}$, положительных по определению, не превысит n . Таким образом, зависимость парадоксальным образом перетекает в независимость. Эта процедура в настоящее время известна как разложение Деблина. Дальнейшие примеры составных сумм S_{N_t} связаны с именами Ф.Дж. Анскомба (см. [Anscombe, 1952]) и Д. Блэквелла. (см. [Blackwell, 1953]) и от-

носятся к началу 1950-х гг. С тех пор такие суммы стали основным инструментом последовательного анализа, теории надежности, теории страховых рисков, теории массового обслуживания и многих других областей прикладной теории вероятностей.

Мы классифицируем (см. Рис. 1) составные суммы следующим образом. Когда случайные векторы $(T_i, X_i): = (X, T), i = 1, 2, \dots$, н.о.р. и их компоненты $T_i, i = 1, 2, \dots$, положительны, соответствующая составная сумма S_{N_t} называется простой и ее базис называется простым.

Когда случайные векторы $(T_i, X_i), i = 1, 2, \dots$, н.о.р. и их компоненты $T_i, i = 1, 2, \dots$, не обязательно положительны, соответствующая составная сумма S_{N_t} является сложной и называется J -суммой. Ее базис называется J -базисом.

Когда случайные векторы $(T_i, X_i), i = 1, 2, \dots$, являются зависимыми (следовательно, метод характеристических функций, который переводит свертки в произведения, не работает в полной мере) а случайные величины $T_i, i = 1, 2, \dots$, положительны, соответствующая сложная составная сумма S_{N_t} называется D суммой. Ее базис называется D -базисом.

Когда случайные векторы $(T_i, X_i), i = 1, 2, \dots$, являются зависимыми, а случайные величины $T_i, i = 1, 2, \dots$, не обязательно положительны, соответствующая сложная составная сумма называется JD -суммой. Ее базис называется JD -базисом.

4. Аппроксимация распределения простой составной суммы.

Обозначим плотность и функцию распределения стандартного нормального распределения через

$$\varphi_{(0,1)}(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \Phi_{(0,1)}(x) := \int_{-\infty}^x \varphi_{(0,1)}(z) dz.$$

Для краткости положим $\mu_T := ET, \mu_X := EX, \sigma_T^2 := DT, \sigma_X^2 := DX$, и $\kappa_{XT} := E((X - \mu_X)(T - \mu_T))$. Для $i, j > 0, i + j > 2$ положим $\kappa_{XT}^{(i,j)} := E((X - EX)^i (T - ET)^j)$. Очевидно, что $\kappa_{TX}^{(i,j)} = \kappa_{XT}^{(j,i)}$. Положим также:

$$\begin{aligned} M_S &:= \frac{\mu_X}{\mu_T}, D_S^2 := \frac{\mu_T^2 \sigma_X^2 - 2\mu_T \mu_X \kappa_{XT} + \mu_X^2 \sigma_T^2}{\mu_T^3} \\ A_S &:= \left(\left(\kappa_{XT}^{(3,0)} - 3\sigma_X^2 \kappa_{XT} \mu_T^{-1} + 6\sigma_X^2 \sigma_T^2 \mu_X \mu_T^{-2} - 6\sigma_T^2 \kappa_{XT} \mu_X^2 \mu_T^{-3} \right) \mu_T^{-1} \right. \\ &\quad \left. - 3 \left(\kappa_{XT}^{(2,1)} - 2\kappa_{XT}^2 \mu_T^{-1} + \sigma_X^2 \sigma_T^2 \mu_T^{-1} \right) \mu_X \mu_T^{-2} \right. \\ &\quad \left. + 3 \left(\kappa_{XT}^{(1,2)} - \sigma_T^2 \kappa_{XT} \mu_T^{-1} \right) \mu_X^2 \mu_T^{-3} - \left(\kappa_{XT}^{(0,3)} - 3\sigma_T^4 \mu_T^{-1} \right) \mu_X^3 \mu_T^{-4} \right) D_S^{-3} \\ B_S &:= \mu_X (\sigma_T^2 \mu_T^{-2} + 1) D_S^{-1} \end{aligned}$$

Приведем пример аналитического результата, относящегося к распределению простой составной суммы.

ТЕОРЕМА 1. Пусть функция плотности случайного вектора (T, X) ограничена сверху конечной константой, $D_S^2 > 0, E(T^4) < \infty, E(X^4) < \infty$. Тогда

$$\begin{aligned} \sup_{x \in \mathbb{R}} | P\{S_{N_t} - M_S t \leq x D_S t^{1/2}\} - \Phi_{(0,1)}(x) \\ + \frac{1}{6} (A_S (x^2 - 1) + 3B_S) \varphi_{(0,1)}(x) t^{-1/2} | = O(t^{-1}), t \rightarrow \infty. \end{aligned}$$

Этот результат, называемый асимптотическим разложением в предельной теореме для распределения составной суммы S_{N_t} , позволяет уменьшить величину t , при котором аппроксимация такого распределения удовлетворительна. В этом результате учитывается корреляция между компонентами случайного вектора (T, X) и третьи моменты этих компонент.

Для использования приближения (1) достаточно вычислить входящие в M_S, D_S^2, A_S, B_S моменты. Это можно делать аналитически, вычисляя интегралы, которыми выражаются моменты $\mu_T, \mu_X, \sigma_T^2, \sigma_X^2$ и проч., и численно, методом имитационного моделирования. Эта численная процедура намного проще той, которая потребовалась бы для расчета распределения сложной суммы S_{N_t} .

5. Метод доказательства Теоремы 1.

Воспользовавшись формулой полной вероятности, имеем

$$\begin{aligned} P\{S_{N_t} \leq x\} &= \sum_{n=1}^{\infty} P\{S_{N_t} \leq x, N_t = n\} \\ &= P\left\{ \sum_{i=1}^n X_i \leq x, \sum_{i=1}^n T_i \leq t < \sum_{i=1}^{n+1} T_i \right\} \\ &= \sum_{n=1}^{\infty} \int_0^t P\left\{ \sum_{i=1}^n X_i \leq x \mid \sum_{i=1}^n T_i = t - z \right\} f_T^{*n}(t - z) \\ &\quad \times P\{T_{n+1} > z\} dz. \end{aligned}$$

Подынтегральные выражения подходят для применения интегро-локальных центральных предельных теорем с неравномерным остаточным членом для н.о.р. двумерных случайных векторов.

Обобщения. Теорема 1 является примером применения общего подхода к исследованию составных сумм. Впервые этот подход был предложен в [Малиновский, 1993]. Существенно более общие результаты получены для трех типов сложных составных сумм (см. Рис. 1): \mathcal{J} -сумм, \mathcal{D} -сумм и \mathcal{JD} -сумм. Эти результаты будут изложены отдельно.

Литература

1. Anscombe, F.J. (1952) Large-sample theory of sequential estimation, Proc. Cambridge Phil. Soc., 48, 600–607.
2. Baldwin, C.Y., and Clark, K.B. (2000) Design Rules. Vol. 1: The Power of Modularity. MIT Press, Cambridge, MA.
3. Blackwell, D. (1953) Extension of a renewal theorem, Pacific J. Math., 3, 315–320.
4. Brooks, F.P. (1975) The Mythical Man-Month: Essays on Software Engineering. Addison-Wesley, Reading.
5. Callebaut, W, and Rasskin-Gutman, D. (2005) Modularity: Understanding the Development and Evolution of Natural Complex Systems. The Vienna Series in Theoretical Biology. MIT Press, Cambridge, MA.
6. Dœblin, W. (1938) Sur deux problèmes de M. Kolmogoroff concernant les chaînes d'énumérables, Bull. Soc. Math. de France, 66, 210–220.
7. Holling, C.S. (1973) Resilience and stability of ecological systems, Annual Review of Ecology and Systematics, 4, 1–23.
8. Ito, Y. (2008) Modular Design for Machine Tools. McGraw-Hill, New York.
9. Langlois, R.N. (1999) Modularity in technology, organization, and society, University of Connecticut Working Paper 1999–05.
10. [Малиновский, В.К. (1993) Предельные теоремы для остановленных случайных последовательностей I. Теор. вер. применен., 38, 4, 890–826.
11. Малиновский, В.К. (2020) Модели долгосрочного страхового планирования. М.: Янус-К.
12. Malinovskii, V.K. (2021) Level-Crossing Problems and Inverse Gaussian Distributions. Chapman and Hall/CRC, Boca Raton.
13. Malinovskii, V.K. (2021) Risk Measures and Insurance Solvency Benchmarks. Chapman and Hall/CRC, Boca Raton,
14. Parnas, D.L. (1972) On the criteria to be used in decomposing systems into modules, Communications of the ACM, 15, 1053–1058.

*Всеволод Константинович Малиновский, д.ф.-м.н.,
Центральный экономико-математический институт (ЦЭМИ РАН), 117418,
Нахимовский просп., 47, Москва, Россия
E-mail address: admin@actlab.ru, Vsevolod.Malinovskii@mail.ru
URL: <http://www.actlab.ru>*

Ключевые слова

Модульный анализ, количественные методы регулирования, суммы случайных величин, сложные составные суммы, предельные теоремы, имитационное моделирование.

Vsevolod Malinovsky, Modular analysis of complex systems and complex composite sums

Keywords

Modular analysis, quantitative approach to regulation, sums of random variables, complex compound sums, limit theorems, simulation.

DOI: 10.34706/DE-2023-01-02

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract

The paper deals with modular analysis of complex compound sums of random variables that arise in the study of various complex systems in economics, technology, etc. The compound sums are of a considerable theoretical interest, e.g., as an auxiliary tool in the analysis of Markov dependencies, and of a considerable practical interest, since a number of well-known dynamical models, starting from standard models of renewal theory, are reduced to them. The complex compound sums underlie a number of multi-periodic, usually comprising from three to seven periods, dynamic models. In such models, which are of interest for the analysis of various quantitative approaches to regulation, the models related to one period are usually static and multi-component.

1.3. РЕАЛИЗАЦИЯ ГЕОМЕТРИЧЕСКОГО ПРЕОБРАЗОВАНИЯ ОТРАЖЕНИЯ ДЛЯ ПРОТОТИПА ВЫСОКОПРОИЗВОДИТЕЛЬНОГО ГЕОМЕТРИЧЕСКОГО ЯДРА

Егунов В.А.¹, Филимонов О.Ю.²

¹ВолгГТУ, г. Волгоград

²ООО «Сингулярис Лаб», г. Волгоград

В данной работе представлены результаты по оптимизации вычислений аффинных преобразований для полигональных моделей, которые широко используются в геометрических ядрах САПР. Рассмотрены основные форматы хранения данных полигональных моделей, приводится матричное представление полигональных моделей. Рассматриваются различные варианты выполнения преобразования отражения над полигональными моделями, описывается процесс векторизации вычислений. Приводится сравнение различных вариантов выполнения.

Введение

Существует определённый спектр задач, для решения которых требуется использование вычислительной геометрии. Примером таких задач может служить вычисление объёма произвольного трёхмерного тела, булевы операции над полигональными моделями, триангуляция многоугольника или более простые задачи по преобразованию тел в пространстве. Такие задачи встречаются в 3D проектировании, а именно в CAD, CAM и CAE системах, в робототехнике, ГИС и т.д.

В докладе, подготовленном коллективом Института статистических исследований и экономики знаний НИУ ВШЭ [Абдрахманова, 2022] отмечается, что как в последние годы, так и в настоящее время, в нашей стране успешно развивались ключевые элементы цифровой экономики. Более того, отмечаются устойчивые тренды цифровой трансформации экономики и социальной сферы. Базой для выполнения данной трансформации являются, в том числе, программные продукты для проектирования и компьютерного инжиниринга – CAD, CAM, CAE и пр. [Полянсков, 2013].

Для таких задач ключевой параметр – это оценка сложности используемых алгоритмов [Дешко, 2021]. Так, разница между алгоритмом с квадратичной и линейно-логарифмической сложностью может выливаться в дополнительные часы или дни вычислений. Для оптимизации времени работы алгоритма, если отсутствует альтернатива в виде идентичного алгоритма с более низкой сложностью, можно использовать различные методы – распараллеливание, векторизация [Егунов, 2020], кэш-оптимизация [Егунов, 2019], вычисление на GPU и т.д.

В данной работе рассматривается эффективность реализаций аффинного преобразования отражения для полигональных моделей.

Геометрические ядра.

Как правило, для решения вышеописанных задач используется геометрическое ядро, если речь идёт о САПР, или его более лёгкий аналог. Геометрическое ядро – набор библиотек с программным интерфейсом, с помощью которого можно пользоваться функциями геометрического моделирования. Наиболее известные геометрические ядра – это, к примеру, Parasolid, Open CASCADE. Последнее геометрическое ядро имеет открытый исходный код, так что для него существует ряд API для различных языков программирования. Из отечественных разработок можно отметить ядро C3D, которое лежит в основе Компас-3D, и RGK (Russian Geometric Kernel) [Стешина, 2015], в процессе разработки которого были написаны несколько статей [Шаповалов, 2013], также задействованы различные компании и университеты [Filimonov, 2021]. Более подробно о самих геометрических ядрах и их роли в мире САПР можно узнать здесь [Цымбал, 2022].

Одной из важных характеристик геометрических ядер является возможность работать с разным представлением данных, что определяет возможности ядра по чтению и загрузке моделей, хранящихся в различных геометрических форматах. Речь здесь может идти как о форматах, разработанных специально для конкретной САПР, так и об общедоступных форматах. Геометрические форматы первой группы, как правило, содержат не только саму модель, но и сцену, куда включается освещение, тени, другие модели и т.д. Из общедоступных форматов можно перечислить следующие: STL, OBJ, COLLADA, FBX, STEP. Стоит отметить, что наиболее популярными являются STL [Stroud, 2000] и COLLADA. У некоторых общедоступных форматов существует проблема универсальности, так как для них нет точного описания спецификации.

Матричное описание полигональной сетки

Полигональная модель описывается набором полигонов и нормалей [Гонахчян, 2014]. В качестве полигона может выступать любой плоский многоугольник, однако на практике используется треугольный полигон. Так, полигональная модель с N полигонами содержит N нормалей и $N * 3$ вершин. Таким

образом, полигональную модель можно представить в виде матриц, показанных на рисунке 1 (широкая матрица) и рисунке 2 (высокая матрица).

Оба матричных представления имеют свои достоинства и недостатки, поэтому в разных алгоритмах будет использоваться представление, описание которого по структуре и по расположению данных в памяти вычислительной машины является более предпочтительным для рассматриваемого алгоритма. К примеру, для широкого представления в кэш попадёт массив однотипных координат (x, y или z), а для высокого представления в кэше будет находиться набор вершин, что для линейных преобразований таких, как вращение и отражение, может значительно снизить время выполнения алгоритма.

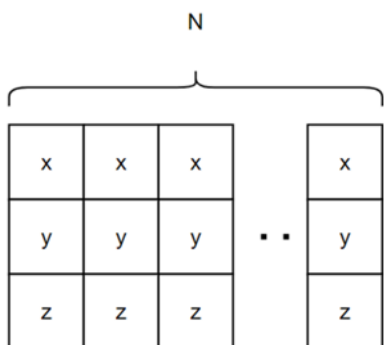


Рисунок 1 Представление в виде матрицы 3xN, N – число вершин

Непосредственно само полигональное описание 3D-модели может использоваться во многих областях: начиная от 3D-творчества в Blender и заканчивая использованием в 3D-печати [Andreev, 2021].

Оптимизация преобразования отражения

В данной работе в качестве оптимизируемой операции рассматривалось преобразование отражения. Данная операция выполняется путем применения матрицы отражения (оператор Хаусхолдера) к набору векторов, которые могут быть записаны в виде матрицы 3xN или Nx3.

$$R = H(n)\bar{x} \tag{1}$$

где H(n) – матрица Хаусхолдера (2), n – нормаль к гиперплоскости, от которой происходит отражение, \bar{x} – отражаемый вектор.

$$H(n) = I - 2nn^T \tag{2}$$

В качестве отражаемого объекта может выступать и набор векторов. Двустороннее преобразование Хаусхолдера часто используется для приведения исходной матрицы к Хессенберговой или двухдиагональной (ленточной) форме. В качестве стороннего используемого метода рассматривается функция LAPACK_?larfx из пакета Intel MKL [Intel, 2015].

Векторизация вычислений

Для повышения производительности операция отражения была векторизована. Для векторизации использовались интринсики (SSE) [Harris, 2012]. Поскольку работа осуществляется с вершинами в трёхмерном пространстве, то использование регистров xmm усложняется тем, что для типа float и double в них можно поместить по 4 и 2 элемента соответственно. Одна вершина описывается 3 точками. Это значит, что для типов float и double требуется одновременная обработка 4 и 2 вершин соответственно для заполнения векторов. Как будет показано далее, наиболее предпочтительным для векторизации является формат Nx3, опишем векторизацию алгоритма с использованием данного формата. На рисунке 3 показано распределение координат 4 вершин в трёх регистрах XMM для типа float в представлении Nx3

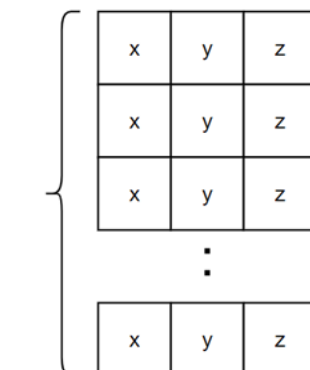


Рисунок 2 Представление в виде матрицы Nx3, N – число вершин

x1	y1	z1
x2	y2	z2
x3	y3	z3
x4	y4	z4

Рисунок 3. – Распределение координат 4 вершин в трёх регистрах XMM для типа float в представлении Nx3

$$H = \begin{bmatrix} a & f & e \\ f & b & g \\ e & g & c \end{bmatrix} \tag{3}$$

Всего требуется вычислить 3 квартета значений, что разбивает один проход на 3 отдельных шага, которые описаны на рисунках 4, 5 и 6. Вектора множителей (к примеру [a, f, e, a]), образованные из матрицы H, циклично сдвигаются на каждом шаге влево так, что последним элементом становится новый первый, то есть вектор [a, f, e, a] на первом шаге станет [f, e, a, f] на втором. Аналогичный сдвиг будет выполнен и на шаге 3. Выполнение векторизации для типа double осуществляется аналогично, за тем исключением формирования вектора множителей.

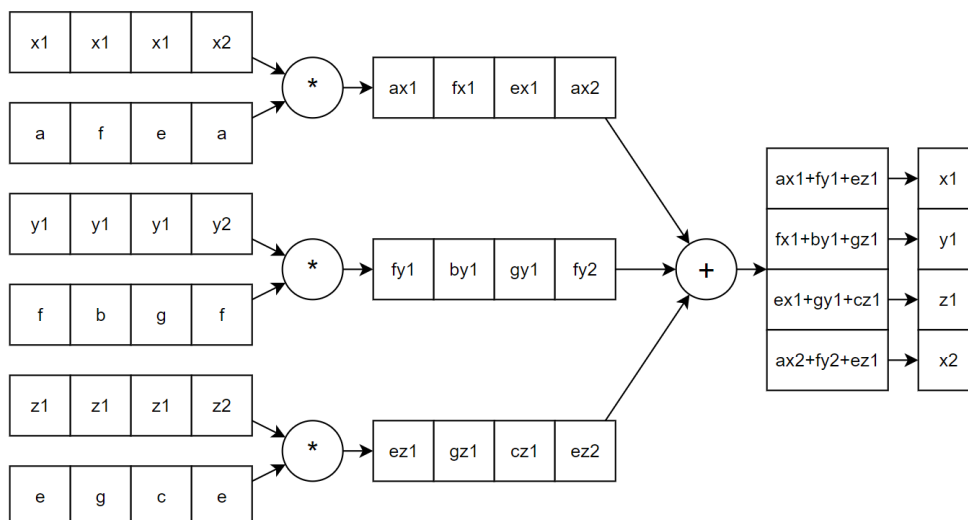


Рисунок 4 – Схема выполнение 1/3 части шага для векторизованной операции отражения.
Тип данных – float

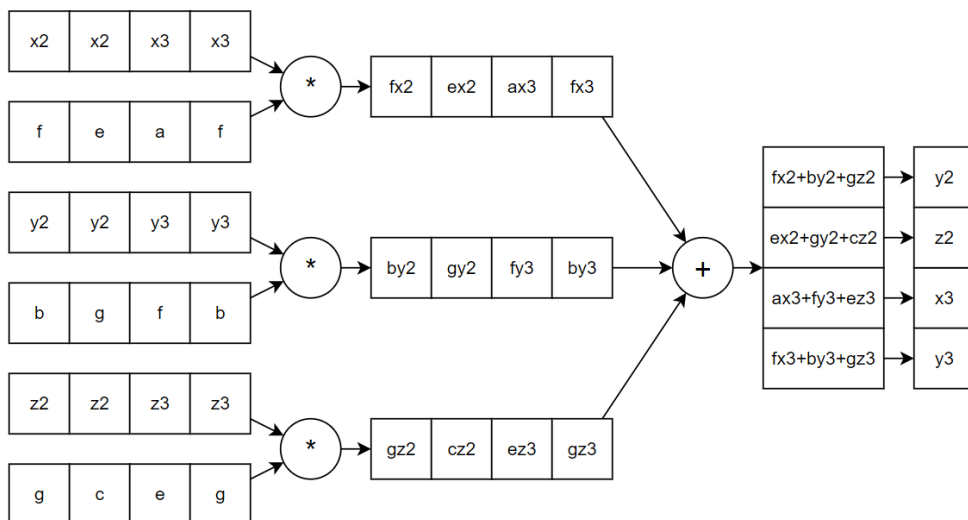


Рисунок 5 – Схема выполнение 2/3 части шага для векторизованной операции отражения.
Тип данных – float

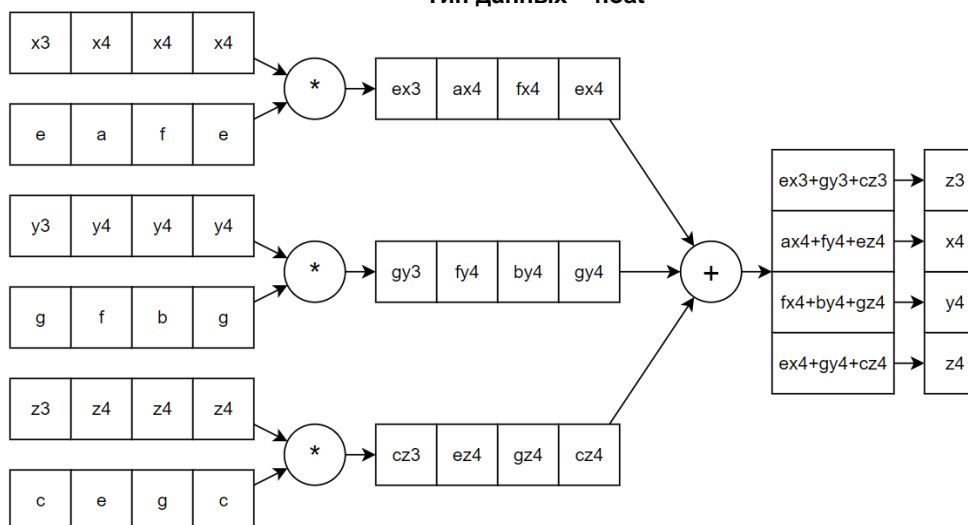


Рисунок 6 – Схема выполнение 3/3 части шага для векторизованной операции отражения.
Тип данных – float

Результаты исследования

Результаты вычислительных экспериментов представлены на рисунках 7, 8, 9. На рисунках представлены обозначения в формате name_format_type. Здесь name – название алгоритма, format – формат хранения данных, type – тип используемых данных. Name принимает значения:

- standart – матричная форма реализации отражения Хаусхолдера;
- hessenberg – векторная форма реализации отражения Хаусхолдера;
- vectorized_standart – векторизованный вариант матричной формы реализации отражения Хаусхолдера;
- vectorized_hessenberg – векторизованный формат векторной формы реализации отражения Хаусхолдера;
- mkl – реализация Intel MKL.

Format может принимать значения 3xN и Nx3, сами форматы описаны выше. Type может принимать значения double и float, что соответствует типам используемых данных (программы были написаны на C++).

Например, standart_nx3_float означает реализацию матричной формы реализации отражения Хаусхолдера, формат хранения – «высокая матрица», тип данных – float.

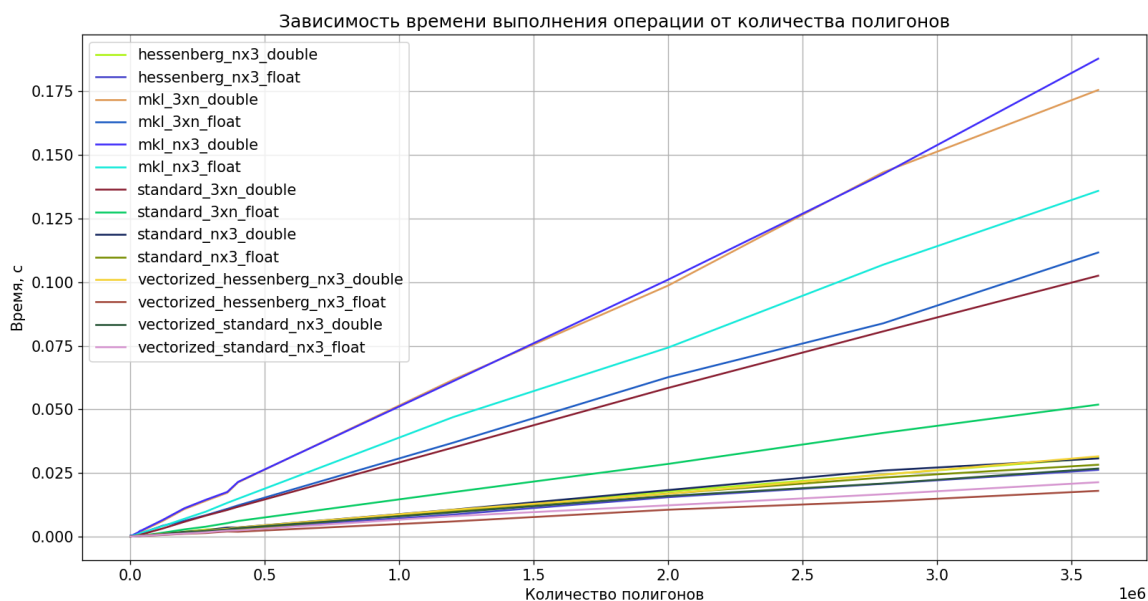


Рисунок 7 – Зависимость времени выполнения операций от количества полигонов

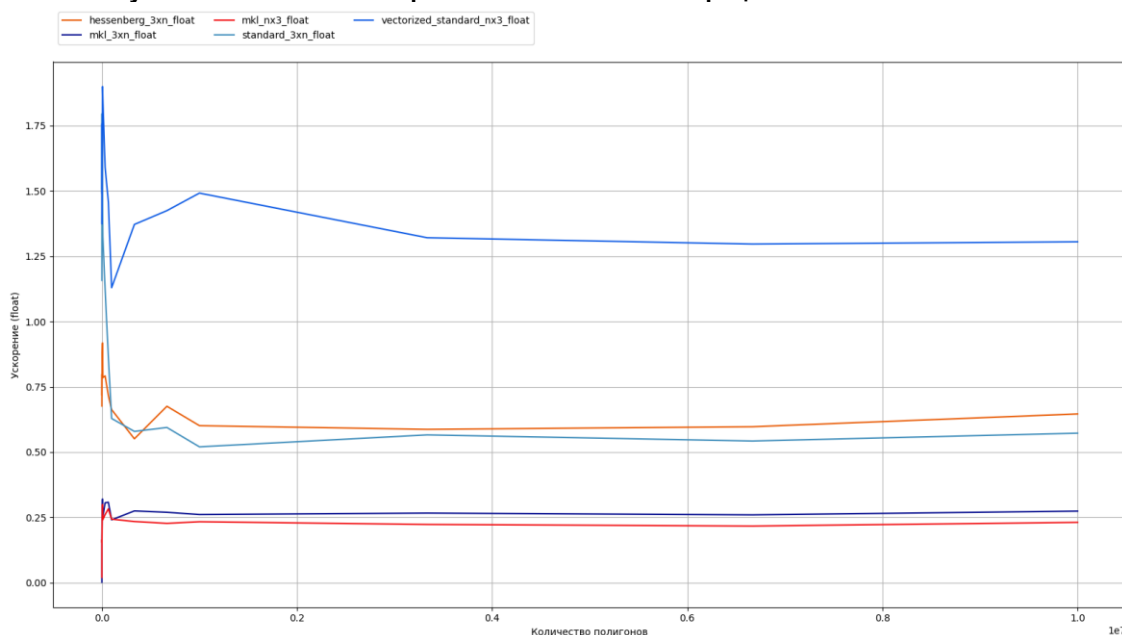


Рисунок 8 – Зависимость ускорения от количества полигонов. Тип float

Анализируя зависимости на рисунке 7, можно сделать следующие выводы:

- в общем случае формат хранения Nx3 является более предпочтительным;
- векторная форма реализации отражения Хаусхолдера является более предпочтительной;
- векторизация уменьшает общее время выполнения программы для всех методов;
- функция библиотеки Intel MKL в данном случае отработала плохо.

Самым быстрым алгоритмом является `vectorized_hessenberg_nx3` (в обеих группах алгоритмов – `float` и `double`), самым медленным – `mkl_3xn`.

Далее было проанализировано ускорение полученных программ по отношению к эталонному методу – `standart_nx3_[float|double]`.

Видно, что быстрее работает только векторизованная версия эталонного метода, все остальные реализации проигрывают, причем MKL-реализация проигрывает значительно.

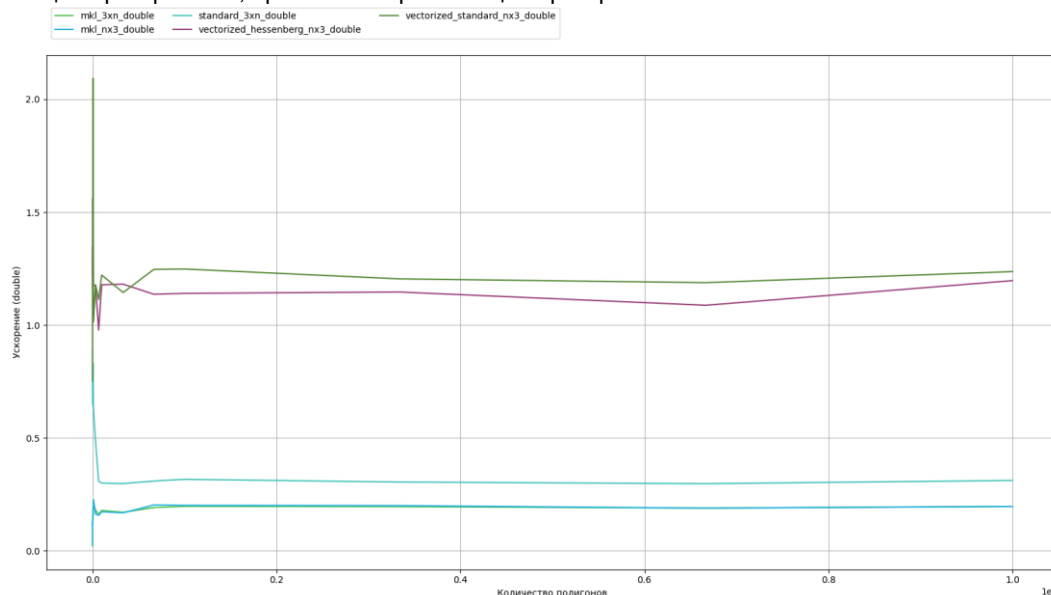


Рисунок 9 – Зависимость ускорения от количества полигонов. Тип double

Здесь также преимущество у векторизованных Nx3-схем, остальные, включая MKL-реализации, показывают гораздо худшие результаты.

В целом можно сказать, что наблюдаются те же тренды, что и на рисунке 7, при этом с некоторого количества полигонов ускорение начинает вести себя линейно. Однако на практике нечасто используются полигональные модели с порядком 10^6 . Поэтому имеет смысл проанализировать ускорения на малом количестве полигонов (до 10000).

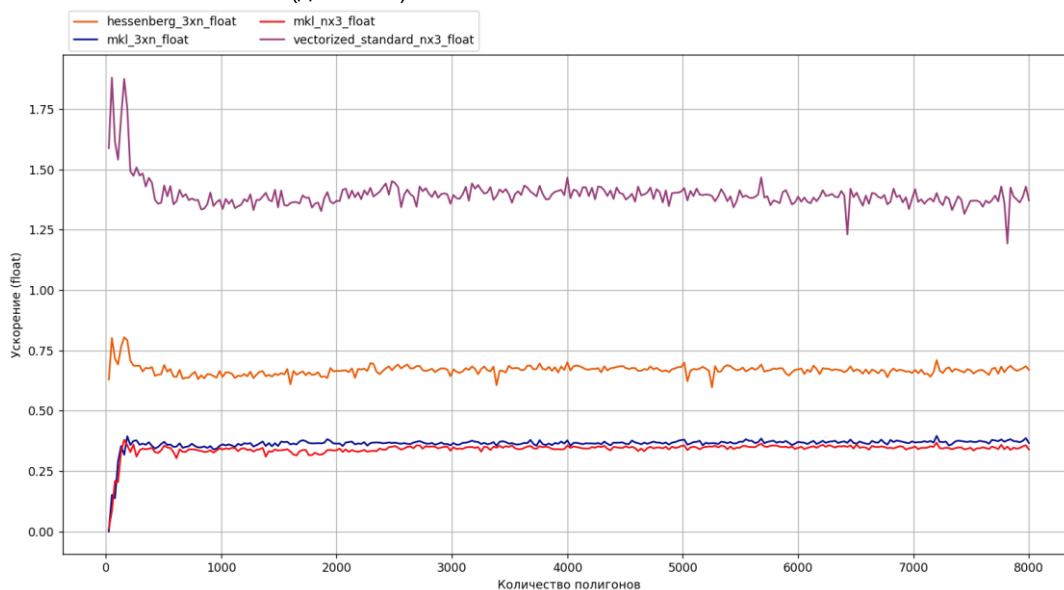


Рисунок 10 – Зависимость ускорения от количества полигонов. Тип float

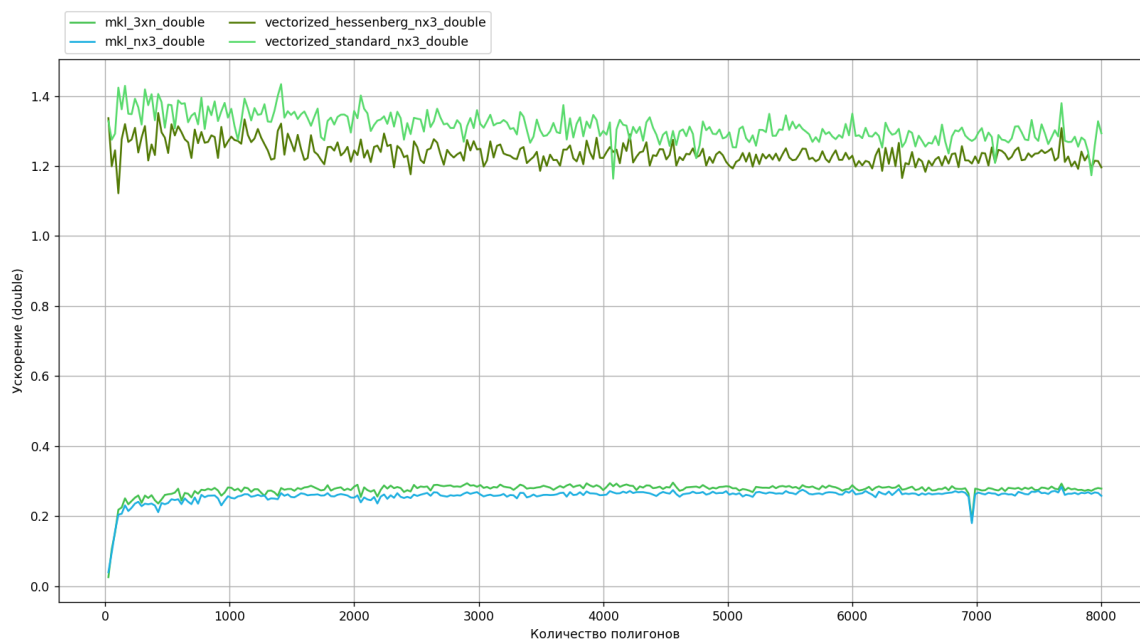


Рисунок 11 – Зависимость ускорения от количества полигонов. Тип double

Из графиков 10 и 11 видно, что ускорение на малом количестве полигонов также ведёт себя линейно.

Выводы

По результатам оптимизации операции отражения можно заключить, что оптимальным форматом представления вершин является Nx3, его использование дает выигрыш во времени до 2,5 раз. Методы библиотеки Intel MKL оказались неэффективными, также неэффективными оказались методы, основанные на обработке полигональной модели в формате 3xN. Гораздо более предпочтительным оказалось использование векторной формы реализации преобразования отражения, при этом векторизация вычислений дает существенный, до 2 раз, прирост производительности.

Так как данные о вершинах полигональных моделей, как правило, хранятся с использованием внешней индексации, полученные методы могут быть использованы для выполнения преобразования отражения в геометрических ядрах. Направление дальнейших исследований будет связано с исследованием реализации других базовых геометрических операций, таких как параллельный перенос, скалирование, вращение, отражение, сдвиг и т.д., для реализации на полигональных моделях.

Литература

1. Абдрахманова Г. И., Васильковский С. А., Вишневский, М. А. Гершман, Л. М. Гохберг и др. Цифровая трансформация: ожидания и реальность: докл. к XXIII Ясинской (Апрельской) междунар. науч. конф. по проблемам развития экономики и общества, Москва, 2022 г. [Текст] / рук. авт. кол. П. Б. Рудник; Нац. исслед. ун-т «Высшая школа экономики». — М.: Изд. дом Высшей школы экономики, 2022. — 221 с. — ISBN 978-5-7598-2658-3 (в обл.). — ISBN 978-5-7598-2468-8 (e-book).
2. Гонахчян В.И. Обзор методов упрощения полигональных моделей на графическом процессоре. Труды Института системного программирования РАН. 2014. Т. 26. № 2. С. 159-174.
3. Дешко И.П., Цветков В.Я. Оценка сложности алгоритма. Славянский форум. 2021. № 3 (33). С. 38-49.
4. Егунов, В.А., Андреев А.Е. Векторизация алгоритмов выполнения собственного и сингулярного разложений матриц с использованием преобразования Хаусхолдера / Прикаспийский журнал: управление и высокие технологии. — 2020. — № 2 (50). — С. 71-85.
5. Егунов, В.А. Кэш-оптимизация процесса вычисления собственных значений на параллельных вычислительных системах / В.А. Егунов // Прикаспийский журнал: управление и высокие технологии. — 2019. — № 1 (45). — С. 154-163.
6. Полянсков Ю.В. Интеграция САПР-, PDM-, ERP-систем в единое информационное пространство производственного предприятия. / Ю.В. Полянсков, А.С. Кондратьева, М.С. Черников, А.А. Блюменштейн. — Известия Самарского научного центра Российской академии наук, т. 15, № 4(3), 2013.
7. Шестина Д.А., Кочан И.Н. T-FLEX PLM: импортозамещение – это реально! Автоматизация в промышленности. 2015. № 1. С. 47-49.
8. Цымбал Д.В., Юров А.Н., Степенко С.А. Анализ геометрических ядер. «Информатика: проблемы, методы, технологии», 2022, стр. 1284–1291.

9. Шаповалов О.В., Сергеев Е.С., Чалышев В.С., Катаев А.В., Крыжановский Д.И., Андреев А.Е. Вопрос распараллеливания в разработке ядра геометрического моделирования. Научный сервис в сети Интернет: все грани параллелизма, 2013.
10. Andreev, A., Andreeva, M., Drobotov, A., Shmeleva, A., Denisov, M. Development of a Discrete Slicer for Additive Manufacturing. Communications in Computer and Information Science, 2021, 1448 CCIS, pg. 267–281.
11. Filimonov, O.Y., Egunov, V.A., Nesterenko, E.N. Constructing Equidistant Curve for Planar Composite Curve in CAD Systems. Communications in Computer and Information Science, 2021, 1448 CCIS, pg. 296–309.
12. Harris D., Harris S. Digital Design and Computer Architecture // 2nd Edition. NY, Morgan Kaufmann, 2012. 712 p.
13. Intel. Math Kernel Library. <https://software.intel.com/en-us/intel-mkl>. (2015).
14. Stroud I., Xirouchakis P.C. STL AND EXTENSIONS. Advances in Engineering Software. 2000. Т. 31. № 2. С. 83–95.

References in Cyrillics

1. Abdrahmanova G. I., Vasil'kovskij S. A., Vishnevskij, M. A. Gershman, L. M. Gohberg i dr. Cifrovaya transformaciya: ozhidaniya i real'nost': dokl. k XXIII YAsinskoj (Aprel'skoj) mezhdunar. nauch. konf. po problemam razvitiya ekonomiki i obshchestva, Moskva, 2022 g. [Tekst] / ruk. avt. kol. P. B. Rudnik; Nac. issled. un-t «Vysshaya shkola ekonomiki». — M.: Izd. dom Vysshej shkoly ekonomiki, 2022. — 221 s. — ISBN 978-5-7598-2658-3 (v obl.). — ISBN 978-5-7598-2468-8 (e-book).
2. Gonahchyan V.I. Obzor metodov uproscheniya poligonal'nyh modelej na graficheskom processore. Trudy Instituta sistemnogo programirovaniya RAN. 2014. Т. 26. № 2. S. 159-174.
3. Deshko I.P., Cvetkov V.YA. Ocenka slozhnosti algoritma. Slavyanskij forum. 2021. № 3 (33). S. 38-49.
4. Egunov, V.A., Andreev A.E. Vektorizaciya algoritmov vypolneniya sobstvennogo i singulyarnogo razlozhenij matric s ispol'zovaniem preobrazovaniya Hauskholdera / Prikaspijskij zhurnal: upravlenie i vysokie tekhnologii. — 2020. — № 2 (50). — С. 71-85.
5. Egunov, V.A. Kesh-optimizaciya processa vychisleniya sobstvennyh znachenij na parallel'nyh vychislitel'nyh sistemah / V.A. Egunov // Prikaspijskij zhurnal: upravlenie i vysokie tekhnologii. — 2019. — № 1 (45). — С. 154-163.
6. Polyanskov YU.V. Integraciya SAPR-, PDM-, ERP-sistem v edinoe informacionnoe prostranstvo proizvodstvennogo predpriyatiya. / YU.V. Polyanskov, A.S. Kondrat'eva, M.S. Chernikov, A.A. Blyumen-shtejn. — Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk, t. 15, № 4(3), 2013.
7. Steshina D.A., Kochan I.N. T-FLEX PLM: importozameshchenie – eto real'no! Avtomatizaciya v promyshlennosti. 2015. № 1. S. 47-49.
8. Cymbal D.V., YUrov A.N., Stepenko S.A. Analiz geometricheskikh yader. «Informatika: problemy, metody, tekhnologii», 2022, str. 1284–1291.
9. SHapovalov O.V., Sergeev E.S., CHalyshev V.S., Kataev A.V., Kryzhanovskij D.I., Andreev A.E. Vopros rasparallelivaniya v razrabotke yadra geometricheskogo modelirovaniya. Nauchnyj servis v seti Internet: vse grani parallelizma, 2013.

Ключевые слова

Триангуляция, полигональные модели, матричное представление, оптимизация вычислений, интринсики, аффинные преобразования, преобразование Хаусхолдера, векторизация

Егунов Виталий Алексеевич – доцент кафедры «ЭВМ и системы»
Волгоградского государственного технического университета, Волгоград, Россия,
ORCID: 0000-0001-9087-3275, vegunov@mail.com

Филимонов Олег Юрьевич – инженер-программист ООО «Сингулярис Лаб», Волгоград, Россия
filimonovoleg@bk.ru

Keywords

Triangulation, polygonal models, matrix representation, optimization of calculations, intrinsics, affine transformations, Householder transformation, vectorization

Vitaly Egunov, Oleg Filimonov, Implementation of geometric reflection transformation for a prototype of a high-performance geometric kernel

DOI: 10.34706/DE-2023-01-03

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract. This paper presents the results on optimization of calculations of affine transformations for polygonal models, which are widely used in geometric CAD kernels. The main formats of data storage of polygonal models are considered, a matrix representation of polygonal models is given. Various variants of the reflection transformation over polygonal models are considered, the process of vectorization of calculations is described. A comparison of different execution options is given.

1.4. ПРИМЕНЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ СРЕД ДЛЯ УСКОРЕНИЯ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

Шатравин В. Шашев Д. В.

Томский государственный университет (ТГУ), Томск, Россия

Одним из основных факторов, ограничивающих применение современных алгоритмов машинного обучения в технических системах, является несовершенство используемого аппаратного обеспечения. Особенно остро проблема стоит для крупных нейронных сетей в маломощных и автономных системах, имеющих жесткие ограничения к массе и энергопотреблению. Большинство предлагаемых на сегодняшний день аппаратных ускорителей нейронных сетей либо имеют высокое энергопотребление и массу, либо поддерживают лишь очень ограниченное множество алгоритмов. Решением этой проблемы может быть применение перестраиваемых аппаратных ускорителей, которые поддерживают динамическую настройку на реализацию требуемых алгоритмов. Одним из способов построения таких ускорителей могут быть решения на основе концепции перестраиваемых вычислительных сред (ПВС). В данной работе представлена реализация рекуррентных архитектур нейронных сетей на примере сети Хопфилда и сети долгой краткосрочной памяти (LSTM) на ускорителях, построенных на основе ПВС. Приведены формулы оценки быстродействия разработанных моделей на основе результатов симуляций на FPGA. Полученные оценки показывают высокое быстродействие предложенных моделей в сравнении с существующими аналогами при значительно большей занимаемой на полупроводнике площади. Согласно оценкам, расчёт одного шага LSTM сети с 25 скрытыми нейронами занимает 223 нс. Результаты позволяют сделать вывод о большом потенциале применения перестраиваемых сред для реализации рекуррентных сетей и необходимости дальнейших оптимизаций предложенных моделей.

Введение

Активное развитие алгоритмов машинного обучения открывает перед техническими системами новые возможности в области решения задач распознавания образов, обработки естественного языка, предсказательной аналитики и многих других. В первую очередь это достигается при помощи глубоких свёрточных и рекуррентных нейронных сетей. В рамках данной работы наибольший интерес представляют рекуррентные сети, отличительной особенностью которых является наличие обратных связей между нейронами. Обратные связи создают эффект памяти, что позволяет рекуррентным сетям накапливать информацию, то есть учитывать результаты прошлых итераций при обработке новых входных сигналов. Такие сети наиболее эффективны при работе с данными, представленными в виде последовательности – устная речь, текст, динамика котировок, результаты измерений датчика. На сегодняшний день рекуррентные сети представляют собой широкий класс разнообразных архитектур, среди которых наиболее известной являются сети с долгой краткосрочной памятью (LSTM).

В то же время хорошо известна проблема вычислительной сложности алгоритмов машинного обучения. Современные нейронные сети могут насчитывать сотни миллиардов параметров, что делает применение классических вычислительных устройств на основе центральных процессоров неэффективным с точки зрения скорости вычислений и затрачиваемой энергии.

В связи с этим на практике большую популярность получили специализированные под нейросетевые алгоритмы вычислительные устройства, архитектура которых учитывает особенности осуществляемых преобразований. Такие вычислительные устройства называют аппаратными ускорителями нейронных сетей. Аппаратные ускорители показывают высокую эффективность в алгоритмах машинного обучения благодаря распараллеливанию матричных операций и оптимизации доступа к памяти. В качестве аппаратных платформ таких устройств выступают графические процессоры (GPU), программируемые логические интегральные схемы (FPGA) и интегральные схемы специального назначения (ASIC) [Chen; 2019; He, 2021; Ghimire, 2022]. К последним, в частности, относятся тензорные процессоры компаний Google и Huawei, ускоритель Eyeriss, а также многие другие. Каждая из перечисленных платформ имеет свои особенности и предпочтительные области применения.

В рамках данной статьи будут рассматриваться аппаратные ускорители для широкого класса маломощных мобильных и автономных систем, к которым относятся БПЛА, смартфоны, мобильные роботы, спутниковые системы, умные датчики интернета вещей и многие другие. Для таких систем наиболее характерно применение вычислителей на основе FPGA и ASIC в связи с их высокой производительностью при умеренном энергопотреблении. Однако анализ показывает, что большинство существующих решений специализируются на узком классе архитектур нейронных сетей и поддерживают относительно небольшой набор алгоритмов. В то же время сложные мобильные системы могут нуждаться сразу в нескольких нейронных сетях с различными архитектурами, что приводит к необходимости

сти использования сразу нескольких ускорителей, каждый из которых будет специализирован под свой класс нейронных сетей.

Одним из путей решения этой проблемы может быть применение перестраиваемых аппаратных ускорителей, которые поддерживают широкий класс алгоритмов, но в каждый момент времени реализуют только некоторые из них. В основе архитектуры таких ускорителей могут лежать перестраиваемые вычислительные среды (называемые также однородными), которые обеспечивают не только высокую степень параллельности вычислений, но и динамическую низкоуровневую настройку отдельных участков, и высокую надёжность.

Однако применение аппаратных ускорителей на перестраиваемых средах для реализации нейронных сетей конкретных архитектур требует разработки специализированных алгоритмов. В работах [Shatravin, 2021, 2022] предложены алгоритмы для реализации глубоких сетей прямого распространения, а также сигмоидной функции активации нейронов. В данной работе будут представлены алгоритмы реализации на перестраиваемых средах рекуррентных нейронных сетей на примере сети Хопфилда и сети с долгой краткосрочной памятью (LSTM).

1. Рекуррентные нейронные сети

Рекуррентные нейронные сети – широкий класс архитектур нейронных сетей, отличительной особенностью которых является использование полученных на выходе сети результатов при обработке новых входных сигналов. Это достигается благодаря наличию обратных связей, через которые выходной сигнал после некоторых преобразований попадает обратно на вход сети, где проходит обработку вместе с новым входным сигналом. Это позволяет рекуррентным сетям эффективно работать с данными, представляющие собой последовательность – динамика изменения параметров некоторой системы, речь, временные ряды и т.д.

По характеру связей между входными и выходными сигналами выделяют несколько типов рекуррентных сетей: «один вход к одному выходу», «один вход ко многим выходам», «многие входы к одному выходу», «многие входы к многим выходам» (рис. 1).

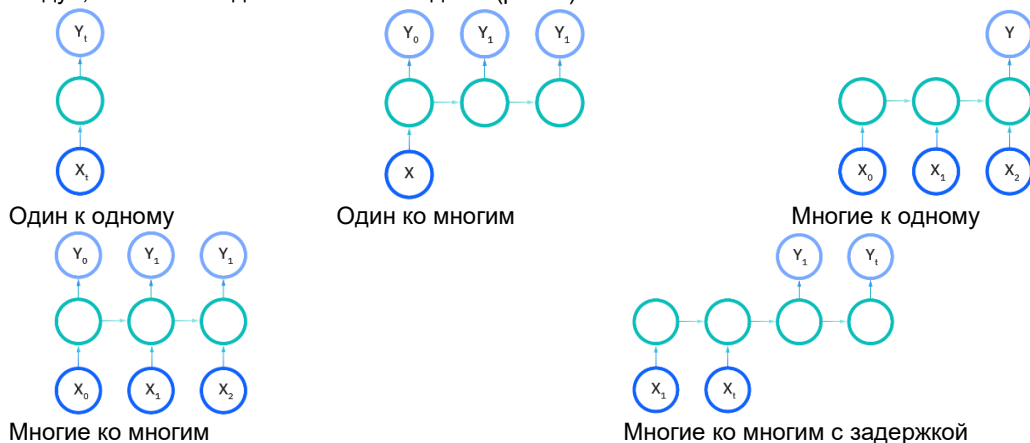


Рис. 1. Типы рекуррентных сетей

В частности, рассматриваемая далее сеть Хопфилда относится к типу «один к одному», так как каждому входному сигналу соответствует одно устойчивое состояние, к которому за некоторое количество шагов сойдётся сеть, а LSTM – к типу «много ко многим», так как на каждом такте на сеть подаётся новый входной сигнал и выводится текущий результат.

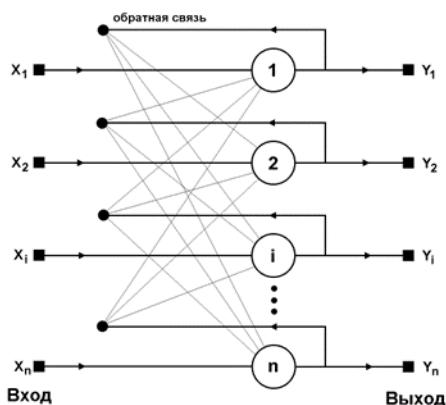


Рис. 2. Рекуррентная сеть Хопфилда

Упомянутая ранее рекуррентная сеть Хопфилда – одна из классических архитектур рекуррентных нейронных сетей. Она состоит из одного слоя нейронов, причём выход каждого нейрона соединён с входами всех остальных нейронов через обратные связи (рис. 2). Сеть Хопфилда имеет симметричную матрицу связей, оперирует двоичными одноуровневыми сигналами и использует пороговую функцию активации нейронов (рис. 3). Попав на вход такой сети, сигнал циклически перемещается в ней до перехода сети в некоторое устойчивое состояние, то есть состояние, при котором значение сигнала (энергия сети) прекращает дальнейшие изменения. Полученный результат соответствует одному из образов, на которые данная сеть была обучена. Иными

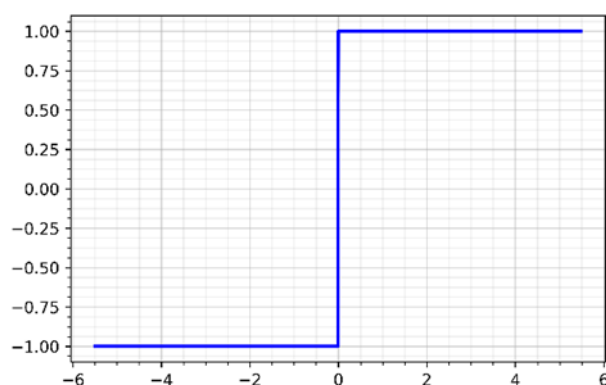


Рис. 3. Пороговая функция активации нейронов в сети Хопфилда

блема «забывания» — результаты с предыдущих тактов относительно быстро заменяются более новыми данными, что препятствует анализу длинных цепочек связанных событий. LSTM сети получили своё название благодаря способности долговременно хранить ключевую информацию с предыдущих тактов и управлять её изменением [Hochreiter, 1997].

Для понимания основных принципов работы LSTM сети рассмотрим её внутреннее устройство (рис. 4) [Olah, 2015]. Долговременное хранение необходимой для работы сети информации обеспечивается введением состояния ячейки (сигнал C_t). На каждом шаге работы сети этот сигнал модифицируется определённым образом: из него устраняется уже неактуальная информация (пурпурный участок сети) и добавляется новая информация, соответствующая текущему входному сигналу x_t (зелёный участок сети).

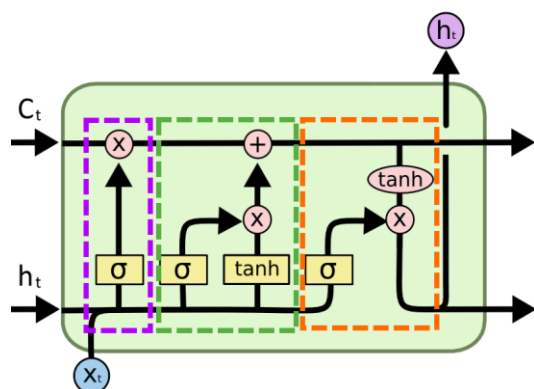


Рис. 4. Строение LSTM сети

словами, сеть Хопфилда выполняет функцию автоассоциативной памяти — восстанавливает исходный образ до одного из заданных шаблонов. В связи с тем, что по мере обработки сигнала энергия сети уменьшается, сеть Хопфилда находит применение в решении задач оптимизации

Гораздо более широкое распространение получили рекуррентные сети с долгой краткосрочной памятью (Long short-term memory, LSTM). LSTM сети относятся к типу «многие ко многим», которые на каждом такте получают новый элемент входной последовательности и возвращают результат, учитывающий не только текущее значение на входе, но и значения, вычисленные на предыдущих тактах. Для многих рекуррентных архитектур характерна проблема

«забывания» — результаты с предыдущих тактов относительно быстро заменяются более новыми данными, что препятствует анализу длинных цепочек связанных событий. LSTM сети получили своё название благодаря способности долговременно хранить ключевую информацию с предыдущих тактов и управлять её изменением [Hochreiter, 1997]. Для понимания основных принципов работы LSTM сети рассмотрим её внутреннее устройство (рис. 4) [Olah, 2015]. Долговременное хранение необходимой для работы сети информации обеспечивается введением состояния ячейки (сигнал C_t). На каждом шаге работы сети этот сигнал модифицируется определённым образом: из него устраняется уже неактуальная информация (пурпурный участок сети) и добавляется новая информация, соответствующая текущему входному сигналу x_t (зелёный участок сети). После этого из текущего входного сигнала x_t и состояния ячейки C_t формируется новый выходной сигнал h_t (оранжевый участок сети). Этот процесс повторяется для каждого нового элемента входной последовательности X . Можно видеть, что описанная архитектура сети состоит из нескольких слоёв нейронов с функциями активации сигмоида и гиперболический тангенс, а также операций поточечного сложения и умножения. LSTM сети активно используются при решении широкого класса задач, в том числе для обработки естественного языка, текстов, хронологически упорядоченных событий. В связи с большой сложностью необходимых в данных задачах сетей остро стоит вопрос их аппаратного ускорения.

2. Аппаратные ускорители на перестраиваемых вычислительных средах

Как было сказано ранее, одним из перспективных направлений развития аппаратных ускорителей нейронных сетей является разработка реконфигурируемых ускорителей на основе перестраиваемых вычислительных сред (ПВС). Благодаря способности к перестраиваемости такие ускорители могут поддерживать широкий класс нейросетевых алгоритмов, динамически настраиваясь на требуемые в данный момент модели сетей. В работах [Shatravin, 2021, 2022] показано, как на ускорителях на основе ПВС могут быть реализованы нейронные сети прямого распространения, а также некоторые основные функции активации. Симуляции разработанных моделей демонстрируют высокое быстродействие. На данный момент эта теория активно исследуется, разрабатываются адаптации других распространённых нейросетевых алгоритмов. Для обсуждения реализации рекуррентных нейронных сетей необходимо предварительно рассмотреть основные аспекты предлагаемых моделей аппаратных ускорителей.

Лежащие в основе ускорителей перестраиваемые вычислительные среды представляют собой модель вычислительного устройства, состоящего из большого

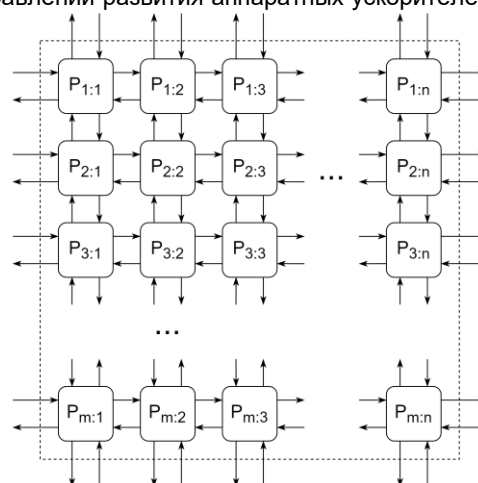


Рис. 5. Перестраиваемая вычислительная среда

количества простых вычислительных элементов (ВЭ). В рамках данной работы будут рассматриваться двумерные среды с квадратными элементами (рис. 5). Все ВЭ одинаковы, но каждый из них может быть независимо от других настроен на выполнение определённой операции из заранее заданного базиса операций. Для хранения текущей настройки (операции и дополнительных параметров) каждый ВЭ имеет небольшой объем внутренней памяти. Соседние элементы соединены друг с другом симметричными связями, что позволяет им обмениваться результатами вычислений и передавать сигнал через себя в другие участки среды.

Применение ПВС к построению аппаратных ускорителей обусловлено их многочисленными преимуществами. Благодаря тому, что каждый элемент среды функционирует независимо от других, среда может выполнять алгоритмы с высокой степенью параллельности, что играет большую роль в алгоритмах, поддерживающих распределённые вычисления и распараллеливание, к которым в том числе относятся и нейросетевые алгоритмы. Способность ПВС к динамической перенастройке позволяет не только реализовать в рамках одного вычислителя широкий класс алгоритмов, но и обеспечивает их изменение в процессе функционирования устройства. Изотропность среды допускает выполнение вычислений на любом её участке, что может быть полезно в случае частичного повреждения устройства, а также допускает одновременное выполнение нескольких различных алгоритмов на разных участках одной среды. Немаловажным преимуществом однородности вычислителей являются упрощение промышленного производства и масштабирования.

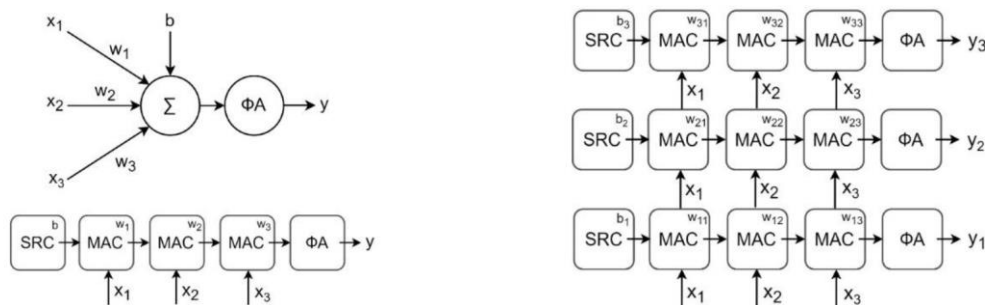


Рис. 6. Нейрон из элементов среды (слева) и полносвязный слой из трёх нейронов (справа)

В упомянутых ранее работах [Shatravin, 2021, 2022] предлагается реализация нейрона в виде цепочки из вычислительных элементов ПВС, а полносвязного слоя – как объединения таких цепочек (рис. 6). Этот подход обеспечивает высокую гибкость при реализации требуемых моделей нейронных сетей, так как изменяя длину цепочки элементов можно получить нейрон требуемой конфигурации. В то же время функции активации нейронов могут быть реализованы как в виде отдельной операции базиса элемента (как в случае с очень широко применяемой функцией линейного выпрямителя ReLU), так и в виде определённым образом настроенной группы элементов (сигмоида, гиперболический тангенс и т.д.). Пример реализации сигмоиды на предложенной модели ускорителя показан на рисунке 7 [Shatravin, 2022]. Алгоритм настройки такого ускорителя описан в работе [Шатравин, 2022].

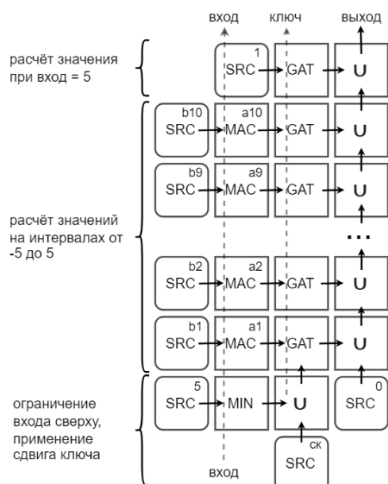


Рис. 7. Реализация сигмоидной функции активации на ПВС

произвольным количеством слоёв на среде ограниченного размера.

В данной работе предложен способ реализации рекуррентных сетей Хопфилда и LSTM на перестраиваемых ускорителях на основе вычислительных сред.

Для достижения лучших результатов может быть использован сегментированный режим функционирования среды. В этом режиме среда разбивается на несколько участков (сегментов), каждый из которых настраивается на реализацию одного из слоёв сети (рис. 8). Входной сигнал последовательно перемещается от сегмента к сегменту, проходя соответствующую обработку, а пройденные сегменты перенастраиваются на последующие слои. Такой подход позволяет уменьшить накладные расходы на перенастройку элементов, улучшить утилизацию ресурсов среды, устранить внешний обмен промежуточными результатами, а также реализовать сеть с про-

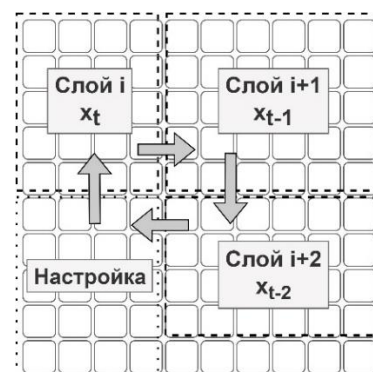


Рис. 8. Сегментированный режим функционирования среды

3. Реализация рекуррентной сети Хопфилда на ПВС

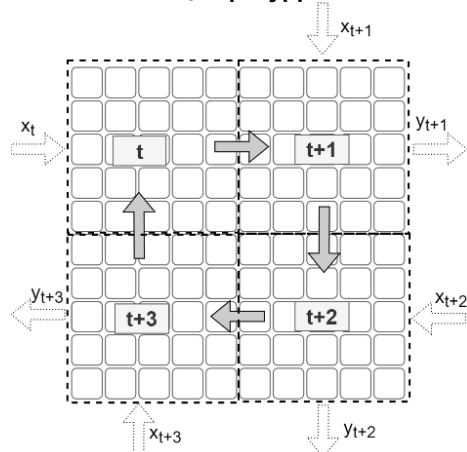


Рис. 9. Развёрнутое представление рекуррентной сети

В основе аппаратной реализации рекуррентных сетей на ПВС может использоваться их развёрнутое представление (рис. 9). Согласно такому представлению, рекуррентная сеть заменяется на глубокую нейронную сеть, каждый слой которой соответствует одному шагу работы рекуррентной сети. Все слои развёрнутого представления одинаковы, причём каждый эквивалентен всей рекуррентной сети. Тогда, с учётом описанного ранее сегментированного режима, рекуррентная сеть может быть реализована на среде согласно рисунку 10. Каждый сегмент среды реализует один и тот же слой, но с различным направлением передачи результата. Сигнал циклически перемещается между сегментами так долго, как требуется согласно выбранной архитектуре среды. Через внешние границы сегментов может подаваться новый входной сигнал и выводиться промежуточный результат. Так как все сегменты одинаковы, отсутствует необходимость в их перенастройке, что исключает накладные расходы и позволяет достичь максимальной скорости функционирования.

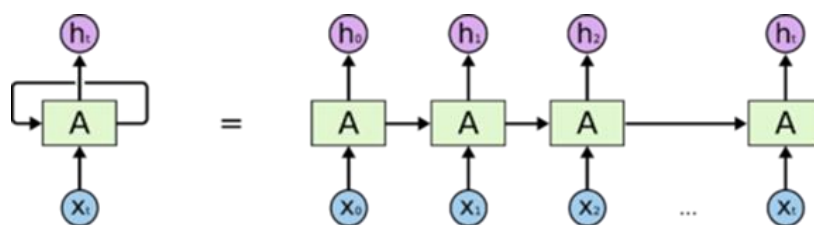


Рис. 10. Реализация развёрнутого представления рекуррентной сети на ПВС

Из рисунка 2 очевидно, что развёрнутое представление сети Хопфилда с n нейронами будет представлять собой глубокую неполносвязную сеть, каждый слой которой будет состоять из n нейронов. Сеть является неполносвязной, так как, согласно её архитектуре, каждый нейрон одного слоя не связан с соответствующим ему нейроном следующего слоя. Реализация сети Хопфилда для $n = 4$ на среде, разделённой на четыре сегмента, приведена на рисунке 11. Все сегменты среды реализуют одну и ту же конфигурацию и различаются лишь направлением передачи сигнала. Для наглядности один из сегментов выделен серым.

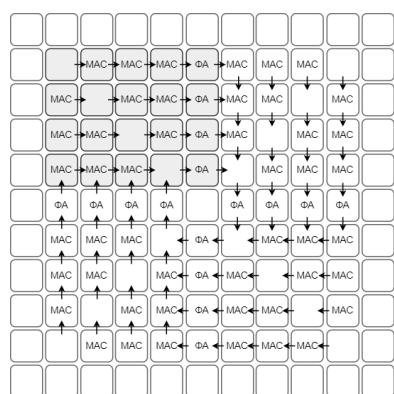


Рис. 11. Реализация сети Хопфилда с четырьмя нейронами, серым выделен один шаг развёрнутого представления сети.

Напомним, что в качестве функции активации сеть Хопфилда использует пороговую функцию (рис. 3). Несмотря на простоту, для её реализации на среде требуется решить важный вопрос – включить ли эту функцию в базис операций ВЭ, либо реализовать её на среде при помощи уже имеющихся операций. Включение функции в базис позволит затратить на её реализацию всего один ВЭ. Такое решение целесообразно, если ожидается её активное применение в реализуемых моделях. В остальных случаях более предпочтительной может быть реализация пороговой функции при помощи других операций базиса. Один из возможных вариантов реализации на пяти ВЭ представлен на рисунке 12. Он опирается на применение операции «затвор» (GAT), которая сравнивает поступающий на ключевой вход ВЭ сигнал с пороговым значением, и, при их совпадении, пропускает через себя основной сигнал (на картинке изображен слева), в ином случае на выходе элемента будет ноль. Пороговое значение затвора задано таким образом, что отпирание происходит при установленном в единицу значении знакового бита входного сигнала. Операции «источник сигнала» (SRC) формируют необходимые уровни, а «объединение» (U) и «сложение с накоплением» (MAC) осуществляют поворот и объединение сигналов.

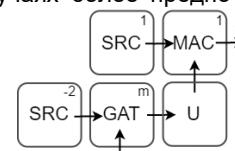


Рис. 12. Реализация пороговой функции активации на среде

4. Реализация на ПВС LSTM сети

Как было показано ранее, LSTM сеть имеет более сложное строение по сравнению с сетью Хопфилда. В связи с этим для реализации LSTM сетей нами рекомендуется разделение среды не на четыре, а на два сегмента, причём один из сегментов повернут относительно другого на 180° (рис. 13). Такое разбиение позволяет значительно уменьшить необходимый размер среды, что играет важную роль при реализации сетей с большим количеством нейронов и сложным состоянием ячейки (C_t).

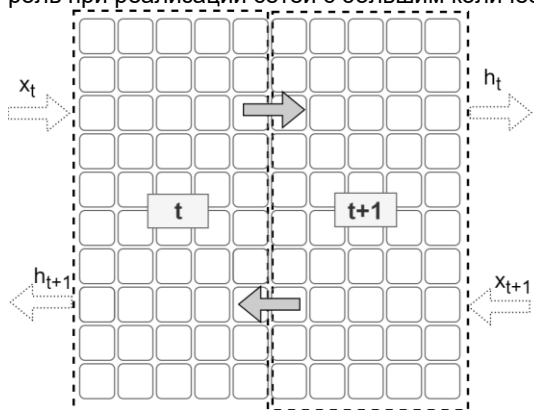


Рис. 13. Двухсегментное разбиение среды для реализации LSTM

Рассматриваемой архитектуры ПВС отсутствуют операции, использование которых позволило бы получить приемлемую реализацию поточечного умножения. Решением этой проблемы может быть внедрение в базис новой операции «умножение» (MUL), которая бы выполняла соответствующее преобразование над двумя поступающими на вход элемента сигналами. Такое решение является приемлемым, так как внутри каждого элемента уже реализуется операция умножения с накоплением (MAC), которая реализуют всю необходимую логику умножения. Таким образом, внедрение операции MUL отразится лишь на маршрутизации входных сигналов внутри элемента среды, не усложняя его вычислительный блок.

В итоге благодаря небольшой модификации базиса операций элемента ПВС становится возможной реализация LSTM сетей на описанной архитектуре ПВС. Реализация одного из двух сегментов такой среды схематично представлена на рисунке 14. Цвета блоков диаграммы соответствуют участкам сети с рисунка 4. Можно видеть, что выходной сигнал предыдущего шага (h_t) выходит из среды не на предыдущем, а на текущем шаге, одновременно используя для расчёта нового выходного значения (h_{t+1}). В то же время входной сигнал следующего шага (x_{t+1}) проходит текущий сегмент насквозь без каких-либо изменений. Такая конструкция объясняется ограничениями передачи сигнала внутри предложенной архитектуры среды. Обеспечение полной свободы перемещения сигнала внутри среды увеличило бы сложность её элементов, что приводит либо к увеличению размера требуемой среды, либо к уменьшению производительности при сохранении исходного размера. Тем не менее, такое решение не окажет значительного влияния на быстродействие модели в связи с низкими задержками передачи сигнала элементами среды (как будет показано далее, задержка передачи сигнала составляет менее 0.5 нс на один элемент среды).

5. Оценка быстродействия разработанных моделей.

Очевидно, что одной из ключевых характеристик аппаратных средств ускорения нейронных сетей является их быстродействие. В связи с тем, что на практике применяются сети разной конфигурации, целесообразно оценить быстродействие как функцию от параметров сети. В частности, от количества нейронов на разных слоях. Для этого необходимо определить зависимость от этих параметров отдельных компонентов ускорителя, оценить их быстродействие, и затем определить функциональную зависимость всей модели.

По рисункам 11 и 14 можно сделать вывод, что ключевые компоненты предлагаемых моделей это неполносвязный слой (блок операций MAC), функции активации (сигмоида, гиперболический тангенс), блоки поточечного сложения и умножения. Немаловажную роль играют также элементы, осуществляющие передачу сигнала между этими компонентами, так как передача сигнала внутри ПВС тоже занимает определённое время.

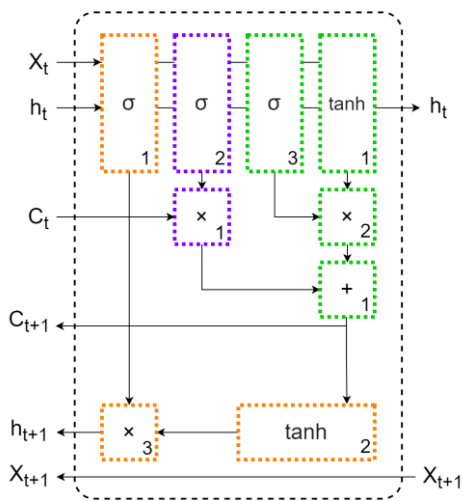


Рис. 14. Сегмент среды (правый), реализующий LSTM сеть

При выполнении оценки была проведена симуляция описанных компонентов на программируемых логических интегральных схемах (FPGA). FPGA являются широко применяемым инструментом как для исследования и прототипирования новых устройств, так и для внедрения в конечные устройства [Chang, 2016, D. Ghimire, 2022]. В рамках данного исследования была выбрана FPGA фирмы Intel Cyclone V (5CGFXC9E7F35C8), а симуляция проводилась в среде Quartus Prime (version 20.1.0, build 711 SJ Edition) при помощи встроенного пакета временных симуляций Timing Analyzer Tool. Были оценены время передачи сигнала по цепочке элементов среды разной длины, время расчёта операции умножения с накоплением (MAC) цепочкой элементов и время расчёта сигмоидной функций активации. Так как гиперболический тангенс может быть рассчитан из сигмоиды при помощи дополнительных операций умножения и сложения, его быстродействие может быть рассчитано из уже известных данных. Подробно процесс и результаты измерений описаны в [Shatravин, 2021, 2022].

Проведённые симуляции показали, что на реализацию одного элемента среды требуется 296 логических ячеек FPGA. Средняя длительность передачи сигнала через элемент менее 0.5 нс. Выполнение операции умножения с накоплением требует около 1.1 нс, а полная реализация сигмоиды 18.5 нс. Таким образом, расчёт гиперболического тангенса требует около 21.8 нс., а пороговая функция активации 2.7 нс. Данные результаты получены для элементов, работающих с 16-разрядными двоичными числами с фиксированной точкой (по 8 разрядов под целую и дробную части).

Рассмотрим модель среды, реализующую нейронную сеть Хопфилда (рис. 11). Она состоит из четырёх сегментов, каждый из которых представляет собой неполносвязный слой из n нейронов и пороговой активации. Тогда длительность наибольшего пути распространения сигнала в таком сегменте составляет:

$$t_{H1}(n) = 0.5n + 1.1(n - 1) + 0.5 + 2.7. \quad 1)$$

Так как все сегменты одинаковы, то на один полный круг в среде сигналу требуется ровно в четыре раза больше времени. Если принять $n = 25$ нейронов, то $t_{H1}(25) = 42.1$ нс, а $t_{H4}(25) = 4t_{H1}(25) = 168.4$ нс. При этом потребуются среда квадратной формы с длиной стороны $2n + 1 = 51$ вычислительных элементов, то есть всего 2601 вычислительный элемент.

Рассмотрим LSTM сеть. Она имеет значительно более сложную структуру и включает в себя несколько разных полносвязных слоёв (рис. 14). Для LSTM сети можно выделить несколько ключевых параметров: размер входного сигнала (l_x), размер внутреннего состояния (l_c) и размер выходного сигнала (l_h). Четыре верхних полносвязных слоя ($\sigma_1, \sigma_2, \sigma_3, \tanh_1$) принимают на вход одновременно входной сигнал и выходной сигнал с предыдущего шага, то есть длина их входного вектора составляет $l_x + l_h$. При этом их выходы имеют разный размер. Выход σ_1 имеет длину l_h , выходы $\sigma_2, \sigma_3, \tanh_1$ равняются l_c . Расположенный в нижнем правом углу \tanh_2 имеет вход длины l_c , а выход l_h . Из этих данных очевидным образом следуют размеры блоков поточечного сложения и умножения. Тогда можно видеть, что наибольшая длительность вычисления результата на каждом компоненте выражается:

$$t_{\sigma_1} = 0.5l_h + 1.1(l_h + l_x) + 18.5, \quad 2)$$

$$t_{\sigma_2} = t_{\sigma_3} = 0.5l_c + 1.1(l_h + l_x) + 18.5, \quad 3)$$

$$t_{\tanh_1} = 0.5l_c + 1.1(l_h + l_x) + 21.8, \quad 4)$$

$$t_{\tanh_2} = 0.5l_h + 1.1l_c + 21.8, \quad 5)$$

$$t_{mul_1} = t_{mul_2} = t_{add_1} = 0.5 * 2(l_c - 1) + 1.1, \quad 6)$$

$$t_{mul_3} = 0.5 * 2(l_h - 1) + 1.1. \quad 7)$$

В то же время надо иметь в виду, что для ПВС характерна высокая параллельность и независимое функционирование отдельных элементов в составе среды. Это значит, что формирование результата в компоненте σ_2 не ожидает завершения расчета σ_1 , также как \tanh_1 не ожидает завершения $\sigma_1, \sigma_2, \sigma_3$. Они все функционируют одновременно (с учётом задержки поступления сигналов на их входы). Из этого следует, что полная длительность расчёта результата на одном сегменте LSTM среды будет меньше суммы длительностей на всех её компонентах. Для определения фактической длительности получения результата необходимо найти и измерить критический, то есть наиболее долгий путь перемещения сигнала. Из предложенной модели среды и известных размеров её компонентов можно сделать вывод, что критический путь будет проходить слои с σ_1 по σ_3 насквозь (передача сигнала, $t_{in \rightarrow \tanh}$), затем будет осуществляться расчёт компонентов $\tanh_1 \rightarrow mul_2 \rightarrow add_1 \rightarrow \tanh_2 \rightarrow mul_3$ с учётом задержек на конструктивно расположенных между ними элементах, осуществляющих передачу сигнала ($t_{add \rightarrow \tanh_2}, t_{\tanh_2 \rightarrow mul_3}$). Тогда:

$$t_{LSTM1} = t_{in \rightarrow \tanh_1} + t_{\tanh_1} + t_{mul_2} + t_{add_1} + t_{add \rightarrow \tanh_2} + t_{\tanh_2} + t_{\tanh_2 \rightarrow mul_3} + t_{mul_3}, \quad 8)$$

$$t_{in \rightarrow \tanh_1} = 0.5(l_h + 2l_c), \quad 9)$$

$$t_{add \rightarrow \tanh_2} = 0.5(r_{\tanh} + l_c), \quad 10)$$

$$t_{\tanh_2 \rightarrow mul_3} = 0.5(2l_c - r_{\tanh}), \quad 11)$$

где r_{\tanh} – количество столбцов элементов среды, необходимых для реализации гиперболического тангенса. Примем равным 15.

Таким образом, если задать $l_x = 10$, $l_c = 25$, $l_h = 5$, то в соответствии с (8) полное время расчёта одного сегмента будет составлять:

$$t_{LSTM1} = 27.5 + 50.8 + 25.1 + 25.1 + 20 + 51.8 + 17.5 + 5.1 = 222.9 \text{ нс.} \quad (12)$$

Так как оба сегмента одинаковы, то полный круг внутри среды, соответствующий двум последовательным LSTM слоям, будет проходиться сигналом за 445.8 нс. При этом потребуется среда из $105 \times 90 = 9450$ вычислительных элементов.

Анализ альтернативных реализаций LSTM на FPGA показывает соизмеримость или преимущество предложенных моделей в отношении производительности. Исследование [He, 2021] описывает реализацию ускорителя для LSTM сети с 512 нейронами на скрытом слое (размер состояния ячейки C_t) с задержкой на обработку одного сигнала в 47.8 мкс, что при кратном масштабировании в 10.5 раз превышает полученную в данной работе оценку ускорителя на основе ПВС. [Сао, 2019] при помощи нескольких техник оптимизации (квантование данных, устранение некоторых весов и других) и параллельному расчёту поточечных операций для LSTM сети с 200 нейронами в скрытом слое удалось добиться времени выполнения 1.3 мкс, что несколько выигрывает у рассчитанных в данной работе оценок ценой уменьшения точности. Быстродействие предложенного в исследовании [Ferreira, 2016] ускорителя составляет 1.14 мкс для сети с 64 скрытыми слоями, что в 2 раза уступает предложенной нами реализации. Работа [Chang, 2016] описывает аппаратный ускоритель, требующий 932 мс на расчёт одной итерации сети со 128 скрытыми нейронами, что превышает результаты данной работы почти на 6 порядков. Следует заметить, что все рассмотренные аналоги имеют жёсткую неперестраиваемую структуру, что делает преимущества моделей на основе ПВС ещё более весомыми.

Ключевым недостатком моделей на основе ПВС является большая занимаемая площадь полупроводника. Это объясняется как накладными расходами на обеспечение перестраиваемости среды (каждый элемент среды поддерживает полный базис операций, хотя в любой момент времени выполняет лишь одну), так и высокой степенью параллельности и пространственного распределения алгоритмов. Приведённое ранее значение (296 логических ячеек FPGA на реализацию одного элемента среды) позволяет оценить количество требуемых ячеек для среды заданного размера. Уменьшение этой величины позволит увеличить производительность среды благодаря более плотному размещению её элементов на полупроводнике. Это одно из основных направлений развития исследований аппаратных ускорителей на ПВС.

Заключение

Сегодня алгоритмы машинного обучения играют важную роль во многих информационных и технических системах. Особый интерес представляют глубокие свёрточные и рекуррентные сети. Однако свойственная им высокая вычислительная сложность ограничивает их применение в широком классе маломощных мобильных и автономных систем. Одним из путей решения этой проблемы является применение аппаратных ускорителей нейронных сетей на основе перестраиваемых вычислительных сред. Такие среды обеспечивают не только высокое быстродействие благодаря распараллеливанию и пространственному распределению алгоритмов, но и высокую гибкость, что позволяет изменять выполняемые устройством алгоритмы во время его функционирования. Эта способность к динамической перенастройке отдельных участков открывает широкие возможности для разработки глубоких нейронных сетей с большим количеством слоёв.

В данной работе показано, как представление рекуррентных нейронных сетей в виде глубоких сетей может быть использовано при реализации перестраиваемого аппаратного ускорителя на ПВС. Предложены алгоритмы реализации классической рекуррентной сети Хопфилда и широко используемой сети с долгой краткосрочной памятью (LSTM). Приведены формулы оценки быстродействия LSTM сети в зависимости от её конфигурации. В частности, длительность вычисления одного шага сети с 10 входными сигналами, 25 скрытыми нейронами и 5 выходными сигналами составляет 223 нс., что значительно меньше, чем у большинства рассмотренных аналогов, к тому же лишённых способности к перенастройке. Слабым местом предложенных ускорителей на основе ПВС является большая занимаемая площадь на кристалле полупроводника. Это объясняется накладными расходами на обеспечение перестраиваемости, пространственным распределением алгоритма, высокой степенью параллельности. Так как плотность размещения элементов среды напрямую влияет на её производительность, вопрос уменьшения размера отдельного элемента будет исследоваться в рамках дальнейших работ. Несмотря на эту особенность, аппаратные ускорители на основе ПВС демонстрируют высокие показатели по ключевым параметрам, что подтверждает перспективность их дальнейшего исследования.

Благодарности:

Исследование выполнено за счет гранта Российского научного фонда № 21-71-00012, <https://rscf.ru/project/21-71-00012/>

Литература

1. Шатравин В., Шашев Д. В. Разработка алгоритма настройки перестраиваемой вычислительной среды в составе аппаратного ускорителя искусственных нейронных сетей / В. Шатравин, Д. В. Шашев // Цифровая экономика. – 2022. – 20(4). – с. 11–18.

2. Cao S. Efficient and Effective Sparse LSTM on FPGA with Bank-Balanced Sparsity / S. Cao [et al.] // Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. – 2019. – pp. 63-72.
3. Chang A. X. M. Recurrent Neural Networks Hardware Implementation on FPGA / A. X. M. Chang, B. Martini, E. Culurciello // arXiv:1511.05552v4. – 2016. – pp. 1-7.
4. Chen Y. H. Eyeriss v2: A flexible accelerator for emerging deep neural networks on mobile devices / Y. H. Chen [et al.] // IEEE Emerging and Selected Topics in Circuits and Systems (Jetcas). – 2019. – 9. – pp. 292-308.
5. Ferreira J. C., Fonseca J. An FPGA Implementation of a Long Short-Term Memory Neural Network / J. C. Ferreira, J. Fonseca // International Conference on ReConfigurable Computing and FPGAs (ReConFig). – 2016. – pp. 1-8.
6. Ghimire D., Kil, D., Kim, S.-h. A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration // J. Electronics, MDPI. – 2022, 11. – vol. 945. – pp. 1-23.
7. He D. An FPGA-Based LSTM Acceleration Engine for Deep Learning Frameworks / D. He [et al.] // Electronics. – 2021. – 10(6). – pp. 1-15.
8. Hochreiter S., Schmidhuber J. Long Short-term Memory / S. Hochreiter, J. Schmidhuber // Neural computation. – 1997. – 9(8).
9. Shatravин V., Shashev D., Shidlovskiy S. Sigmoid Activation Implementation for Neural Networks Hardware Accelerators Based on Reconfigurable Computing Environments for Low-Power Intelligent Systems // MDPI: Applied Sciences. – 2022. – 12(10).
10. Shatravин V., Shashev D. V., Shidlovskiy S.V. Applying the Reconfigurable Computing Environment Concept to the Deep Neural Network Accelerators Development // International Conference on Information Technology (ICIT) – 2021. – pp. 842-845.
11. Understanding LSTM Networks. Colah's blog [Электронный ресурс]. – URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (дата обращения 12.10.2022).

References in Cyrillics

1. Shatravин V., Shashev D. V. Development of configuration algorithm for reconfigurable computing environment in neural networks hardware accelerators / V. Shatravин, D. V. Shashev // Digital Economy. – 2022. – 20(4). – pp. 11-18.

Шатравин Владислав, аспирант, млад. науч. сотр. ТГУ, Томск

(shatravин@stud.tsu.ru)

Шашев Дмитрий Вадимович, к.т.н, доцент ТГУ, Томск

(dshashev@mail.tsu.ru)

Ключевые слова

Рекуррентные нейронные сети, LSTM, перестраиваемые вычислительные среды, реконфигурируемые аппаратные ускорители.

Vladislav Shatravин, Acceleration of recurrent neural networks with computing environments

Keywords

Recurrent neural networks, LSTM, reconfigurable computing environments, reconfigurable hardware accelerators.

DOI: 10.34706/DE-2023-01-04

JEL: C63 – Вычислительные методы, моделирование.

Abstract

The application of modern machine learning algorithms in technical systems is limited by the hardware they use. The problem is particularly serious when using large neural networks in low-power and autonomous systems that have severe weight and power consumption restrictions. Majority of modern neural network hardware accelerators either have high both power consumption and weight or they are highly specialized for a small set of algorithms. One possible solution to the problem is the use of dynamically reconfigurable hardware accelerators that can change the implemented algorithms at run time. The accelerators can be based on the principles of reconfigurable computing environments (RCE). This paper presents implementations of Hopfield and long short-term memory (LSTM) recurrent networks on RCE-based accelerators. The performance evaluations of the developed models were determined through simulations on FPGA. Estimates show the high performance of the presented models in comparison with analogues, however, the requirements for the area on a chip are also higher. According to estimates, an LSTM network with 25 hidden neurons will be calculated in 223 ns. The results obtained allow to conclude that there is a high potential for using RCE-based accelerators for recurrent networks and the need for further optimization.

1.5. ИССЛЕДОВАНИЕ СТРУКТУРНЫХ ИЗМЕНЕНИЙ В СЕГМЕНТЕ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ СУПЕРКОМПЬЮТЕРОВ

Варшавский Л. Е., ЦЭМИ РАН, г. Москва, Россия

В статье исследуется динамика структуры парка высокопроизводительных суперкомпьютеров из списка Top-500 по производительности и по используемой тактовой частоте процессоров. Рассматриваются сценарии роста производительности суперкомпьютеров США и Китая, построенные на основе разработанных эконометрических моделей.

1. Введение.

В настоящее время высокопроизводительные вычисления превратились в важную отрасль, обслуживающую все основные направления науки и экономики. Практически все крупные компании используют суперкомпьютеры как для повышения своего научно-производственного потенциала, так и для повышения экономической эффективности принимаемых решений. Суперкомпьютеры используются и в режиме оперативного управления производственными процессами (в частности, при создании и эксплуатации «цифровых близнецов»), при управлении мультиагентными системами в экономике и технике в режиме реального времени, а также при планировании [Макаров и др., 2022].

Суперкомпьютеры являются важным средством укрепления военного потенциала и национальной безопасности, способствующим повышению эффективности различных видов вооружений (ядерного, ракетно-космического и др.), а также систем связи и шифрования.

Всё это обуславливает актуальность исследования перспектив развития суперкомпьютерной техники, а также стратегий её развития в ведущих странах. В статье проводится анализ динамики структурных изменений в парке высокопроизводительных суперкомпьютеров из списка Top-500 по таким показателям, как производительность и тактовая частота процессоров. Исследуются сценарии роста производительности суперкомпьютеров США и Китая, построенные на основе разработанных эконометрических моделей.

2. Моделирование структуры парка высокопроизводительных суперкомпьютеров по производительности

Переход к новому поколению высокопроизводительной техники во многих случаях сопровождается стабилизацией структуры различных типов оборудования по производительности. Так, в частности,

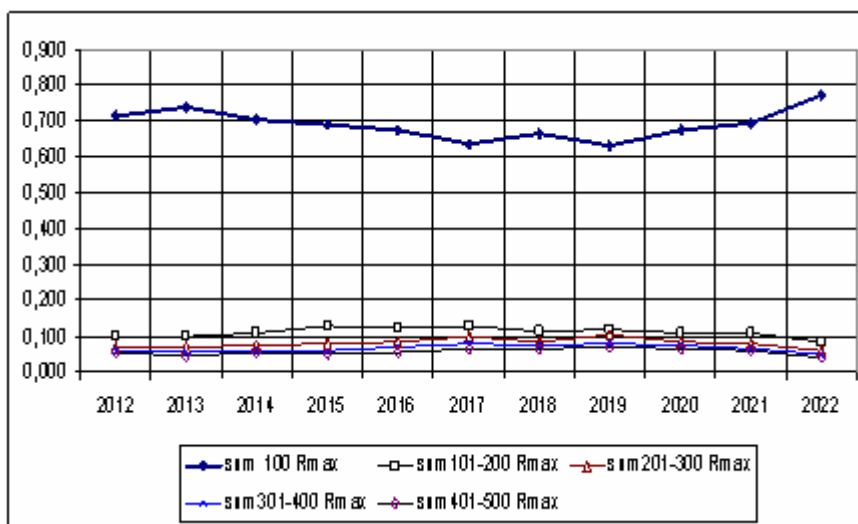


Рис. 1. Динамика доли 5 групп суперкомпьютеров в списке Top-500, где первая группа охватывает наиболее мощные суперкомпьютеры (1–100), вторая—менее мощные (101–200), остальные три— суперкомпьютеры с порядковыми номерами, соответственно 201–300, 301–400, 401–500. (построено по данным <https://top500.org>).

происходило в энергетике (в традиционной и атомной) и в авиации. Так происходит и в отрасли высокопроизводительных вычислений. По мере освоения петафлопсной производительности, за период с 2013 по 2019 г. происходило некоторое уменьшение доли группы-100 наиболее производительных суперкомпьютеров в списке 500 наиболее производительных суперкомпьютеров (Top-500, <https://top500.org>). Однако с 2020 г., в связи с резким ростом производительности суперкомпьютеров-чемпионов и переходом к эксафлопсной

эре, началось увеличение этой доли (с 63% в 2019 г. до 77% в 2022 г., рис.1). Похожая динамика имела место в соотношении между суммарными производительностями группы¹ из первых 100 суперкомпьютеров и группы из последних 100 наименее мощных устройств (рис. 2).

Для оценки выравнивания и однородности структуры количественных показателей суперкомпьютеров может быть использован показатель энтропии:

$$H_t = - \sum_{i=1}^5 s_i \text{LN}(s_i), \tag{1}$$

где s_i — доля i -ой группы суперкомпьютеров в исследуемом суммарном показателе.

В настоящей работе для оценки изменения структуры производительности суперкомпьютеров R_{\max} по данным июньских выпусков <https://top500.org> построена агрегированная модель, основанная на использовании связи между темпом роста производительности суперкомпьютеров-чемпионов и непосредственно энтропией:

$$H_t = \frac{\alpha_0 H_{t-1}}{1 + \alpha_1 H_{t-1}} + \gamma \text{LN}(R_{\max t} / R_{\max t-1}). \tag{2}$$

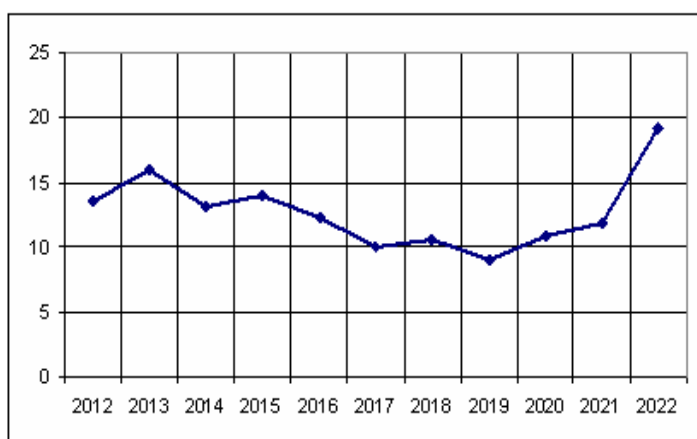


Рис. 2. Динамика соотношения между суммарными производительностями группы из первых 100 суперкомпьютеров и группы из последних 100 наименее мощных устройств в списке Top 500 (построено по данным (<https://top500.org>))

Оценки параметров модели, полученные с использованием нелинейного метода наименьших квадратов (МНК), составляют: $a_0=4.297$; $a_1=2.705$; $\gamma=-0.291$, а коэффициент вариации — 0.062. Равновесное значение энтропии при нулевом входе составляет $H_\infty = 1.219$, что соответствует доле 1-ой группы, равной 0.527. Положение равновесия является устойчивым, что свидетельствует о возможной стабилизации структуры за счёт приближения производительности менее мощных суперкомпьютеров к производительности суперкомпьютеров первой группы при исчерпании повышения мощности суперкомпьютеров-чемпионов в отдаленной перспективе.

Расчеты на перспективу при сценарии роста производительности суперкомпьютера-рекордсмена до 10 и 100 эксафлопс (10^{18} флопс) в 2025 и 2030 гг., соответственно (такие уровни являются целевыми для разработчиков национальных лабораторий США²), показывают, что в связи с резким увеличением производительности суперкомпьютеров-чемпионов в начальный период (со среднегодовым темпом роста 2.09) происходит снижение энтропии. Затем, по мере уменьшения темпов роста производительности суперкомпьютеров-чемпионов (до 1.58), энтропия увеличивается (рис. 3). Иными словами, в период повышенного темпа роста производительности суперкомпьютера-рекордсмена происходит повышение концентрации вычислительных мощностей в первой группе, а в дальнейшем, по мере снижения темпа роста — некоторое снижение.

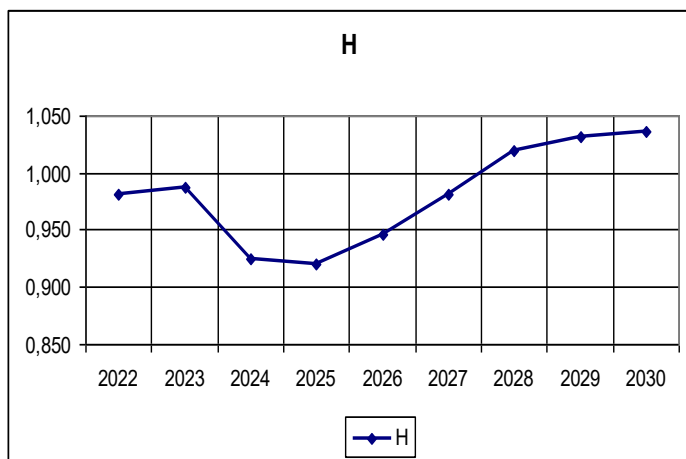


Рис. 3. Прогнозная динамика энтропии (1), рассчитанная на основе модели (2)

¹ В соответствии с индексом Linpack — R_{\max} , характеризующим скорость решения линейных уравнений большой размерности.

² <https://www.nextplatform.com/2022/06/30/so-you-think-you-can-design-a-20-exaflops-supercomputer/>

3. Исследование структуры парка высокопроизводительных суперкомпьютеров ведущих стран по тактовой частоте процессоров

Тактовая частота процессоров, используемых в суперкомпьютерах, является одним из важных факторов, определяющих быстродействие этих устройств. До начала 2000-х гг., пока возможно было обеспечивать пропорциональное уменьшение топологических размеров микросхем и тактовой частоты (в этом состоит так называемое правило Деннарда), инженерам удавалось добиваться роста производительности устройств, главным образом, за счет их миниатюризации и увеличения частоты [Dennard, 1974]. Однако уже в середине первого десятилетия XXI века, из-за обострения проблемы отвода тепла и возникновения в чипах токов утечки (как из-за работы на высоких частотах, так и за счет проявления квантовых эффектов, в частности, туннельного эффекта [Варшавский 2013] [Варшавский 2022b]), рост производительности микросхем и компьютеров в целом стал связываться с увеличением числа ядер в процессорах.

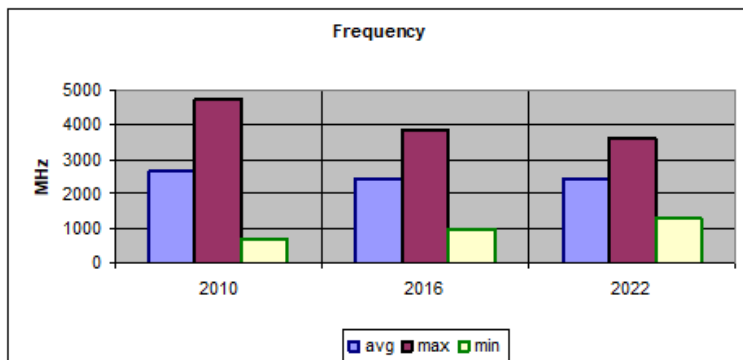


Рис. 4 Динамика средней (avg), максимальной (max) и минимальной (min) тактовой частоты в МГц, используемой в суперкомпьютерах из списка Top-500 (построено по данным <http://top500.org>)

Показательной является судьба микропроцессора Tejas, разработанного компанией Intel. Тактовую частоту процессора предполагалось повысить к 2010 г. до 10–15 ГГц. Однако из-за повышенного тепловыделения Intel в 2005 г. объявил об отказе от производства микропроцессора Tejas и в 2006 г. перешел к выпуску двухядерного процессора Core³, а впоследствии — многоядерных процессоров [3, 4] (впоследствии было установлено, что имеет место степенная зависимость потребляемой процессором мощности от его тактовой частоты [Митропольский, 2015], [Варшавский 2022b]).

Переход к многоядерности, обеспечивающий параллелизм и ускорение вычислений, существенно сдерживает увеличение частоты, ведущей к росту потребления энергии и увеличению тепловыделения⁴.

В связи с вышеуказанными проблемами, средняя тактовая частота процессоров, используемых в высокопроизводительных суперкомпьютерах, за последние годы стабилизировалась на уровне 2427 МГц.

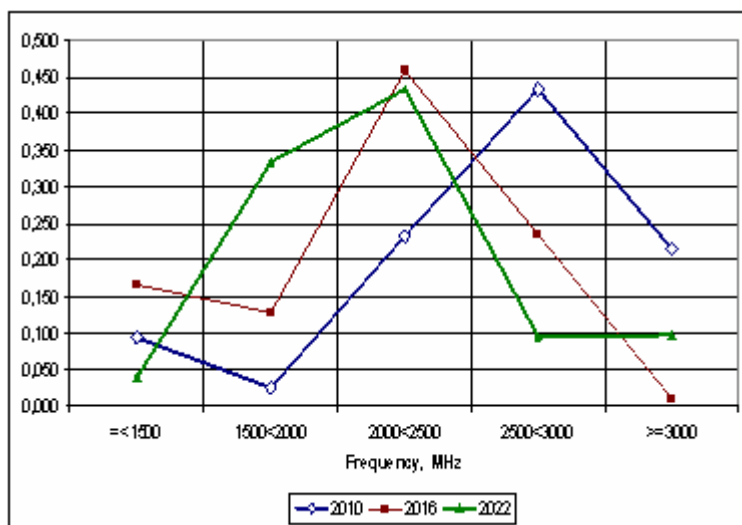


Рис. 5 Удельный вес суперкомпьютеров с различными диапазонами тактовой частоты процессоров в МГц в общем списке top500 (построено по данным <http://top500.org>)

В то же время, за период с 2010 по 2022 г. максимальная используемая частота снизилась с 4700 МГц до 3600 МГц, а минимальная частота — повысилась с 700 МГц до 1300 МГц (рис. 4). Интересно отметить, что в 2005 минимальная используемая частота в суперкомпьютерах из списка Top-500 составляла 333 МГц. Если в 2010 г. максимальная доля суммарной производительности Rmax суперкомпьютеров из списка Top-500 приходилась на диапазон частот 2500–3000 МГц, то в 2016 и 2022 г. — на диапазон частот 2000–2500 МГц. При этом доля суммарной производительности суперкомпьютеров, использующих диапазон частот 1500–2500 МГц, повысилась с 0.258 в 2010 г. до 0.767 в 2022 г. (рис. 5).

Вместе с тем, в ряде наиболее мощных суперкомпьютеров используются тактовые частоты как до 1500 МГц (китайский Sunway с частотой 1450 МГц), так и свыше 3000 МГц (американские Summit и Sierra, с частотами соответственно 3070 и 3100 МГц). Естественно, что одним из главных средств достижения высокой производительности суперкомпьютеров при использовании низких тактовых частот является увеличение числа процес-

³ URL: <http://pe4.info/1-7.php> (Дата обращения -1.03.2002).

⁴ URL: <https://www.nextplatform.com/2023/01/19/more-cpu-cores-isnt-always-better-especially-in-hpc/> (Дата обращения 21.01.2023).

сорных ядер. Так, в суперкомпьютере Sunway производительностью 93 Петафлопс свыше 10.8 млн. ядер, в то время, как в суперкомпьютерах Summit и Sierra производительностью, соответственно, 148.6 и 94.6 Петафлопс — 1.57 и 1.41 млн. ядер (эти суперкомпьютеры входят в первую десятку из списка Top-500). Следует отметить, что в наиболее производительных американских суперкомпьютерах большая часть ядер (свыше 90%) приходится на ускорители (в основном на графические процессоры GPU), характеризующиеся повышенным потреблением энергии по сравнению с центральными процессорами (CPU). Во многом с этим связано то, что суперкомпьютеры Summit и Sierra, использующие повышенные тактовые частоты, характеризуются не самой высокой энергоэффективностью (по данным <http://top500.org> за май 2022 г., соответственно, 14.72 и 12.72 гигафлопс/Вт).

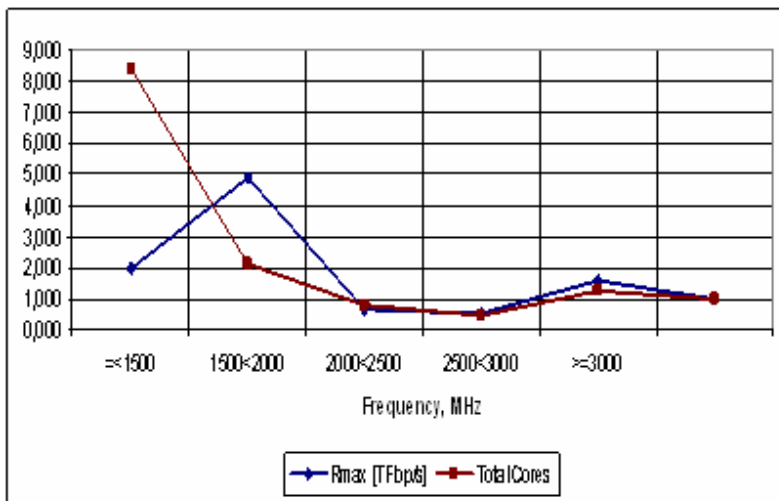


Рис. 6. Эластичности суммарных производительности (R_{max}) и числа ядер (Total Cores) суперкомпьютеров в списке top500 при различных диапазонах тактовой частоты процессоров в МГц (построено по данным <http://top500.org>)

листов, повышение быстродействия устройств оперативной памяти, в частности, памяти с высокой пропускной способностью (high bandwidth memory — HBM) ⁴.

4. Исследование динамики производительности высокопроизводительных суперкомпьютеров ведущих стран (на примере США и Китая)

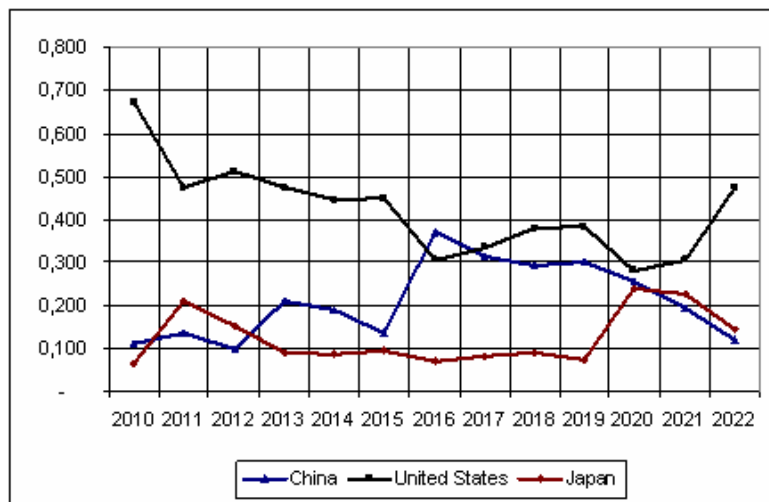


Рис. 7 Динамика доли суперкомпьютеров 3-х стран в общей производительности 500 самых высокопроизводительных суперкомпьютеров (без учета 2-х эксафлопсных суперкомпьютеров Китая, введенных в 2021 г., и не включенных в список Top-500; построено по данным <http://top500.org>).

наиболее производительных суперкомпьютеров устойчиво повышалась с 29% до почти 54%, а Китая — устойчиво снижалась с 36% до 4,5%. В 2022 г.в эту группу входило 33 американских суперкомпьютеров, в то время, как китайских — всего 2.

На рис.6 по данным майского выпуска <http://top500.org> за 2022 г. приведены зависимости эластичностей суммарных производительности и числа ядер суперкомпьютеров, использующих разные частоты (т.е. отношение долей суммарных производительности и числа ядер к долям числа суперкомпьютеров в списке Top-500, использующих разные частотные диапазоны).

В связи с исчерпанием возможности повышения тактовой частоты процессоров и высокой энергоемкостью процессоров-ускорителей, в настоящее время одним из способов повышения производительности и энергоэффективности суперкомпьютеров считается, по мнению специалистов, в частности, памяти с высокой пропускной способностью (high bandwidth memory — HBM) ⁴.

В рейтингах высокопроизводительных суперкомпьютеров Top-500 большая часть суммарной производительности этого оборудования (примерно 3/4), измеряемой индексом Linpack (R_{max}), традиционно приходится на 3 страны: США, Китай и Японию (рис. 7). В последнее десятилетие началась острая конкуренция за лидерство в исследуемой области между указанными странами и, прежде всего, между США и Китаем.

Суперкомпьютеры этих стран попеременно занимают первое место по производительности в списке Top-500. Интересно отметить, что за период с 2016 по 2022 гг. доля США в суммарной производительности группы из 100

Особый интерес представляет развитие суперкомпьютерных технологий в Китае, успешно осуществляющего импортозамещение в рамках своей программы. Первые суперкомпьютеры производились на основе импортных компонентов. В начале 2010-х гг. китайский суперкомпьютер производительностью более 1 Петафлопс стал наиболее мощным в списке Top-500. В последующие годы ведущее место в указанных рейтингах попеременно занимали китайские суперкомпьютеры с процессорами собственной архитектуры и производства, а также американские и японские. В значительной степени китайским инженерам удалось добиться импортозамещения основных компонентов. Особенность последних моделей китайских суперкомпьютеров-чемпионов состоит в использовании десятков тысяч процессоров, включающих в себя сотни ядер, с относительно небольшой тактовой частотой. В одном из последних суперкомпьютеров эксафлопсной производительности (Sunway Oceanlite), который Китай, стремясь избежать дополнительных санкций США, не включил в список Top-500, используются процессоры с 390 ядрами (общее число ядер в системе составляет порядка 40 млн.), работающими с тактовой частотой 1.45 ГГц. Использование сравнительно низкой тактовой частоты позволяет несколько компенсировать повышенный расход энергии и выделение тепла, связанных с резким увеличением числа ядер 5, [Варшавский, а].

Следует отметить, что в настоящее время процессоры с такой частотой используются в относительно небольшом числе суперкомпьютеров (в 2022 г. в 6% из списка Top-500).

Большая часть высокопроизводительных вычислительных систем использует процессоры с частотами в диапазоне 2–3 ГГц (см. п. 3). Переход к архитектуре с повышенным числом ядер в наиболее мощных китайских суперкомпьютерах связан, в частности, со стремлением избежать санкций США на поставку процессоров повышенной производительности (включая процессоры-ускорители, в т.ч. на основе графических процессоров), а также оборудования для производства микросхем с малым топологическим размером (14 нм и меньше).

В настоящей работе на основе данных из <http://top500.org>, а также с учетом ввода в 2021 г. двух китайских эксафлопсных суперкомпьютеров, не вошедших в Top-500, разработано несколько вариантов модели гонки производительности высокопроизводительных суперкомпьютеров США и Китая.

Так, один из вариантов представляет собой модель векторной авторегрессии VAR вида:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \begin{pmatrix} \xi_{1t} \\ \xi_{2,t-1} \end{pmatrix}, \quad (3)$$

где $X = \begin{pmatrix} x_{it} \\ x_{it-1} \end{pmatrix}$ — векторы, составленные из натуральных логарифмов суммарных производительностей суперкомпьютеров США ($i=1$) и Китая ($i=2$), A, B — матрицы и векторы соответствующей размерности ($i=1,2$), ξ_{it} — натуральный логарифм производительности суперкомпьютера-чемпиона в году t ,

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 2\lambda_1 & -\lambda_1^2 & 0 & a_{21} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ a_{12} & 0 & 2\lambda_2 & -\lambda_2^2 \end{pmatrix}, \quad B_i = \begin{pmatrix} 0 & 0 \\ b_{i1} & b_{i2} \end{pmatrix}; \quad (4)$$

На основе нелинейного МНК получены следующие оценки параметров векторной авторегрессии (3)–(4):

$$\lambda_1=0.882; \quad a_{21}=0.000; \quad \lambda_2=0.847; \quad a_{12}=0.0455; \quad b_{11}=0.417; \quad b_{12}=0.423; \quad b_{21}=b_{22}=0.000.$$

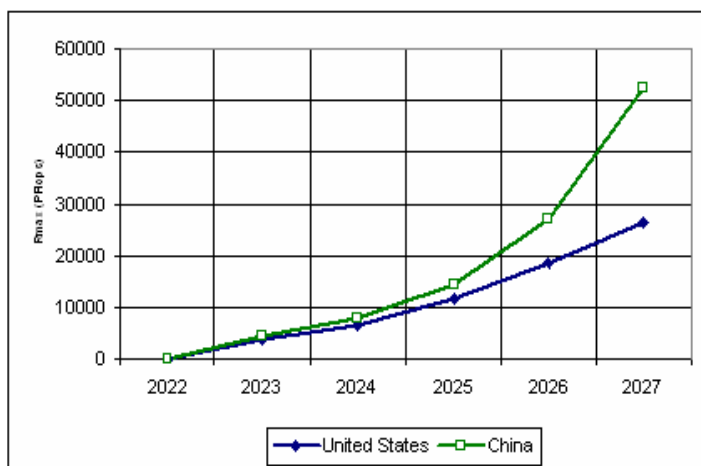


Рис. 8 Динамика суммарной производительности суперкомпьютеров США и Китая, рассчитанная на основе модели (3)–(4).

Структура матриц построенной модели (3)–(4) свидетельствует о том, что можно попытаться получить необходимые динамические соотношения путём раздельного оценивания динамики взаимосвя-

Коэффициенты вариации для зависимостей с переменными x_{it} составили 0.044 (при $i=1$) и 0.071 (при $i=2$). На рис. 8 приведен прогнозный сценарий динамики суммарной производительности суперкомпьютеров на 5-летний период при условии роста производительности суперкомпьютеров-чемпионов до 10 эксафлопс в 2025 г. и далее до 100 эксафлопс в 2030 г. Расчеты по этой модели показывают, что уже в 2025 г. суммарная производительность суперкомпьютеров Китая может составлять 14.4 эксафлопс, а США — 11.8 эксафлопс. В 2027 г. суммарная производительность китайских суперкомпьютеров будет в 2 раза выше американских.

⁵ <https://www.tomshardware.com/news/two-chinese-exascale-supercomputers> (Дата обращения 15.07.2022).

зей для США и Китая. Другая построенная нами модель состоит из двух распределенных запаздываний, связывающих: 1). натуральные логарифмы производительности суперкомпьютера-чемпиона ξ_{1t} (входная переменная) и суммарной производительности суперкомпьютеров США x_{1t} ; 2). суммарные производительности суперкомпьютеров Китая x_{2t} и США x_{1t} (входная переменная). Эти распределенные запаздывания имеют следующий вид:

$$x_{1t} = 2\lambda_{11} x_{1t-1} - \lambda_{11}^2 x_{1t-2} + b_{11} \xi_{1t} + b_{12} \xi_{1t-1} \quad (5)$$

$$x_{2t} = 2\lambda_{22} x_{2t-1} - \lambda_{22}^2 x_{2t-2} + a_{21} x_{1t-1}. \quad (6)$$

На основе нелинейного МНК получены следующие оценки параметров распределенных запаздываний (5)–(6): $\lambda_{11}=0.812$; $b_{11}=0.524$; $b_{12}=-0.501$; $\lambda_{22}=0.835$; $a_{21}=0.503$, причём коэффициенты вариации для $x_{i,t}$ составили, соответственно, 0.040 и 0.073.

На рис. 9 приведен прогнозный сценарий динамики суммарной производительности суперкомпьютеров на 5-летний период при таком же сценарии роста производительности суперкомпьютеров-чемпионов, как и для предыдущей модели. Расчеты показывают, что при действии зависимостей (5)–(6) в перспективе, уже в 2025 г. суммарные производительности суперкомпьютеров в США и Китае будут составлять, соответственно, 14.8 и 12.2 экзафлопс, а в 2027 г. — примерно по 45 экзафлопс, после чего начнётся существенное превышение производительности китайских суперкомпьютеров.

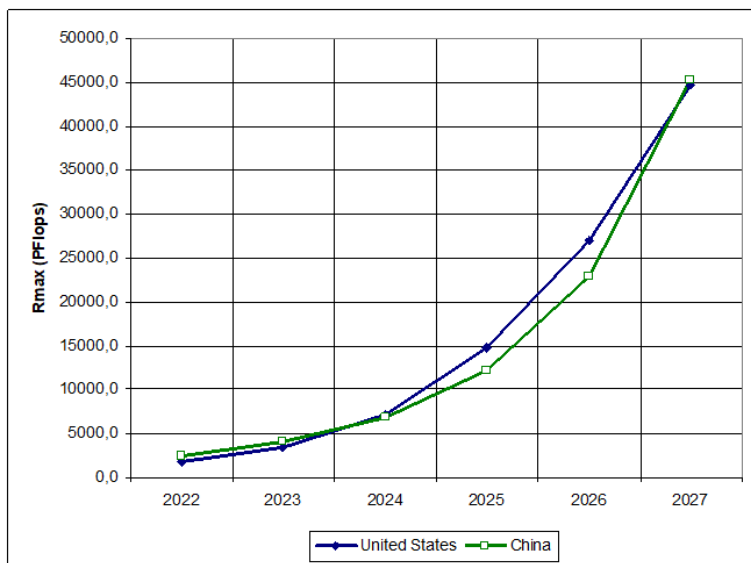


Рис. 9. Динамика суммарной производительностей суперкомпьютеров США и Китая, рассчитанная на основе модели (5)–(6).

Другая построенная нами модель состоит из двух распределенных запаздываний, связывающих: 1). натуральные логарифмы производительности суперкомпьютера-чемпиона в списке Top-500 ξ_{1t} (входная переменная) и суммарной производительности суперкомпьютеров США x_{1t} ; 2). натуральные логарифмы производительности суперкомпьютера-чемпиона Китая ξ_{2t} (входная переменная) и сум-

марной производительности суперкомпьютеров Китая x_{2t} . Эти распределенные запаздывания имеют следующий вид:

$$x_{1t} = 2\lambda_{11} x_{1t-1} - \lambda_{11}^2 x_{1t-2} + b_{11} \xi_{1t} + b_{12} \xi_{1t-1} \quad (7)$$

$$x_{2t} = 2\lambda_{22} x_{2t-1} - \lambda_{22}^2 x_{2t-2} + b_{21} \xi_{2t} + b_{22} \xi_{2t-1}. \quad (8)$$

На основе нелинейного МНК получены следующие оценки параметров распределенных запаздываний:

$$\lambda_{11}=0.812; b_{11}=0.524; b_{12}=-0.501; \lambda_{22}=0.628; b_{21}=0.450; b_{22}=-0.252,$$

причём коэффициенты вариации для уравнений (7) и (8) составили, соответственно, 0.040 и 0.062. На рис. 10 приведен прогнозный сценарий динамики суммарной производительности суперкомпьютеров на 5-летний период при таком же сценарии роста производительности суперкомпьютеров-чемпионов ξ_{1t} в модели (5), как и для предыдущей модели. В то же время в модели (8) принимались прогнозные значения входной переменной ξ_{2t} , соответствующие возможностям суперкомпьютеров с многоядерной архитектурой типа Sunway Oceanlite (оцененным экспертно), равные в 2025 г. 4 экзафлопс (при числе ядер 80 млн.), и в 2030 г. 7 экзафлопс (при числе ядер 100 млн.)⁶. Расчеты показывают, что при действии зависимостей (7)–(8) в перспективе, уже в 2025 г. суммарные производительности суперкомпьютеров в США и Китае будут составлять, соответственно, 14.8 и 23.5 экзафлопс, а в 2027 г. — 45 и 69 экзафлопс, после чего продолжится усиление отрыва суммарной производительности суперкомпьютеров Китая от США.

⁶ По некоторым оценкам максимальная производительность китайского суперкомпьютера-чемпиона с многоядерной архитектурой процессоров может составить только 4 экзафлопс, и то без учета последних санкций США, введенных в октябре 2022 года.

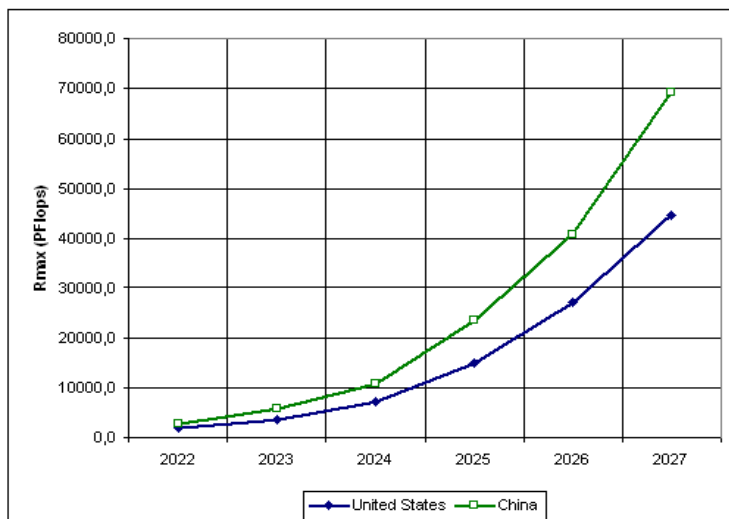


Рис. 10. Динамика суммарной производительностей суперкомпьютеров США и Китая, рассчитанная на основе модели (7)–(8).

5. Выводы

Проведенный анализ показывает, что значительные структурные сдвиги в производительности суперкомпьютеров из списка 500 происходят в моменты ввода суперкомпьютеров-чемпионов. В дальнейшем краткосрочном периоде происходит выравнивание структуры суперкомпьютеров по производительности.

За последние 15 лет по мере роста производительности и числа ядер произошло существенное сужение диапазона используемой в суперкомпьютерах тактовой частоты,

Использование различных построенных эконометрических моделей в прогнозных сценариях роста суммарной производительности наиболее мощных суперкомпьютеров разных стран приводит к разным результатам. В связи с возможностью резких изменений производительности суперкомпьютеров-чемпионов уже в ближайшие 3–5 лет, а также с возможностью введения разного рода санкций, эконометрический подход целесообразно объединять с экспертными оценками.

Литература

1. Варшавский Л.Е. Проблемы повышения энергоэффективности аппаратных средств в области информационных технологий//Труды ИСА РАН. 2013. т. 63. вып. 3. С. 3–19.
2. Варшавский Л.Е. а. Экономико-статистический анализ тенденций развития суперкомпьютеров//Концепции, 2022, №1 (41), с.70–77.
3. Варшавский Л.Е. б. Социально-экономические проблемы развития информационно-коммуникационных технологий (ИКТ).– М.: ЦЭМИ РАН, 2022. – с.158.
4. Макаров В.Л., Бахтизин А.Р., Сушко Е.Д., Сидоренко М.Ю., Хабриев Б.Р. Суперкомпьютерные технологии в общественных науках М.: Департамент научных изданий, ГАУГН, 2022. – 386 с. ISBN 978-5-604-80427-8.
5. Митропольский Ю.И. Элементарная база и архитектура будущих суперкомпьютеров//Микроэлектроника. 2015.—Том 44.—№3, с. 163–179.
6. Dennard R. H., Gaensslen F. H., Yu H-N, Rideout V.I., Bassous E., and LeBlanc A. R., Design of ion-implanted MOSFET's with very small physical dimensions// IEEE Journal of Solid-State Circuits. 1974. SC-9, pp. 256–268.

References in Cyrillics

1. Varshavskij L.E. Problemy povysheniya jenergojeffektivnosti apparatnyh sredstv v oblasti informacionnyh tehnologij//Trudy ISA RAN. 2013. t. 63. vyp. 3. S. 3–19.
2. Varshavskij L.E. a. Jekonomiko-statisticheskij analiz tendencij razvitija superkomp'juterov//Konceptii, 2022, #1 (41), s.70–77.
3. Varshavskij L.E. b. Social'no-jekonomicheskie problemy razvitija informacionno-kommunikacionnyh tehnologij (IKT).– M.: CJeMI RAN, 2022. – s.158.
4. Makarov V.L., Bahtizin A.R., Sushko E.D., Sidorenko M.Ju., Habriev B.R. Superkomp'juternye tehnologii v obshhestvennyh naukah M.: Departament nauchnyh izdaniy, GAUGN, 2022. – 386 s. ISBN 978-5-604-80427-8.
5. Mitropol'skij Ju.I. Jelementnaja baza i arhitektura budushhih superkomp'juterov//Mikrojelektronika. 2015.—Tom 44.—#3, s. 163–179.

Ключевые слова

Суперкомпьютеры, структура, эконометрические модели.

Варшавский Леонид Евгеньевич – д.э.н., главный научный сотрудник
Федерального государственного бюджетного учреждения науки Центральный экономико-
математический институт Российской академии наук"
г. Москва, Россия. E-mail: hodvar1@gmail.com

Leonid Varshavsky. Modeling the dynamics of the structure and performance of high-performance supercomputers.

Keywords

Supercomputers, structure, econometric models

DOI: 10.34706/DE-2023-01-05

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract.

The article discusses the dynamics of the structure and performance of a fleet of high-performance supercomputers from the Top 500 list. Econometric models of the interrelated development of the performance of supercomputers in the USA and China are considered. The predictive properties of several econometric models constructed in the work are investigated. The structure of the fleet of high-performance supercomputers is analyzed by the processor clock frequency used.

1.6. АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ САРКАЗМА В ТЕКСТАХ НА РУССКОМ ЯЗЫКЕ

Гурин А.А.¹, Жуков Т.А.²
^{1,2}РЭУ им. Г.В. Плеханова. Москва, Россия

В данной статье описываются подходы и методы для определения саркастических выражений на русском языке. Было разработано и внедрено решение, позволяющее определять сарказм для микроблоговой платформы Twitter.

Введение

С греческого сарказм (σαρκασμός) переводится, как «разрывать плоть». Сарказм является одним из видов сатирического изобличения, язвительной насмешкой, высшей степенью иронии, которая основана не только на усиленном контрасте подразумеваемого и выражаемого, но и на немедленном намеренном обнажении подразумеваемого [Лапина, Маталина, Секачев, 2008]. Высказывания и художественные произведения, написанные с сарказмом, утверждают одно, но дают ясно понять, что подразумевают противоположное, например, при помощи издевательской гиперболы или интонации. К сарказму часто прибегали античные ораторы в политической борьбе, чтобы обвинить оппонента.

Сарказм встречается в произведениях, относящихся к эпохе Древней Греции, например, в поэме Гомера «Одиссея» во второй песне: *«Меж тем женихи, изобильный обед учреждая,*

*Многими колкими сердце его оскорбляли речами.
 Так говорили одни из ругателей дерзко-надменных:
 «Нас Телемах погубить не на шутку замыслил; быть может,
 Многих он в помощь себе приведет из песчаного Пилоса, многих
 Также из Спарты; о том он, мы видим, заботится сильно.
 Может случиться и то, что богатую землю Эфиру
 Он посетит, чтоб, добывши там яду, смертельного людям,
 Здесь отравить им кратеры и разом нас всех уничтожить».*

В эпохе Древнего Рима тоже встречается сарказм. Так, в словаре латинских выражений Сомова В.П. [Сомов, 1992] встречаются саркастические выражения, например:

An nescis, mi fili, quantilla prudentia mundus regatur? — Сын мой, разве ты не знаешь, как мало надо ума, чтобы управлять миром?

Esse spectaculum dignum, ad quod respiciat intentus operi suo Deus — Вот зрелище, достойное того, чтобы на него оглянулся Бог, созерцая своё творение.

Англо-ирландский писатель-сатирик Джонатан Свифт использовал в своем творчестве приемы иронии и сарказма, так, в 1729 году появилась сатира «Скромное предложение», где автор обращается к правительству и предлагает просто съесть бедных детей Ирландии, чтобы решить проблемы голода и перенаселения. Это саркастическое предложение было адресовано правящему английскому высшему классу того времени. *Один очень образованный американец, с которым я познакомился в Лондоне, уверял меня, что маленький здоровый годовалый младенец, за которым был надлежащий уход, представляет собою в высшей степени восхитительное, питательное и полезное для здоровья кушанье, независимо от того, приготовлено оно в тушёном, жареном, печёном или варёном виде. Я не сомневаюсь, что он также превосходно подойдёт и для фрикасе или рагу.*

Кроме того, сарказм встречается и в классической литературе. Он используется для резкой и жесткой критики лица или объекта действительности. Например: А.С. Грибоедов в комедии «Горе от ума» описывает ситуацию, где отец Софьи Павел Афанасьевич Фамусов, кокетничая со служанкой Лизой, говорит: *«Скромна, а ничего кроме проказ и ветру на уме...».* Другой пример: Н. В. Гоголь в поэме «Мертвые души» пародийно описывает смерть бедного городского прокурора: *«Послали за доктором, чтобы пустить кровь, но увидели, что прокурор был уже одно бездушное тело. Тогда только с соболезнованием узнали, что у покойника была, точно, душа, хотя он по скромности своей никогда ее не показывал».* Таким образом, сарказм — это издевка в завуалированной форме, содержащая уничижительную оценку лица, предмета или явления действительности. Сегодня саркастические выражения можно встретить в социальных сетях, микроблогах, сервисах отзывов и на многих других платформах.

Распознавание сарказма в текстовых данных является одной из сложных задач обработки естественного языка, и в последнее время она стала интересной областью исследований из-за ее важности для интеллектуального анализа эмоциональной окраски текстовых сообщений в социальных сетях [Долбин, 2018]. Данная задача неразрывно связана с проблемой анализа тональности в целом. На сегодняшний день существуют различные методы анализа текста на естественном языке, которые предоставляют возможности по извлечению признаков и установлению семантических связей [Рубцова, 2014], [Gurin, 2020], [Mishra, Dey, Bhattacharyya, 2017]. Однако, автоматическое распознавание тональности сильно отличается от естественного, в связи с этим задача определения сарказма на естественном языке все еще является актуальной.

С точки зрения анализа текста задача определения сарказма может быть полезна при анализе ответов пользователей для правильного выбора стратегии общения с виртуальным помощником или для автоматического анализа и верной интерпретации комментариев, отзывов, размещенных в социальных медиаресурсах [Долбин, 2018].

Существует ряд исследований, посвященных автоматическому распознаванию сарказма в тексте [Bouazizi, Ohtsuk, 2016], [Son, Kumar, Sangwan, 2019], [Долбин, 2018]. Данные комплексные решения можно разделить на две группы. Первая группа включает решения, построенные на лингвистической структуре сарказма. Под этим понимается синтаксический разбор предложений по определенным правилам, вследствие чего представляется возможным определить сарказм. Более детально принцип метода изложен в работах [Bharti, Babu, Jena, 2015], [Riloff, Qadir, Surve, 2013].

Вторая группа решений основана на методах машинного обучения. В данную группу входят подходы с использованием машинного обучения с учителем, глубокое обучение на основе сверточных нейросетей, преднатренированные языковые модели, такие как Word2Vec, GloVe, BERT и другие [Devlin, Chang, 2018], [Евдокимова, 2022].

Для оценки качества построенной модели с использованием методов машинного обучения применяются следующие метрики: Accuracy, Precision, Recall и F1.

Метрики рассчитываются по следующим формулам:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

$$\text{Precision} = TP/(TP+FP)$$

$$\text{Recall} = TP/(TP+FN)$$

$$F1 = 2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}))$$

Здесь TP – количество документов, по которым классификатор принял правильное решение среди сообщений, в которых присутствует сарказм, TN – количество документов, по которым классификатор принял правильное решение среди сообщений, в которых отсутствует сарказм. FP – количество документов, по которым классификатор принял ошибочное решение, классифицировав их, как содержащие сарказм, FN – количество документов, по которым классификатор принял ошибочное решение, классифицировав их, как не содержащие сарказм.

Решения и исследования в области определения сарказма представлены на английском, испанском, итальянском, французском, японском, китайском, корейском и арабском языках [Baroiu, 2022], [Edoardo, 2020], [Elsa, Segura-Bedmar, 2021], [Khalid, 2020]. В работах [Bjarke, Mislove, Oegaard, 217], [Buckley, Paltoglou, Cai, 2019] представлены решения для английского языка. Среди них DeepMoji, созданный в Массачусетском технологическом институте [Bjarke, Mislove, Oegaard, 217]. Данное решение было обучено на 1,2 миллиарде сообщений в Twitter, содержащих смайлики, для обучения использовалась модель глубокого обучения, которая учитывает многие нюансы того, как язык используется для выражения эмоций. Например, она хорошо улавливает сарказм и сленг. Предварительно обученная модель DeepMoji доступна для использования и тестирования, а также реализована для фреймворков Keras и pyTorch. Точность определения сарказма по метрике Accuracy составляет деления тональности текста и эмоций [Buckley, Paltoglou, Cai, 2019]. На базе данного программного обеспечения существует надстройка для определения сарказма на английском языке. Точность определения сарказма по метрике Accuracy составляет 65% [Eisner, Rocktaschel, Augenstein, 2021]. Существуют работы, в которых используются архитектуры с LSTM (долгосрочная, краткосрочная память) [Arshad, 2022] и преднатренированные векторные представления слов (GloVe, FastText) [Jeff, Socher, 2014], [Arshad, 2022].

В работе [Reynier, Francisco, Delial, 2021] была построена модель определения сарказма для испанского языка, точность определения по метрике F1 составляет 68,3%. В работе [Tommaso, Nicole, Viviana, 2018] представлены решения для итальянского языка, оценка качества работы модели по метрике Precision составляет 51,8%. В работе [Radu, 2021] создано решение для французского языка, точность определения сарказма составляет 74,01% по метрике Accuracy. Для японского языка тоже существуют решения. Так, в работе [Sayed, Agrawal, Darkunde, 2020] точность определения сарказма составляет 79% по метрике Precision. В исследовании [Zhu, 2015] точность определения сарказма для китайского языка составляет 73,6%, оценка выполнялась по метрике F1. Для корейского языка в исследовании [Raijmakers, 2021] была построена модель, которая обеспечивает точность в 65,4% по метрике F1. В работе [Mohammed, Eltyeb, Elsamani, 2021] было создано решение для арабского языка, точность определения сарказма 87,23% по метрике F1.

Выбор модели и метода реализации зависит от конкретной задачи и набора данных, на котором происходит обучение модели.

Постановка задачи и исходные данные

Целью данной работы является создание и реализация алгоритма распознавания саркастических выражений на русском языке.

Для достижения цели был выполнен ряд задач, а именно:

- были собраны наборы данных для обучения моделей для анализа тональности текста и определения сарказма;
- подготовлены данные для использования в моделях (проведена очистка от специальных символов, все слова приведены в начальную форму);

- построена модель, определяющая тональность предложений с разбиением на 3 класса (негативный, нейтральный и позитивный);
- обучена модель определения сарказма.

Так, для построения модели определения сарказма была собрана выборка русскоязычных текстов, которые включают хэш-тэг «сарказм» из социальной сети Twitter. Выборка содержит два набора данных. Первый набор предназначен для тренировки и проверки модели, он включает в себя: 6598 сообщений, из которых 5938 предназначены для обучающей выборки и 660 для проверки качества построенной модели. Второй набор предназначался для проверки построенной модели, он не использовался при ее построении и содержит 3044 сообщения. Все сообщения прошли стадию подготовки данных. В них заранее были удалены специальные символы, осуществлен перевод к нижнему регистру всех слов и приведение их в начальную форму с помощью библиотеки `ru morphology2` [Korobov, 2015].

Также, для построения функции анализа тональности был использован общедоступный набор данных `PolSentiLex` [Koltsova, Alexeeva, Kolcov, 2016]. Данный набор содержит 26849 текстов. Из них 10736 относятся к негативному, 13953 — к нейтральному и 2160 — к позитивному классам. Для выделения признаков в тексте использовались языковые модели `Word2Vec`, `Bert` и `ruBert` [Kuratov, Arkhipov, 2019], [Devli, Chang, 2019], [Евдокимова, 2022].

Для дальнейшего выделения саркастически настроенных сообщений использовался словарь тональности слов [Кулагин, 2022]. Схема работы приложения представлена на рисунке 1.

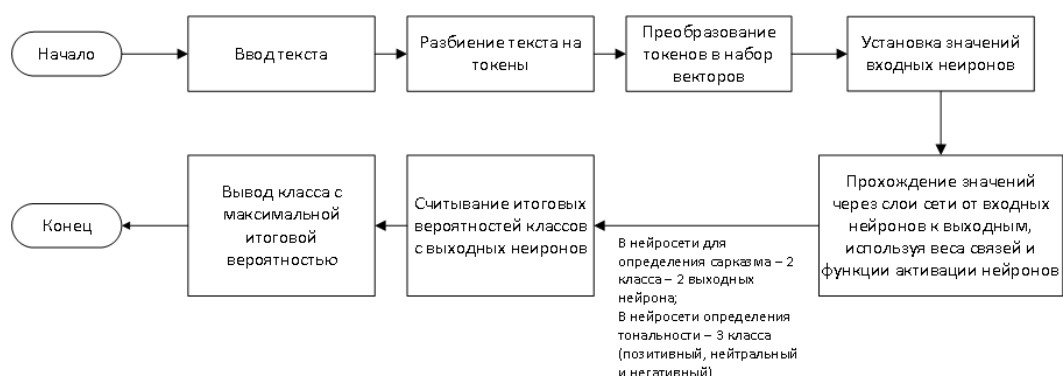


Рис. 1. Схема работы приложения для определения сарказма

Построение решения для определения тональности текста с использованием `Word2Vec`

В ходе исследования и изучения существующих подходов к распознаванию тональности сообщений была разработана нейросетевая модель, которая является сверточной нейросетью и использует эмбединги `Word2Vec` [Arshad, 2022]. Архитектура данной сверточной нейросети представлена в таблице 1.

Таблица 1. Архитектура сверточной нейросети

Слой	Размерность выхода	Количество параметров	Присоединен к
Входной слой	(N, 2800)	Без представления	Не присоединен
Эмбединг (Представление слов в виде вектора)	(N, 2800, 500)	40353500	Входному слою
Первый слой	(N, 2799, 200)	200200	Эмбедингу
Второй слой	(N, 2798, 200)	300200	Эмбедингу
Третий слой	(N, 2797, 200)	400200	Эмбедингу
Четвертый слой	(N, 2796, 200)	500200	Эмбедингу
Пятый слой	(N, 2795, 200)	600200	Эмбедингу
Снижение размерности выхода для первого слоя	(N, 200)	0	Первому слою
Снижение размерности выхода для второго слоя	(N, 200)	0	Второму слою
Снижение размерности выхода для третьего слоя	(N, 200)	0	Третьему слою
Снижение размерности выхода для четвертого слоя	(N, 200)	0	Четвертому слою
Снижение размерности выхода для пятого слоя	(N, 200)	0	Пятому слою
Объединение слоев	(N, 1000)	0	Снижению размерности выхода для слоев: 1,2,3,4,5
Метод исключения	(N, 1000)	0	Объединению слоев
Функция активации	(N, 128)	128128	Методу исключения

Всего в модели было использовано 42483015 параметров, из них обучаемых — 2129515, необучаемых — 40353500.

Данная модель использует 500-мерный вектор параметров слова (embedding) для снижения размерности параметров сети [Kuzmenko, Kutuzov, 2016]. Параметры обучения модели: обучение происходило на 49 941 примерах, валидация осуществлялась на 5 550 примерах, использовалось 3 эпохи обучения, среднее время эпохи обучения на CPU составляет 2,1 часа, общее время обучения на CPU — 6,3 часа.

Проведено обучение и тестирование моделей, основанных на модели НС с различными гиперпараметрами для сетей со сверточными слоями, представлено в таблице 2. В результате тестирования наилучший результат на тестовой выборке показала модель под номером 6.

Табл.2. Результаты обучения и тестирования Word2Vec моделей

Параметр/Модель	1	2	3	4	5	6	7
Количество фильтров	5	10	5	5	5	5	5
Размерность фильтра	200	200	200	200	100	400	200
Метод исключения	0,1; 0,2	0,1; 0,2	0,3	0,5	0,1;0,2	0,1;0,2	0,1; 0,2
Дополнительные признаки							Эмодзи
Точность на обучающем наборе (в %)	75,38	75,39	80,84	75,06	78,73	78,44	77,15
Точность на валидационном наборе (в %)	73,37	73,32	72,40	72,72	72,90	73,96	73,69
Точность на тестовом наборе (в %)	74,13	73,19	73,64	73,47	73,45	74,23	74,11
Время обучения (GPU)	0:14:28	0:31:13	0:20:06	0:21:00	0:08:20	0:26:15	0:13:20
Количество эпох	3	4	6	6	4	4	4
Размерность пакета	32	32	32	32	32	32	32

При оценке точности использовалась мера Ассигасу, которая вычисляется по формуле: Точность = P/N . Здесь P – количество документов, по которым классификатор принял правильное решение, а N – размер выборки.

Также была обучена модель с эмодзи эмбедингом, с итоговой точностью на тестовой выборке в 74,11%, не повысив результат по сравнению с базовым показателем в 74,13% во второй модели. В ней использован 300-мерный эмбединг эмодзи из emoji2vec [Eisner, Rocktaschel, 2021].

До построения вектора признаков текста собранный набор данных проходит предобработку, которая включает в себя несколько этапов, а именно:

- удаление специальных символов (точек, запятых, тире и др);
- преобразование всех слов и букв в слова к одному регистру, как правило, к нижнему;
- удаление стоп-слов, которые не несут в себе никакой информации о содержании;
- процесс лемматизации всех слов. То есть преобразование всех слов к начальной форме.

Технология Word2Vec работает с большим текстовым корпусом и по определенным правилам присваивает каждому слову уникальный набор чисел, так называемый семантический вектор. Библиотека NLTK [Bird, Klein, Loper, 2022] предоставляет инструменты для обработки естественного языка. В их числе есть метод, который позволяет исключать некоторые слова, которые не имеют эмоционального окраса. Смысл этой операции состоит в том, чтобы понизить количество признаков, извлеченных из текста, тем самым сделать модель более точной. Таким образом, был выполнен эксперимент для 6 модели с использованием стоп-слов и без их использования. Результаты данного эксперимента приведены в таблице 3.

Табл.3. Сравнение процесса обучения моделей с фильтрацией стоп-слов

Модель	6 модель без фильтрации стоп-слов		6 модель с фильтрацией стоп-слов	
	обучающая	валидационная	обучающая	валидационная
1 эпоха	66,20%	71,26%	64,84%	69,55%
2 эпоха	72,45%	72,92%	71,37%	72,77%
3 эпоха	75,47%	73,17%	74,33%	71,48%
4 эпоха	78,44%	73,96%	78,19%	72,43%
Максимум по		73,96%		72,77%

При тестировании полученных моделей были получены следующие результаты: при исключении стоп-слов из списка NLTK точность распознавания на тестовой выборке снижается до 72,10% (на 2,1%) по сравнению с моделью без исключений этих стоп-слов (74,24% точности) и на валидационной выборке до 72,77% (на 1,19%) против 73,96%.

Построение решения для определения тональности текста с использованием языковой модели BERT

BERT — это языковая модель, основанная на архитектуре «трансформер», предназначенная для предобучения языковых представлений с целью их последующего применения в широком спектре задач обработки естественного языка [Devlin, Chang, 2019]. По аналогии с Word2Vec используется для выделения признаков. Отличия BERT от Word2Vec заключаются в следующем:

- Модели Word2Vec генерируют вложения, которые не зависят от контекста: то есть — для каждого слова существует только одно векторное (числовое) представление. Разные значения слова (если они есть) объединяются в один вектор. BERT генерирует эмбединг, который позволяет использовать более одного векторного представления для одного и того же слова в зависимости от контекста, в котором это слово используется;
- Вложения Word2Vec не учитывают позицию слова. Модель BERT принимает в качестве входных данных позицию каждого слова в предложении перед вычислением его встраивания;
- Модель Word2Vec изучает вложения на уровне «слова». То есть модель Word2Vec была обучена на корпусе из 1 миллиона уникальных слов, тогда модель сгенерирует 1 миллион вложений слов — по одному вектору для каждого слова в словаре. Однако такие представления не могут генерировать векторы для слов, встречающихся за пределами словарного пространства. Так, Word2Vec не поддерживает слова вне словаря, что является одним из его основных недостатков. BERT, с другой стороны, изучает репрезентации в «подслово». Подслова рассматриваются как среднее между вложениями на уровне символов и вложениями на уровне слов. Таким образом, модель BERT может иметь словарный запас около 50 тысяч слов несмотря на то, что она обучена на корпусе из 1 миллиона уникальных слов [Zhu, 2015]. Этот вид моделирования стал очень популярным, потому что модель может генерировать векторное представление для любого слова и не ограничивается словарным пространством.

Таким образом, было проведено обучение нейросетей, основанных на Ru-BERT эмбединге с использованием фреймворка deepravlou [Blog DeepPavlov, 2021] и BERT [Devlin, Chang, 2019]. Результаты тестирования для различных эмбедингов BERT: многоязычная – 75% точности, русская (deepravlou), дополненная токенами эмодзи – 78,8% точности. Значения гиперпараметров для наилучшей модели, результаты представлены в таблицах 4 и 5.

Табл.4. Результаты обучения и тестирования deepravlou RuBERT моделей

Параметр\Модель	1	2	3	4
Максимальная длина сообщения в токенах	64	200	512	64
Вероятность сохранения веса связей между слоями	0,5	0,5	0,5	0,2
Скорость изменения весов при обучении	1,0E-05	1,0E-05	1,0E-05	1,0E-05
Максимальное количество итераций обучения без уменьшения	5	5	5	5
Точность на обучающем наборе (в %)	84,260	85,020	84,670	83,570
Точность на валидационном наборе (в %)	78,610	79,440	78,280	78,460
Точность на тестовом наборе (в %)	78,820	79,284	78,820	78,310
Время обучения (GPU)	0:18:36	1:14:44	1:21:01	0:19:40
Количество эпох обучения	1	1	1	1
Размерность пакета	64	32	8	64

Табл.5. Результаты обучения и тестирования deepravlou RuBERT моделей (продолжение)

Параметр\Модель	5	6	7	8
Максимальная длина сообщения в токенах	64	64	64	64
Вероятность сохранения веса связей между слоями	0,5	0,5	0,5	0,5
Скорость изменения весов при обучении	5,0E-06	2,0E-05	1,0E-05	1,0E-05
Максимальное количество итераций обучения без уменьшения	5	5	2	10
Точность на обучающем наборе (в %)	78,830	78,430	78,900	78,360
Точность на валидационном наборе (в %)	85,660	85,500	84,430	84,100
Точность на тестовом наборе (в %)	78,790	78,710	79,000	78,600
Время обучения (GPU)	1:45:59	0:17:31	0:46:10	0:17:42
Количество эпох обучения	2	1	1	1
Размерность пакета	64	64	64	64

Для обучения использовался токенизатор «bert preprocessor» со словарями токенов, включенных в эмбединги. По итогам тестирования была определена наилучшая модель под номером 2 с точностью на тестовой выборке 79,284%. Матрица ошибок для каждого класса представлена в таблице 6.

Табл.6. Матрица ошибок для определения тональности текста
Актуальные результаты

Прогнозируемые результаты	Актуальные результаты		
	Негативный класс	Позитивный класс	Нейтральный класс
Негативный класс	1813	132	658
Позитивный класс	150	3590	1089
Нейтральный класс	731	1072	9263

Модель учета сарказма

В процессе разработки была составлена обучающая и тестовая выборки из собранного набора данных. Обучающая выборка строится по следующему принципу: если с помощью словаря окраски слов [Кулагин, 2022] текст оценен положительно по 3 классам (негативному, позитивному, нейтральному) и если в тексте положительных слов больше отрицательных, но при этом текст размечен как негативный, то в нем присутствует сарказм. При подсчете слов был использован словарь тональностей русского языка [Кулагин, 2022]. Блок-схема алгоритма определения сарказма в выборке для обучения модели представлена на рисунке 2.

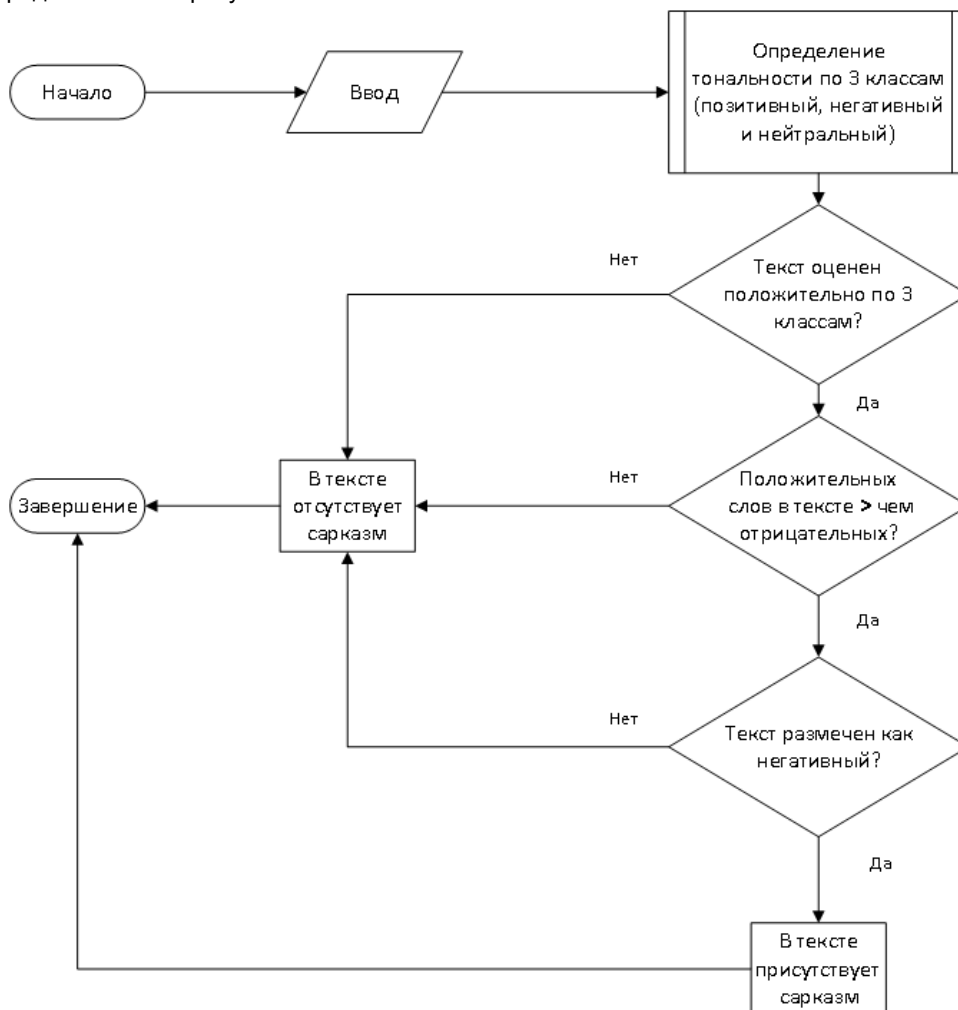


Рис. 2. Блок-схема алгоритма определения сарказма для выборки

Модель была обучена со следующими гиперпараметрами: размерность пакета = 16, максимальная длина сообщения в токенах = 128, вероятность сохранения веса связей между слоями = 0.5, скорость изменения весов при обучении = $1E-05$, максимальное количество итераций обучения без уменьшения = 5. По результатам тестирования итоговая точность модели составила: 78,57%.

После этого выборки для модели обнаружения сарказма в сообщениях были перестроены с учетом обоих вариантов сарказма, как отрицательного, выраженного положительными словами, так и положительного, выраженного отрицательными. Затем модель сарказма была переобучена на этих новых выборках. Итоговая точность распознавания: обучающая выборка — 95,49%, валидационная выборка — 90,75%, тестовая выборка — 69,19%. Матрица ошибок представлена в таблице 7.

Табл.7. Матрица ошибок саркастических выражений

		Актуальные результаты	
		Присутствует сарказм	Отсутствует сарказм
Прогнозируемые результаты	Присутствует сарказм	1048	464
	Отсутствует сарказм	474	1058

По итогам выполненных работ были сделаны следующие выводы.

Для обучения и оценки моделей распознавания сарказма необходима размеченная вручную, отдельная выборка, в которой имеется большинство комбинаций слов и других символов, которые можно интерпретировать как саркастические. Без данной выборки невозможно достоверно оценить любой из предлагаемых методов распознавания сарказма.

Также проблема распознавания сарказма очень часто зависит от контекста, в котором передается информация и картины мира говорящего, что делает невозможным правильное распознавание без использования большого количества дополнительной информации, которая не содержится в обучающей выборке.

Исключение или удаление стоп-слов, а также добавление пространства со смайликами негативно сказывается на точности определения модели распознавания тональности текста.

Для дальнейшего тестирования модели было решено собрать другой список саркастических выражений, которые встречаются в различных произведениях, фильмах, книгах, и протестировать функцию определения сарказма.

Построенная функция определения сарказма доступна по адресу: <http://passare.ru/sarcasm.php>

Набор данных для тестирования доступен по ссылке: <https://disk.yandex.ru/d/B825xhig9HC97w>

Точность на собранном наборе данных составляет 68 %. Всего 423 примера, верно определено 288 примеров, ошибочно — 135. В таблице 8 представлена численная оценка качества построенной модели.

Табл.8. Оценка качества модели.

Метрика	Точность в %
Accuracy	69,18
Precision	69,31
Recall	68,85
F1	69,08

Правильно были определены следующие примеры (10 штук из выборки):

У вас ничего не получается с первого раза. Парашютный спорт как раз для вас.

Продолжайте закатывать глаза. Может, там вы найдете свой мозг.

Бог юморист. Не верите? Посмотрите в зеркало.

Не каждая серая масса обладает мозгом. Правда?

Я не съел ваш кусок торта. Я просто спасал вашу фигуру.

Сходят с ума только те, у кого он есть.

Может быть, вам дать еще ключ от квартиры, где деньги лежат?

Я хлопал не потому, что мне понравилось выступление, а по той причине, что оно закончилось.

Несмотря на выражения моего лица, вы продолжаете говорить?

К счастью для вас, зеркала не могут смеяться.

Ошибочно были определены следующие примеры (10 штук из выборки):
Продолжайте говорить. Я всегда зеваю, когда мне интересно.
Ваше мнение очень интересно. Давайте поговорим об этом в следующем месяце.
Безусловно, я работаю меньше вас. Просто я делаю все правильно с первого раза.
Вам следует записывать все интересные мысли и идеи. На одном стикере для заметок вам хватило бы места.

Ты мне так же нравишься, как понедельники.
Я терпелив, ведь здесь слишком много свидетелей.
Я понимаю, что твои друзья круче, ведь они невидимы.
Ваша пылкая речь не делает вас правым.
Когда мои глаза закрыты, ты лучше всего выглядишь.
Если вы уйдете, то я не буду скучать.

Заключение

В процессе выполнения работы были достигнуты следующие результаты:

- был разработан алгоритм определения сарказма в русскоязычных текстах;
- построена модель определения тональности сообщений на русском языке по трем классам: позитивному, негативному и нейтральному;
- реализован алгоритм на языке Python 3.7 для определения тональности текстов и сарказма;
- разработано несколько подходов на языке Python 3.7 определения тональности сообщений: на основе сверточных нейронных сетей с использованием Word2Vec эмбедингов с итоговой точностью 74,23%; на основе фреймворка deeppavlov и эмбединга RuBERT с итоговой точностью 79,44%;
- разработана модель определения сарказма в сообщениях из Twitter с точностью 69,19%.

Однако, все еще остается ряд вопросов, предложений и улучшений к уже существующим решениям. Так, например, существуют решения для английского языка, которые определяют, стоит ли понимать буквально смысл постов в Twitter, Instagram и прочих социальных медиа. Для этого решение анализирует не только текст, но и изображение, прикрепленное к публикации [Emspak, 2022]. Другой пример основывается на анализе лексических индикаторов, языковых маркеров и информации о контексте, всех предыдущих твитов и действия пользователя в соцсети. Ввиду того, что изучение только текста недостаточно. Немаловажный контекст обеспечивают изображения. Это особенно актуально для таких соцсетей, как Twitter, Instagram и Tumblr, в которых изображения изначально несут более важную смысловую нагрузку, чем текст. Для обнаружения сарказма в Twitter, Instagram и Tumblr также предложены две различные вычислительные структуры, которые объединяют анализ текстовой и визуальной информации. Первый подход основан на методе опорных векторов, как и в большинстве подобных исследований. Этот метод был дополнен для работы не только с текстовой, но и с визуальной информацией. Вторым подходом основан на глубинном обучении нейросети на базе изображений ImageNet. Таким образом, лучшие результаты (80–88% распознавания) позволяют получить сочетание двух методов [Emspak, 2022].

Данное исследование выполнено в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ на тему "Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и стилизации многомерных данных, прогнозирования временных рядов и проектирования рекомендательных систем", номер проекта FSSW-2023-0004.

Литература

1. Долбин, А. В. Распознавание сарказма в задаче определения тональности текста на естественном языке. Молодой ученый. 2018. № 49 (235). — С. 13–17.
2. Евдокимова И.С., Евдокимов Д.А. Модель идентификации саркастических предложений в естественно-языковом тексте. Интернаука: электрон. научн. журн. 2022. № 22 (245).
3. Кулагин Д. И. Открытый тональный словарь русского языка КартаСловСент URL: https://github.com/dkulagin/kartaslov/tree/master/dataset/emo_dict (Дата обращения: 10.01.2023).
4. Лапина, И. К. Большой энциклопедический словарь. / И. К. Лапина, Е. Н. Маталина, Р. Г. Секачев. — М.: АСТ, Астрель, 2008. — 553 с.
5. Рубцова Ю.В. Методы автоматического извлечения терминов в динамически обновляемых коллекциях для построения словаря эмоциональной лексики на основе микроблоговой платформы Twitter. Доклады Томского государственного университета систем управления и радиоэлектроники. 2014, № 3 (33). — С.140-144.
6. Сомов В. П. По-латыни, между прочим. Словарь латинских выражений. — М.: Гитис, 1992. — 230 с.
7. Arshad S, «Sentiment Analysis / Text Classification Using CNN URL: <https://towardsdatascience.com/cnn-sentiment-analysis-1d16b7c5a0e7>. (Дата обращения: 20.01.2023).
8. Baroiu. A. Automatic Sarcasm Detection. Systematic Literature Review. Information 2022, vol.13, p. 399.

9. Bharti S. K, Babu K. S, Jena S. K. Parsing-based Sarcasm Sentiment Recognition in Twitter Data. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, 2015, p. 1373-1380.
10. Bird S, Klein E, Loper E. Natural Language Processing with Python — O'Reilly Media Inc, 2009. — ISBN 0-596-51649-5
11. Bjarke F, Mislove A, Søgaard A, Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion, and sarcasm. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017, p. 1615–1625.
12. Bouazizi M, Ohtsuki T. O. A pattern-based approach for sarcasm detection on Twitter. IEEE Access. 2016, vol. 4, p. 5477-5488.
13. Buckley M, Paltoglou K, Cai G, Kappas, A. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, issue 61 (12), 2019, p. 2544–2558.
14. DeepPavlov: an open source conversational AI framework. (б.д.) [Электронный ресурс]. – URL: <http://deeppavlov.ai/> (Дата обращения: 20.01.2023).
15. Edoardo S. An Exploration of Sarcasm Detection Using Deep Learning. Corso di laurea magistrale in Ingegneria Informatica (Computer Engineering), 2020, p.68.
16. Eisner B, Rocktäschel T, Augenstein I, Bošnjak M, and Riedel S, emoji2vec: Learning Emoji Representations from their Description. URL: <https://deepemoji.mit.edu/> (Дата обращения: 12.01.2023).
17. Elsa S, Segura-Bedmar I. Sarcasm Detection with BERT. Procesamiento del Lenguaje Natural, Revista nº 67, septiembre de 2021, p. 13–25.
18. Emspak J. Computers Can Sense Sarcasm? Yeah, Right– URL:<https://www.scientificamerican.com/article/computers-can-sense-sarcasm-yeah-right/> (Дата обращения: 15.01.2023).
19. Gurin A.A. Methods for Automatic Sentiment Detection. Proceedings of the 10th International Scientific and Practical Conference named after A. I. Kitov "Information Technologies and Mathematical Methods in Economics and Management (IT&MM2020)". Moscow, Russia, October 15-16, 2020. CEUR Workshop Proceedings. – 2021. – vol. 2830.
20. J. Devlin Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. Google AI Blog. Retrieved 2019-11-27. (Дата обращения: 10.01.2023).
21. Jeff P, Socher R, Manning. 2014. GloVe: Global Vectors for Word Representation. Computer Science Department, Stanford University, Stanford. IEEE Access. 2014, p.14.
22. Khalid A, Hämäläinen M, ¡Qué maravilla! Multimodal Sarcasm Detection in Spanish: A Dataset and a Baseline. Computation and Language arXiv preprint :2105.05542.
23. Koltsova O.Y, Alexeeva S.V, Kolcov S.N. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. Компьютерная лингвистика и интеллектуальные технологии. 2016. С. 277–287.
24. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages. Analysis of Images, Social Networks and Texts, 2015, p. 320–332.
25. Kuratov. Y, Arkhipov. M. (2019). Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language. arXiv preprint:1905.07213.
26. Mishra A, Dey K, Bhattacharyya P. Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. vol. 1, 2017. p. 377–387.
27. Mohammed M, Eltyeb S, A. Elsamani, Automatic Sarcasm Detection in Arabic Text:A Supervised Classification Approach, International Journal of New Technology and Research (IJNTR) ISSN: 2454-4116, vol.7, Issue 8, August 2021 p. 32-42.
28. Radu T, FreSaDa: A French Satire Data Set for Cross-Domain Satire Detection, arXiv:2104.04828v1
29. Raijmakers D, Dirk R, Emojis, Emoticons and Sarcasm Detection for English and Korean Twitter Data, Thesis submitted in partial fulfillment of the requirements for the degree of master of science in data science and society department of cognitive science and artificial intelligence school of humanities and digital sciences Tilburg university.
30. Reynier O,Francisco R,Delia I,Paolo R,Manuel M,Jos´e E, Overview of the Task on Irony Detection in Spanish Variants, CEUR-WS, Vol. 2421.
31. Riloff E, Qadir A, Surve A, De Silva L, Gilbert N, Huang R. Sarcasm as contrast between a positive sentiment and negative situation. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2013, p. 704–714.
32. Sayed S, Agrawal N, Darkunde M. Deep LSTM-RNN with Word Embedding for Sarcasm Detection on Twitter. 2020 International Conference for Emerging Technology (INCET).
33. Son L, Kumar A, Sangwan S, Arora A, Nayyar A. Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. IEEE Access. vol. 7. 2019. p. 23319–23328.
34. Tommaso C, Nicole N, Viviana P, Paolo R, EVALITA Evaluation of NLP and Speech Tools for Italian, Proceedings of the Final Workshop 12-13 December 2018, Naples.

35. Zhu Y, Aligning Books and Movies. Towards Story-like Visual Explanations by Watching Movies and Reading Books. Aligning Books and Movies. Towards Story-like Visual Explanations by Watching Movies and Reading Books. 2015.

References in Cyrillics

1. Dolbin, A. V. Raspoznavanie sarkazma v zadache opredeleniya tonal'nosti teksta na este-stvennom yazyke . Molodoj uchenyj. 2018. № 49 (235). — S. 13-17.
2. Evdokimova I.S, Evdokimov D.A. Model' identifikacii sarkasticheskikh predlozhenij v estestvenno-yazykovom tekste. Internauka: elektron. nauchn. zhurn. 2022. № 22 (245).
3. Kulagin D. I. Otkrytyj tonal'nyj slovar' russkogo yazyka KartaSlovSent URL: https://github.com/dkulagin/kartaslov/tree/master/dataset/emo_dict (Data obrashcheniya: 10.01.2023).
4. Lapina, I. K. Bol'shoj enciklopedicheskiy slovar'. / I. K. Lapina, E. N. Matalina, R. G. Seka-chev. — M.: AST, Astrel', 2008. — 553 с.
5. Rubcova YU.V. Metody avtomaticheskogo izvlecheniya terminov v dinamicheski obnovlyaemykh kollekcijah dlya postroeniya slovarya emocional'noj leksiki na osnove mikroblogovoj platformy Twitter. Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki. 2014, № 3 (33). — S.140-144.
6. Somov V. P. Po-latyni mezhdru prochim. Slovar' latinskih vyrazhenij. — M.: Gitis, 1992. — 230 s.

Ключевые слова

Определение тональности текста, определение сарказма, модель Word2Vec, языковая модель BERT, обучение с учителем.

*Гурин Анатолий Анатольевич,
лаборант-исследователь учебно-научной лаборатории искусственного интеллекта,
нейротехнологий и бизнес аналитики РЭУ им. Г.В. Плеханова,
anatoly196674@gmail.com*

*Жуков Тимур Алекперович,
лаборант-исследователь учебно-научной лаборатории искусственного интеллекта,
нейротехнологий и бизнес аналитики РЭУ им. Г.В. Плеханова,
tim29093@gmail.com*

Keywords

Sarcasm detection in Russian, sentiment analysis, machine learning, BERT, Word2Vec, supervised learning

DOI: 10.34706/DE-2023-01-05

JELclassification C45 – Нейронные сети и связанные темы; M15 Управление информационными технологиями

Abstract

We discuss approaches forwards automatic detection of sarcastic expressions in Russians. A solution for sarcasm detection in Twitter posts is exposed.

1.7. ОПТИМАЛЬНЫЕ ДВУХКОМПОНЕНТНЫЕ ЦЕНЫ В ЭКОНОМИКАХ С ВОЗРАСТАЮЩЕЙ ОТДАЧЕЙ

Козырев А. Н. ЦЭМИ РАН, г. Москва, Россия

Показаны возможности построения и применения оптимальных двухкомпонентных цен в экономиках с возрастающей отдачей на масштаб. Оптимальность цен понимается как наилучшее соотношение между компонентами цены в виде одновременного платежа и ценой непосредственно за продукт. Обоснована необходимость смягчения антимонопольной политики в части использования ценовой дискриминации и ценовых алгоритмов. В качестве основной интерпретации компонентов цены рассматриваются паушальный (единовременный) платеж и роялти в лицензионном договоре (после приведения их к единому виду, то есть к выплатам в денежной форме). Также рассматривается интерпретация с выбором тарифа за подключение к сети и непосредственно за трафик, где плата за трафик тем ниже, чем больше плата за подключение.

1. Введение

Основной замысел этой статьи – показать необходимость смягчения антимонопольной политики в области ценообразования на продукцию наукоемких отраслей, причем представить это на доказательном уровне, используя типичные для микроэкономики промежуточного уровня [Вэриан, 1997]¹ изобразительные средства. Впрочем, такой подход не исключает обращения к примерам из реальной практики, как и применения математики чуть более высокого уровня, когда в этом есть потребность.

В первую очередь речь идет о необходимости снять запреты на использование ценовых алгоритмов и ценовой дискриминации, причем не только прямые запреты, но и опосредованные (через налогообложение). Такие запреты в условиях цифровизации, возрастания доли затрат на создание цифровых продуктов и развитие технологий при снижении почти до нуля доли текущих производственных затрат все сильнее тормозят технический прогресс. И на то есть объективные причины.

Следует подчеркнуть, что речь идет не о чисто российской проблеме, а о проблеме мирового масштаба. В дискуссии середины прошлого века о предельных издержках Рональд Коуз [Coase, 1946] обратил внимание на очевидный и потому упускаемый обычно из вида факт – в теоретических построениях экономистов правительство обычно представляется как всезнающее и всевидящее, хотя на самом деле оно таковым заведомо не является. А потому надо соизмерять издержки от «провалов рынка» с издержками от вмешательства в его работу. Этот тезис не утратил актуальности и сегодня, поскольку история повторяется и здесь.

Если говорить конкретно о проблемах ценообразования в условиях возрастающей отдачи на масштаб, то необходимо различать причины этого явления на уровне общей теории [Kim, 1997] и в каждом конкретном случае, когда применяются регуляторные меры. Обращение к таким конкретным случаям чрезвычайно плодотворно с точки зрения расширения кругозора и заимствования идей. Также уместно напомнить, что в бизнесе идеи рождаются и умирают постоянно. Они умирают, если оказываются преждевременными или нежизнеспособными, причем гораздо раньше, чем на них обратит внимание экономическая теория. До признания экономистами-теоретиками доживают лишь немногие из них. Именно так случилось с идеей применения двойного тарифа, или, точнее, двухкомпонентных цен, при возрастающей отдаче на масштаб [Coase, 1946].

Далее материал распределяется по разделам следующим образом. В непосредственно следующем за настоящим введением разделе 2 дан обзор идей о ценообразовании в условиях возрастающей отдачи на масштаб производства, высказанных разными авторами за достаточно длительный промежуток времени. Часть этих идей прочно вышла в учебники и сознание тех, кто учился по этим учебникам. В разделе 3 обсуждается модель экономики Робинзона Крузо с возрастающей отдачей на масштаб при производстве орехов, в которой есть оптимальное решение, но это решение не может быть реализовано как рыночное равновесие в традиционном понимании. Особое внимание к этой модели связано с тем, что она проста и наглядна, а потому кочует из учебника в учебник. Так, в переведенном на русский язык замечательном учебнике Хэлла Вэриана, выдержавшем уже 20 изданий, именно эта модель используется для демонстрации отмеченной проблемы недостижимости оптимума с использованием рыночных цен. В том же учебнике, но совершенно отдельно рассматриваются двухкомпонентные цены. В разделе 4 настоящей статьи эти сюжеты совмещены. Как и в учебном примере с Робинзоном Крузо, в модели имеется только два вида продуктов, один из которых невозпроизводим, другой производится, причем множество технологически достижимых наборов из двух продуктов имеет невыпуклую форму. Невыпуклость этого множества имеет форму параболы, как и в большинстве учебников, что имеет определенные основания. Предпочтения покупателя подчиняются условию постоянной эластичности замещения (CES). В качестве причины возрастания отдачи на масштаб рассматриваются вложения в развитие технологии, которые могут быть заменены покупкой лицензий. Показано, как при

¹ Перевод сделан с четвертого издания учебника Вэриана, выдержавшего уже 20 изданий.

таким выборе параметров использование двухкомпонентных цен решает проблему, поставленную перед неоклассической теорией в работе [Coase, 1946]. Далее эта модель распространяется на случай, когда лицензиатов много. Показано, что ставки роялти должны быть одинаковы для всех лицензиатов, а единовременные платежи - разными. В Заключении статьи кратко сформулированы выводы из всего предшествующего анализа.

2. Возрастающая отдача на масштаб, конкуренция и монополия.

Новые формы бизнеса, обычно ассоциируемые с цифровой экономикой, изначально ориентированы на получение выгод от внешних эффектов (экстерналий), возрастающей отдачей на масштаб и применения ценовых алгоритмов. Экономическая теория всегда увязывала все перечисленные явления с монополистическим поведением на рынках, что само по себе верно и не подвергается сомнению. Иначе обстоит дело в отношении практических рекомендаций по регулированию экономики в тех случаях, когда рыночный механизм дает сбой. К их числу относятся рекомендации по регулированию цен, включая запрещение ценовой дискриминации и ценовых алгоритмов, хотя то и другое – необходимые условия оптимальности в «экосистемах» современного бизнеса. Проблема имеет длинную историю.

2.1. Вложения в технологии и лицензионная торговля

Возрастающая отдача на масштаб производства может иметь разные, причем вполне конкретные причины, включая сетевые эффекты, обучение персонала в процессе производства и другие причины, не связанные напрямую с вложениями в развитие технологии. Но в данном случае интерес представляет именно тот случай, когда вложения в технологии или иные затраты, не связанные напрямую с текущим производством, приводят к снижению предельных издержек производства на единицу продукции. Вопрос о возмещении этих затрат обсуждался в англоязычной литературе, как минимум, со времен А. Маршала [Marshall, 1920], а в 30-х и 40-х годах 20-го века стал главным предметом спора о предельных издержках и целесообразности вмешательства государства в рыночное ценообразование [Coase, 1946]. Как ни удивительно (точнее, совсем не удивительно), спор продолжается в тех или иных формах и сейчас, о чем можно поговорить отдельно. В данном случае целесообразно сосредоточиться на вложениях в технологии и на лицензионной торговле с реальной практикой применения многокомпонентных цен. Как правило, это единовременный (паушальный) платеж и роялти (текущие выплаты).

Концентрация внимания на вложениях в технологии имеет, как минимум, две важные причины. Во-первых, это позволяет, не теряя содержания, свести все затраты к одной переменной, выражаемой в деньгах. Тем самым задача с выпуском одного однородного продукта сводится к виду, не многим отличному от учебного примера с Робинзоном и орехами. Во-вторых, появляется возможность параллельного рассмотрения разных возможностей вложения в развитие технологии, а именно: можно развить свою технологию, а можно купить и использовать по лицензии уже готовую.

Вариант с покупкой лицензии интересен по целому ряду причин. Главная из них состоит в том, что здесь практически всегда используется двухкомпонентная форма цены (разовый платеж и роялти), а у покупателя лицензии есть реальный выбор. Он может выбирать и поставщика технологии, и предпочтительное для него соотношение между компонентами цены, но уже по согласованию с поставщиком технологии. Так или иначе, здесь выбор тарифа из некоторого семейства возможных тарифов – реальность, а не учебный пример с фирмой по производству орехов «Робинзон Инк».

2.2. Антимонопольная политика

В современной экономической литературе на английском языке одновременно присутствуют диаметрально противоположные точки зрения на антимонопольное регулирование. В том числе всерьез обсуждается утверждение, что с приходом новых технологий «век антимонопольного регулирования закончился» [Basu, 2021]. В частности, в цитируемой работе приводятся доводы, суть которых – наличие почти неограниченных возможностей по обходу антимонопольных мер за счет применения новых технологий. К ним можно добавить доводы иного типа, а именно: значительная часть антимонопольных мер сама по себе деструктивна, поскольку направлена на уничтожение стоимости. В частности, это касается запретов на ценовую дискриминацию и применение ценовых алгоритмов, где как раз и открываются возможности для оптимизации с применением математики.

Антимонопольная политика России, как и большинства современных государств, во многом основывается на представлениях о конкуренции, формируемых при изучении экономикс на экономических факультетах университетов. Именно по этой причине необходимый шаг к отмене запретов – развенчание соответствующих мифов и условных примеров из учебников по экономикс. Для этого вполне логично обратиться к практике и проверить теорию не на условных примерах, кочующих из статьи в статью или из одного учебника в другой, а на примерах из реальной жизни. Но этого, строго говоря, мало, поскольку опровергать ту или иную теоретическую концепцию принято в тех условиях и терминах, в каких она сформулирована ее сторонниками. Разумеется, это не исключает обращения к практике, но история экономической мысли показывает, что аргументов от практики и практиков, как правило, недостаточно для изменения представлений экономистов-теоретиков о том, как устроена экономика. В этом смысле поучительна история идей Рональда Коуза, получившего в 1991 году Нобелевскую премию по экономике за работы [Coase, 1937, 1960], хотя лишь за несколько лет до этого события он считал, что его идеи так и не были поняты [Coase, 1988]. Признание пришло лишь после того, как высказанные в

этих статьях идеи были частично переформулированы и развиты в конструктивном направлении Оливером Уильямсоном [Williamson, 1985] и некоторыми другими экономистами. При номинации на Нобелевскую премию вообще не упоминались две другие работы Коуза [Coase, 1946, 1974], оцениваемые Ричардом Познером [Posner, 1993] даже выше работ 1937 и 1960 годов. Именно эти работы имеют самое непосредственное отношение к вопросам, обсуждаемым в настоящей статье. Как и другие работы Круза, они отличаются разбором реальных, а не условных примеров.

2.3. Спор о предельных издержках и двухкомпонентные цены

В статье [Coase, 1946] Коуз резко выступил против господствующей на тот момент концепции Хотеллинга [Hotelling, 1938], согласно которой монополии в условиях возрастающей отдачи на масштаб должны продавать свою продукцию или услуги по предельным ценам производства, а покрывать превышение средних затрат над предельными ценами предлагалось правительству за счет собираемых налогов. Коуз, как уже говорилось выше, обратил внимание на то, что в этой концепции заложено предположение о всезнающем и всевидящем правительстве, каковым оно по факту никогда не было и не будет в обозримом будущем. А потому предлагаемая Хотеллингом схема неизбежно приводит к необоснованному перераспределению средств. В качестве альтернативы перераспределению средств через налоги Коуз предложил использовать двухкомпонентные цены. В предлагаемой им схеме полная цена складывается из платежей двух видов: один платеж взимается «за вход» или доставку товара на рынок, второй – непосредственно за продукцию или услугу. Такая двухкомпонентная схема реально использовалась и используется во многих конкретных ситуациях. Но смысл платежей, их привязка к конкретным действиям и показателям может быть разным.

Примечательно, что статья [Coase, 1946] гораздо чаще цитируется юристами, чем экономистами. Им она понятнее и ближе, поскольку в качестве доводов им привычно использовать факты и прецеденты решений, а не математические конструкции и графики. Обратная сторона «той же медали» и слабость подхода на основе прецедентов – фактическое неприятие его экономистами, хотя попытки вписать идею Коуза в неоклассическую экономическую теорию были. Такая попытка с использованием теории игр предпринята, например, в статье [Stomper, 2015], причем начинается она с утверждения, что проблема, поставленная Коузом в 1946 году перед неоклассической экономической теорией, так и не была удовлетворительно решена. Впрочем, предлагаемое в [Stomper, 2015] решение тоже сложно назвать удачным. Но вернемся к обсуждению работ Коуза.

В статье [Coase, 1974] на основе неоспоримых фактов показано, что кочующий из публикации в публикацию пример маяка в качестве коллективного блага, создаваемого исключительно за счет налогов, имеет мало общего с действительностью. Как оказалось, в Англии маяки строились преимущественно в частном порядке. Капитаны кораблей всегда находили возможность заплатить за предоставленную услугу, поскольку их психология радикально отличается от психологии трамвайных «зайцев». Чувство собственного достоинства просто не позволяло им не платить. Разумеется, пример с маяком и капитанами можно рассматривать как исключение, но примеров, когда хрестоматийные примеры из учебников оказываются мифами или результатами подлога, не так уж мало. Однако мало кто отважится проверять такие примеры, а потом идти против господствующего мнения. А потому попытка Коуза спасти экономикс от экономистов [Coase, 2012] утопична, хотя и не лишена здравого смысла. Еще более радикальная критика экономикс с позиций экономического реализма [Keen, 2011] тоже не сильно повлияла на общую ситуацию, хотя для развития теории может быть полезна.

3. Невыпуклость технологий в экономике Робинзона Крузо

В экономической теории или, точнее, в её части, известной как экономикс, «провалы рынка», требующие вмешательства регулирующих органов, обычно связывают с эффектами, выглядящими на графиках как невыпуклость некоторого множества. Чаще всего это множество доступных наборов благ или продуктов в широком смысле слова.

Самая популярная в научной и учебной литературе по экономикс невыпуклость – следствие возрастающей отдачи на масштаб при использовании некой технологии. Например, в учебнике [Вэриан, 1997] в этом качестве выступает технология производства кокосов в фирме «Робинзон Инк.». Робинзон тратит время на производство кокосов со все возрастающей эффективностью, но оно ограничено и приобретает тем большую ценность, чем меньше его остается. Естественно предположить, что существует точка оптимума, когда нет смысла тратить ни больше, ни меньше времени на производство кокосов. На рисунке 1 оптимум – это точка касания множества достижимых наборов из кокосов и свободного времени со множеством более желательных наборов. Линия безразличия представляет собой гиперболу, а граница множества достижимых состояний – параболу или,

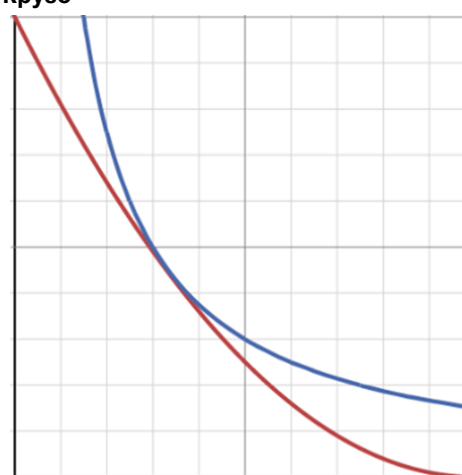


Рисунок 1. Оптимальный выбор между кокосами и свободным временем

выглядит в учебниках по экономикс. В учебнике Вэриана использован несколько иной рисунок, но форма кривых там примерно та же. Такое представление предпочтений и технологических возможностей, как минимум, удобно. К тому же оно согласуется с интуитивными представлениями о предпочтениях и роли технологий. Понятно, что разделить гиперплоскостью выпуклый надграфик (точки над гиперболой) и вогнутый подграфик (точки под параболой) невозможно. Отсюда обычно делается вывод, что невозможно достижение конкурентного равновесия, когда уровень потребления и уровень производства определяются независимо один от другого с ориентацией исключительно на цены, формируемые той самой «невидимой рукой», о которой обязан знать любой экономист. Цитируем Вэриана.

Однако если технология и/или предпочтения — невыпуклые, то цены не несут в себе всей информации, необходимой для выбора эффективного распределения ресурсов. Для этого требуется иметь также информацию о наклонах производственной функции и кривых безразличия в точках, весьма удаленных от точки текущего производства [Вэриан, 1997, с.579].

Цитируемое рассуждение целиком относится не столько к реальной экономике, сколько к ее модельному представлению в духе экономикс, но потом оно проецируется на реальность. Ситуация с одним производителем в виде фирмы Робинзона и одним потребителем в лице того же Робинзона выглядит, как «сказка». Из нее учащиеся должны вынести понимание проблемы, точнее, самой ее сути, не заморачиваясь деталями. Тем не менее, это стандартное для экономикс представление о роли цен в рыночной экономике, а именно: «невидимая рука рынка» передает через цены информацию о предпочтениях потребителей и возможностях производителей, приводя их в равновесие.

Тем не менее, даже в такой предельно упрощенной ситуации вполне можно предположить, что «невидимая рука» предлагает сторонам цены не в таком примитивном виде, а в виде комбинаций из двух или более слагаемых, определяемых по разным правилам. Например, можно предположить, что двухкомпонентная цена выбирается из набора или семейства тарифов, где фиксированный платеж и цена за единицу продукта связаны между собой, при этом более высокому фиксированному платежу соответствует более низкая цена за единицу продукта. Потребитель выбирает и объем потребления, и тариф, причем тариф выбирается им оптимально для данного объема потребления. Тогда бюджетное множество будет невыпуклым, и отделять множество доступных по цене продуктовых наборов от множества более предпочтительных, чем имеющийся продуктовый набор, будут не прямая или гиперплоскости, а какие-то более сложные линии или поверхности. В простейшем случае, как на рисунке 1, вместо прямых логично использовать разделяющие линии в виде фрагментов парабол или чего-то подобного им. Через ту же точку оптимума может проходить более крутая парабола или более плавная гипербола (со сдвигом), отделяя множество более желательных продуктовых наборов от множества технологически доступных продуктовых наборов.

Примечательно, что кривые такого типа могут иметь вполне ясный экономический смысл, но при этом лучше отказаться от «сказки» про Робинзона и кокосы. Лучше рассматривать более реалистичные примеры с двухкомпонентным ценообразованием. Например, это может быть семейство тарифов на связь или на трафик в сети, когда можно выбрать большую плату за подключение и меньшую за использованные минуты или переданные мегабайты. Важно подчеркнуть, что обычно в таких случаях имеется широкий спектр возможностей, причем все они схожи в том, что сначала предполагают затраты без видимой отдачи, потом линейную или близкую к линейной зависимость отдачи от затрат. Большим затратам вначале соответствуют более низкие текущие затраты потом. Это касается не только тарифов на подключение, но и на доступ к технологиям, получаемым по лицензиям или благодаря вложениям в разработки. Доступ к более совершенной технологии обходится дороже, но компенсируется большей эффективностью, если объем ее использования достаточно велик. Графически это можно изобразить, как на рисунке 2. По оси абсцисс откладываются затраты, а по оси ординат — выпуск продукта по используемой технологии. Верхняя огибающая напоминает гиперболу. В контексте сказанного ранее это означает, что гипотеза о параболической форме верхней границы множества технологически достижимых состояний имеет право на существование. Более того, она гораздо ближе к реальности, чем многие предположения, традиционно используемые экономистами при построении моделей, используемых для обоснования вполне конкретных регуляторных мер.

Если граница множества технологически достижимых продуктовых наборов (из двух продуктов) описывается параболой, то и систему цен надо строить аналогичным образом, то есть так, чтобы множества имеющих одну и ту же цену продуктовых наборов описывались параболой. Иначе говоря, разделять множества более предпочтительных, чем имеющийся, и технически достижимых продуктовых наборов следует параболой, проходящими через точку касания таких множеств. То же самое касается и разделения множеств другой размерности. Техника построения поверхностей нужной формы разработана в рамках приложения тропической математики к алгебраической геометрии [Литвинов, 2014].

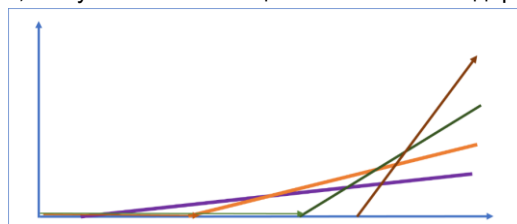


Рисунок 2. Огибающая возможных вариантов вложения в технологии похожа на параболу

4. Оптимизация соотношения выплат

Следуя принципу, заявленному в самом начале статьи, будем искать подходящую модель двух-компонентных цен в предельно простом виде, используя параметрические функции двух переменных с минимально возможным набором управляющих параметров.

4.1. Оптимальные двухкомпонентные цены в экономике Робинзона Крузо

Как и в случае с фирмой Робинзона Крузо по производству и продаже кокосов, в рассматриваемой экономике есть один невозпроизводимый ресурс в ограниченном объеме γ и один производимый продукт, количеству которого соответствует переменная x . Для потребления этот продукт приобретается по ценам или, точнее, по тарифам, которые задает агент, играющий роль «невидимой руки». В нашем случае эту роль играет все тот же Робинзон, хотя могут быть и иные версии.

В распоряжении агента, играющего роль «невидимой руки», имеется семейство тарифов, описываемых двумя числовыми функциями f и g одной переменной t . Здесь t играет роль идентификатора тарифа, функция $f(t)$ задает цену поставляемого продукта при заданном тарифе t , а $g(t)$ – плату за подключение (вход). Повышением тарифа далее именуется повышение t , а не $f(t)$. Естественно предположить, что значение функции $f(t)$ как-то плавно убывает с ростом t , а значение $g(t)$ возрастает. Плата за поставку производимого продукта в объеме x при заданном t составляет

$$f(t)x + g(t).$$

Также предполагается, что роль денег играет невозпроизводимый ресурс, имеющийся изначально в распоряжении потребителя в количестве c . Если потребитель приобретает производимый продукт в количестве x , то в его распоряжении остается

$$y = \gamma - f(t)x + g(t)$$

невозпроизводимого ресурса, то есть он потребляет набор $(x, y) = (x, \gamma - f(t)x + g(t))$.

Логично предположить, что потребитель попытается оптимизировать плату за получение продукта (или услуги) и выбирает не только объем потребления \bar{x} , но и тариф \bar{t} из условия

$$\bar{t} = \arg \min_{t \geq 0} [f(t)x + g(t)].$$

Также логично предположить, что функции f и g выбираются из набора каких-то достаточно простых классов функций. Если функции f и g дифференцируемы, имеют производные \dot{f} и \dot{g} , соответственно, а решение уравнения

$$\dot{f}(t)x + \dot{g}(t) = 0$$

существует и единственно, то можно решить это уравнение и исключить переменную t . В результате получим функцию минимальных возможных затрат ресурса на получение производимого продукта.

Вариант

$$\bar{t} = \arg \min_{t \geq 0} [f(t)x + g(t)] = 0,$$

разумеется, тоже возможен, но не представляет интереса. Он соответствует случаю, когда в технологию или высокий тариф вкладывать бессмысленно. Интересно рассмотреть примеры, когда смысл в таких вложениях есть.

С учетом сказанного выше о стремлении к использованию простых параметрических функций с минимальным числом параметров, можно упростить выражение для суммы затрат на получение производимого продукта в объеме x в виде

$$h(x, t) = \alpha t^{-\beta} x + t,$$

где t – начальные вложения (фиксированный платеж), $\alpha t^{-\beta} x$ – прямые затраты на производство (покупку) продукта. Выбираем t из условия

$$t = \arg \min_{t \geq 0} [\alpha t^{-\beta} x + t].$$

Делаем замену переменной t и получаем функцию с понижающимися затратами на масштаб

$$\hat{h}(x) = h\left(x, (x\alpha\beta)^{\frac{1}{1+\beta}}\right) = (\alpha\beta x)^{\frac{1}{\beta+1}} \left(\frac{\beta+1}{\beta}\right),$$

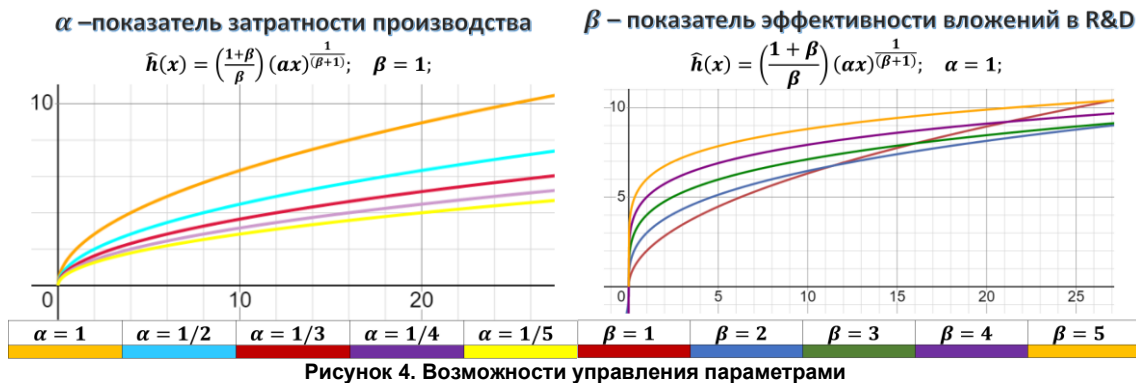
зависящую от x и параметров α, β , но не зависящую от t . Такие функции подходят для описания и тарифов, и затрат на R&D! Меняется лишь экономическая интерпретация переменных.

Невыпуклость множества достижимых состояний или, как вариант, бюджетного множества проще всего показать, принимая $\alpha = 1$, $\beta = 1$ и $\gamma = 1$. Остаток невозпроизводимого продукта или, соответственно, остаток денег составит $y = 1 - 2\sqrt{x}$. Бюджетное множество (см. рисунок 3) ограничено сверху фрагментом параболы с концами $(0, 1)$ – полное отсутствие производства, и $(1/4, 0)$ – на производство тратятся все ресурсы, что соответствует затратам на производство всех имеющихся ресурсов.



Рисунок 3.
Бюджетное множество

Чтобы показать возможности управления тарифами с помощью двух выбранных параметров, рассмотрим варианты, когда один из них фиксирован, а другой меняется. В левой части рисунка 4 фиксирован $\beta = 1$, параметр α меняется. При этом кривые единообразны. В правой его части меняется β , тогда как α остается равным единице. При увеличении β кривая затрат изгибается все круче, а в пределе получается прямой угол. Такое положение соответствует либо фиксированной плате за безлимитный тариф, если речь о тарифе, либо о производстве с затратами только на изготовление цифрового продукта и нулевыми затратами на его тиражирование.



Параметр α играет чисто техническую роль, связывая единицы измерения по оси абсцисс и оси ординат. Роль параметра β более содержательна, его значение определяет соотношение между начальными затратами «за вход» и затратами непосредственно за потребление производимого продукта.

Можно показать возможность отделения множества более предпочтительных наборов благ от множества доступных при тех же начальных запасах и технологии. Переменную y лучше интерпретировать

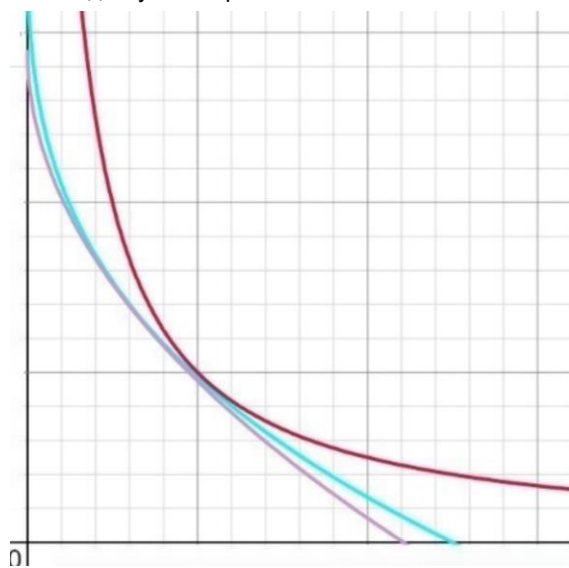


Рисунок 5. Отделимость невыпуклого множества возможностей от выпуклого множества хотелок

как остаток денежных средств у потребителя после приобретения производимого блага в объеме x по оптимальному тарифу при заданных α и β . Чтобы показать отделимость на рисунке, удобно предположить, что речь идет о разделении названных множеств в точке их касания, меняя α , β и γ .

Тут возникает достаточно деликатный вопрос в части интерпретации. Сумма $\left(\frac{1+\beta}{\beta}\right)(\alpha x)^{\frac{1}{\beta+1}} + y$, совпадает с γ , если смотреть на ситуацию с позиций покупателя, но не с позиций продавца. С выручкой продавца от продажи того же объема блага по тому же тарифу эта сумма совпадает только в точке касания разделяемых множеств. То же самое имеет место и в исходной ситуации у Робинзона Крузо с кокосами. Но там это ни у кого не вызывает настороженности, поскольку привычно.

Меняя α , β и γ , можно отделить множество более желательных состояний (надграфик темно розовой гиперболы) от множества технологически достижимых состояний (подграфик фиолетовой параболы) на рисунке 5. Множество достижимых состояний находится внутри бюджетного множества (подграфик голубой параболы). Здесь параметр α можно исключить, если предположить, что все три кривые, включая кривую безразличия (розовая гипербола) смещаются при изменении α , как и кривые в левой части рисунка 4.

4.2. Множественность покупателей

Напомним, что функция $\hat{h}(x)$ была получена заменой t на ее выражение через соответствующее значение переменной x . Аналогично, заменяя x его выражением через t в формуле

$$h(x, t) = \alpha t^{-\beta} x + t,$$

получим

$$\check{h}(t) = \alpha t^{-\beta} \alpha^{-1} \beta^{-1} t^{(1+\beta)} + t = \beta^{-1} t + t.$$

Примечательно, что правая часть равенства не зависит не только от x , но и от α . Интерпретировать это можно так, как будто потребитель производимого блага выбирает не его объем в каких-то натуральных показателях, а сумму, которую он готов потратить на это, включая плату «за вход», в размере t и непосредственно за потребление в размере $\beta^{-1}t$. Иначе говоря, он выбирает бюджет и

тариф, соотнося их с величиной αx . Параметр α можно считать заданным заранее, в том числе, можно считать его равным единице. Логика это, строго говоря, не нарушит. Можно считать, что тариф задается не параметрами, а соотношением начального и последующего платежа.

Рассмотрим теперь ситуацию с одним продавцом и произвольным количеством потребителей равным n . Пусть x_i – объем потребления покупателем i , где $i = 1, n$. Общий объем потребления равен

$$x = \sum_{i=1}^n x_i.$$

Каждый покупатель i выбирает оптимальный для себя тариф t_i из условия

$$t_i = \operatorname{argmin}_{t \geq 0} [\alpha t^{-\beta} x_i + t].$$

Тогда из полученного выше имеем

$$t_i = (x_i \alpha \beta)^{1/(1+\beta)}$$

$$x_i = \alpha^{-1} \beta^{-1} t_i^{(1+\beta)}$$

для каждого $i = 1, n$. При этом полный платеж для i будет равен $t_i + t_i/\beta$, а суммарный платеж составит

$$t + \frac{t}{\beta} = \sum_{i=1}^n \left(t_i + \frac{t_i}{\beta} \right),$$

где t – сумма всех t_i . При этом тариф t оптимален для совокупного потребления x .

Наш двухкомпонентный тариф оказывается дифференцированным, то есть разным для разных покупателей. Но удовлетворяет условию анонимности, то есть зависит не от покупателя или его номера, а лишь от его собственного выбора, который он делает одновременно с выбором объема приобретаемого блага (продукта или услуги). При этом тариф не навязывается. В этом смысле какой-либо ценовой дискриминации здесь нет. Есть одинаковый для всех набор тарифов, из которого каждый покупатель выбирает свой индивидуальный тариф вместе с предполагаемым объемом потребления.

Интересно рассмотреть случай, получаемый при предельном переходе, когда $\beta \rightarrow \infty$. Как уже отмечалось ранее, в пределе вся сумма за потребление блага взимается в виде начального платежа, а непосредственное потребление не оплачивается. В случае, когда речь идет о подключении к сети, это означает безлимитный тариф.

Вот тут и возникает любопытный поворот. При подключении с ограниченным объемом потребления, который потребитель может для себя примерно спрогнозировать и выбрать оптимальный для себя же тариф, разные потребители выбирают разные тарифы. Чем меньше прогнозируемое потребление, тем меньше бюджет потребителя и ниже плата «за вход». Плата непосредственно за потребление пропорциональна плате за вход. Так устроен тариф для всех. Но плата за единицу потребления за всех одинакова, в пределе при $\beta \rightarrow \infty$ она нулевая, но при конечных β .

Интересно взглянуть на эту ситуацию с точки зрения теории коллективных благ, то есть благ, для которых отсутствует конкуренция в потреблении [Samuelson, 1954]. Потребление одним экономическим агентом не мешает потреблению другими агентами. Именно такими благами считаются объекты городской инфраструктуры, телефонные сети, интернет и многое другое. Строго говоря, отсутствие конкуренции в потреблении имеет место ровно до тех пор, пока сеть или городская инфраструктура не перегружена. Именно по этой причине тут надо говорить о коллективных, а не о публичных или общественных благах. Для публичных благ типа математических теорем или законов физики конкуренция в потреблении отсутствует в принципе. А вот плоды прикладной науки, включая технологии и цифровые продукты, обычно становятся публичными благами, поставляемыми в частном порядке.

Именно тут и возникает тот самый вопрос о дифференциации цен и ценовых алгоритмах, о котором говорилось в самом начале статьи. Интуитивно кажется, что цены за доступ к такому благу должны быть равными для всех, как и цены на любые другие блага, или, в крайнем случае, могут быть какие-то льготы для определенных категорий потребителей. Но математика здесь безжалостна, оптимальные цены индивидуальны для каждого потребителя. А это значит, что запрет на дифференциацию цен может сделать нерентабельными вложения в создание и развитие таких продуктов.

Но дифференциация цен в зависимости от покупателя технически сложна, она требует либо достаточно точного знания того, сколько конкретный потребитель готов платить, либо такой организации процесса торговли, когда он сам выберет приемлемую для себя цену. А это уже ценовые алгоритмы, которые антимонопольные ведомства стараются искоренить.

4.3. Двухкомпонентные цены в лицензионной торговле

Как уже говорилось выше (в 2.3.), двухкомпонентному ценообразованию при возрастающей отдаче на масштаб посвящена обширная литература, в том числе целая глава в книге [Villar, 1996], разделы в продвинутых учебниках [Varian, 1984, 1992] и множество статей. Тем не менее, тема не исчерпана. В имеющейся литературе по теме эта идея увязывается либо с внешним регулированием через налоги и субсидии, либо с фиксированной платой сначала просто «за вход», а потом – за конкретные услуги и продукты, но уже по предельным ценам. Плата за вход, как правило, фиксированная. На эту тему см.: классическую статью Уолтера Ои [Oi, 1971i] под названием “A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly”.

В данной работе предлагается иной подход к данной теме. Он основан на идеях оптимизации, изложенных выше, но требует подкрепления фактами, заимствованными из практики торговли лицензиями на использование изобретений и ноу-хау. Двухкомпонентное ценообразование широко применяется в лицензионной торговле технологиями и ноу-хау, но смысл компонентов цены здесь иной, чем в случае посещения Диснейленда или подключения к телефонной сети, хотя общее можно найти и здесь. Во всех случаях первоначальный платеж можно рассматривать как приобретение реального опциона. Эта тема сама по себе заслуживает внимания, по ней тоже есть литература.

Если говорить о лицензионной торговле, то можно говорить о целом пучке смыслов и вариантов привязок к тем или иным конкретным действиям. Размеры начального (паушального) платежа здесь изначально увязываются с размерами текущих платежей (роялти), что выражается в виде зависимости ставки роялти от размеров паушального платежа или наоборот (в виде зависимости начального платежа от ставки роялти) и, разумеется, от ожидаемого объема продаж. Такое понимание цен существенно выпадает из стандартной парадигмы, согласно которой рыночная цена содержит всю информацию, необходимую для принятия решений на локальном уровне. Эта парадигма не подходит даже с оговоркой относительно выпуклости или невыпуклости (см. Цитату из [Вэриан, 1997] выше).

Цитируемое рассуждение целиком относится не столько к реальной экономике, сколько к ее модельному представлению, но потом проецируется на реальность. Между тем, реальными могут быть только цены, тогда как кривые или поверхности безразличия существуют лишь на страницах учебников и в воображении тех, кто эти учебники пишет или читает. Именно цена как совокупность условий заключаемой сделки наблюдаема обеими сторонами этой сделки, а иногда и сторонними наблюдателями. О технологии и предпочтениях у сторон сделки есть лишь их собственные представления, причем эти представления разные у разных сторон и в разной степени фрагментарны. В принципе этот факт никогда не был секретом, но его принято было либо игнорировать при построении математических моделей, как это делается в теории общего равновесия, либо вводить в модель явно, как это делается в теории игр с асимметричной информацией. В реальной экономике ни то, ни другое не ведет к успеху, в чем легко убедиться на практике, если рискнуть. Но вернемся к лицензиям.

Приобретение лицензии на использование технологии – достаточно сложная и трудоемкая процедура, в которой с каждой из сторон участвует команда разных специалистов. На эту тему существует специальная литература. Но сейчас речь не об этом, а о том, что покупка лицензии может рассматриваться как альтернатива вложениям в развитие собственных технологий. А потому представление верхней границы множества технологически достижимых состояний, бюджетного множества при покупке лицензий, а также суммы выплат по двухкомпонентному тарифу в виде фрагмента параболы имеют под собой некое основание. Примечательно, что этот вывод никоим образом не следует из экономической теории и не может из нее следовать. Он следует из частных, но очень конкретных фактов, а именно, из особенностей учета затрат на исследования и развитие в отдельных странах, правила 25% при расчете ставок роялти в лицензионных договорах и его аналитического обоснования.

Если рассматривать варианты вложений компании в развитие технологий, то как основной вариант можно рассматривать вложения в собственные исследования и разработки, а в качестве альтернативного варианта – покупку патентов и лицензий на использование сторонних разработок. При этом следует иметь в виду, что покупка патентов – относительно редкое явление в сравнении с покупкой патентных лицензий, а вложения в собственные разработки не исключают покупку лицензий, как патентных, так и беспатентных. Теоретически, а отчасти и практически эти варианты должны быть одинаково выгодны. На этом и строится аналитическое обоснование правила 25%.

Выбор между первым и вторым вариантом, как и все в реальной экономике, зависит от множества причин. Но все же есть основания считать, что в среднем эти варианты одинаково выгодны. На этом соображении построено хорошо известное «правило 25%» для расчета ставки роялти в лицензионном договоре на производство какого-то продукта, согласно которому ставка роялти должна составлять примерно 25% от EBIT² компании чистой игры, производящей этот и только этот продукт по лицензии. Примечательно, что такое простое правило очень хорошо согласуется с реальными данными по американской промышленности. Средние отраслевые ставки роялти, исчисляемые в процентах от валовой выручки, при пересчете в проценты от EBIT близки к 25%, если брать показатели из финансовой отчетности компаний, предоставляемой SEC³. Это и значит, что они близки к ставкам, рассчитанным по «правилу 25%».

Здесь надо сделать еще одну оговорку относительно правил учета затрат на исследования разработки (R&D) в США. Компаниям разрешается самостоятельно формировать учетную политику в этой части, ее надо просто декларировать и не менять под каждый конкретный случай. Это позволяет распределять затраты на R&D равномерно по годам и так же равномерно вычитать их из налогооблагаемой прибыли. По этой причине показатели из финансовой отчетности, предоставляемой SEC, имеют именно тот смысл, о котором говорилось выше. Они усредняются, а потому нет необходимости учитывать распределенность затрат на разработку и отдачи от них по годам. Разумеется, это совсем не так в компании «чистой игры», использующей одну технологию для производства одного продукта. Но и са-

² Earnings before interests and taxes.

³ SECURITIES AND EXCHANGE COMMISSION – Комиссия по ценным бумагам и биржам США

мо понятие «компания чистой игры» – довольно искусственная конструкция, на практике такую «компанию» приходится выделять из бизнеса реальной компании или, наоборот, достраивать с применением специальных искусственных приемов. Необходимо это, в том числе, для корректного применения «правила 25%» в реальных лицензионных соглашениях, а также при оценке патентов.

Решая конкретные технические проблемы в рамках общей задачи, фирма делает выбор между разработкой собственного решения или покупкой лицензии. Если же сделан выбор в пользу лицензии, то возникает вопрос о том, что конкретно и у кого покупать, а за ним – вопрос о цене и эффективности покупки. Важно подчеркнуть, что обычно в таких случаях имеется широкий спектр возможностей, причем все они схожи в том, что сначала предполагают затраты без видимой отдачи, потом линейную или близкую к линейной зависимость отдачи от затрат. Большим затратам вначале соответствуют более низкие текущие затраты потом. Доступ к более совершенной технологии обходится дороже, но компенсируется большей эффективностью, если объем ее использования достаточно велик. Графически это можно изобразить, как на рисунке 2. По оси абсцисс откладываются затраты, а по оси ординат – выпуск продукта по используемой технологии.

5. Заключение

Подводя итоги сказанному выше, необходимо отметить, что в рамках данной статьи не нашлось места для обоснования тезиса о ценах как о наблюдаемых параметрах, тогда как описание потребительских предпочтений – фантазии на основе очень фрагментарного знания, а описание технологических возможностей – компромисс между этими двумя крайностями, то есть между реальным знанием и прогнозом на базе очень фрагментарных знаний. Об описании технологических возможностей кое-что все же было сказано выше. К этому можно добавить некоторые штрихи относительно описания предпочтений. Способ описания предпочтений в учебниках по экономикс обладает лишь тем достоинством, что он удобен. Особенно удобно описание с применением функций полезности, причем не любых, а таких, которые легко дифференцировать, не заботясь ни о граничных условиях, ни о чем-либо еще.

В реальной экономике функций полезности, разумеется, нет. Нормальный потребитель рассматривает очень ограниченный набор возможностей на очень ограниченном интервале времени. Ему нет дела до обилия блюд из жареных летучих мышей на рынке в Ухане или до цен на жилье в стране, где он не собирается жить. Он выбирает из набора конкретных вариантов, доступных для него в обозримом будущем. А потому возникают задачи по моделированию спроса в таких условиях и разделению с помощью цен таких множеств, где есть элементы дискретности. Таким задачам посвящены работы [Baldwin and Klempereger, 2012, 2015], где множества доступных и более желаемых состояний разделяются с применением поверхностей, для построения которых используется тропическая математика. Там есть ссылки на работы [Danilov and Koshevoy, 2004] и [Danilov, Koshevoy and Murota, 2001], где аналогичные результаты получены раньше другими методами. Но и тропическая математика выросла из идемпотентного анализа, развитого нашими соотечественниками [Маслов, Колокольцев, 1994], а потом успешно применялась для построения производственных функций и для моделирования развития агломераций [Matveenko, 2011, 2012]. К сожалению, эти работы не очень доступны экономистам, поскольку используемая в них математика достаточно сложна. Но и это не все. Формально в современной экономикс используются сколь угодно сложные математические конструкции, но делается это, как правило, не для преодоления реальных сложностей, а в качестве своеобразного упражнения, убеждая тем самым здраво мыслящих экономистов в бесполезности математики. Именно это обстоятельство и послужило причиной того, что в данной статье, а еще раньше в [Kozyrev, 1987], используется математика уровня [Varian, 1984, 1992; Вэриан, 1997]. В свое время сам Вэриан определил этот уровень как «промежуточный», то есть достаточный для того, чтобы показать что-то важное для понимания, но не выйти за рамки доступного для студентов экономистов.

Как уже говорилось выше, плодотворные экономические идеи рождаются чаще всего либо в бизнесе, либо в результате анализа математических конструкций, а потом осваиваются экономической теорией. Одна из таких идей – использование операции максимума вместо обычного сложения в различных математических конструкциях, включая разные варианты уравнения баланса [Макаров, 1973], но не только, применимость этой идеи много шире [Козырев, 1999, 2011]. Сфера и способы применения таких конструкций поразительно многообразны. Они появляются в работах математиков, интересующихся экономикой или, точнее, приложением своих идей в экономике, но не воспринимались экономистами всерьез. Существенно лучше экономистами понимается идея двойственности или, точнее, идея построения системы цен на основе двойственных переменных, поскольку она хорошо коррелируется с идеей предельных или маржинальных цен, которую можно показать на примере замещения малой доли одного продукта малой долей другого продукта. В пределе получается именно то, что нужно.

Для принятия практических решений, как выясняется, гораздо полезнее осмысление реального опыта, возможно, с оглядкой на теорию и ее парадигмы, но никак не слепое следование им. То же самое иногда уместно делать применительно к теории, а именно, менять парадигму, опираясь на осмысление опыта и математику. Такой разворот от бизнеса прямо к математике, минуя остановку в пункте «экономическая теория», стал особенно актуальным благодаря развитию информационных технологий и сетевых сервисов. Сети собирают огромный объем информации не только о возможных поставщиках продукции, но и о покупателях. К тому же они обладают реальными возможностями для управления

вниманием покупателей, а потому о «невидимой руке» рынка можно просто забыть или, как минимум, существенно переосмыслить этот образ. К тому же уместно напомнить, что в лицензионной торговле во все времена такая «невидимая рука» если и действовала, то где-то за рамками любой конкретной сделки. Обсуждая условия таких сделок, стороны всегда старались прогнозировать будущее и обсуждали детали. А потому накопленный в этой области опыт заслуживает самого пристального внимания, как, впрочем, и некоторые новые разделы математики [Маслов, Колокольцев, 1994; Литвинов, 2005].

Литература

1. Вэриан Х. Р. (1997) Микроэкономика. Промежуточный уровень. Современный подход. Учебник для вузов/Пер. с англ. под ред. Н.Л. Фроловой. — М.: ЮНИТИ, 1997. — 767 с.
2. Козырев А.Н. Общее равновесие в экономике с рынками лицензий и продуктов. Тезисы докладов Всесоюзной школы-семинара "Социально-экономические процессы", Кишинёв, 1989. С. 163—164.
3. Козырев А.Н. Алгебраические свойства информации и рынок. Научно-техническая информация. сер.1, 1999, № 5. С. 15—20
4. Козырев А.Н. Модель общего равновесия с эндогенным технологическим прогрессом и двух-этапным ценообразованием Препринт ЦЭМИ РАН WP/2002/0133. — 2002, 58 с.
5. Козырев А.Н. Моделирование НТП, упорядоченность и цифровая экономика// Экономика и математические методы, т. 47, № 4, 2011 г.
6. Литвинов Г. Л. (2005) Деквантование Маслова, идемпотентная и тропическая математика: краткое введение Зап. научн. сем. ПОМИ, 2005, том 326, страницы 145—182 (Mi zns1341)
7. Макаров В.Л., (1973) Баланс научных разработок и алгоритм его решения // Сб.ст. Оптимизация, Новосибирск, 1973, вып.11(28), С.37 — 45
8. Маслов В.П., Колокольцев, (1994) Идемпотентный анализ и его применение в оптимальном управлении. — М.: Физматлит, 1994. — 146 с.
9. Basu, K. (2020) Working Paper New technology and increasing returns: The end of the antitrust century? IZA Policy Paper, No. 146 2019, 2021
10. Basu, K. (2021) New Technology, Increasing Returns, and the End of the Antitrust Century From the book «Law, Economics, and Conflict», 2021 Cornell University Press <https://doi.org/10.1515/9781501759284-007>
11. Baldwin E. and Klemperer P. (2012) Tropical geometry to analyse demand. Technical report, Working paper, Oxford University, 2012
12. Baldwin E. and Klemperer P. (2015) Understanding preferences: "demand types", and the existence of equilibrium with indivisibilities. Technical report, Working paper, The London School of Economics and Political Science, 2015
13. Coase, R., (1937) The Nature of the Firm, *Economica*, Vol. 4, pp. 386-405
14. Coase, R., (1946) The Marginal Cost Controversy. 13 *Economica* 169.169 (1946). The Marginal Cost Controversy," *Economica*, August 1946, 13, (n.s.), 169-82.
15. Coase, R., (1960) The Problem of Social Cost, *Journal of Law and Economics*, October 1960, 3, 1-44.
16. Coase, R., (1974) The Lighthouse in Economics, *Journal of Law and Economics* 17(2):185—213.
17. Danilov V. and Koshevoy, G. Discrete convexity and unimodularity I. *Advances in Mathematics*, 189(2):301{324, 2004.
18. V. Danilov, G. Koshevoy, and K. Murota. Discrete convexity and equilibria in economies with indivisible goods and money. *Mathematical Social Sciences*, 41:251{273, 2001.
19. Hotelling, H. (1938). "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." *Econometrica*, 6(3), 242-69
20. Hotelling, H. (1990). The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates. In: Darnell, A.C. (eds) *The Collected Economics Articles of Harold Hotelling*. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-8905-7_10
21. Keen, S. (2011). *Debunking Economics: The Naked Emperor Dethroned?* ft. Zed Books. ISBN 978-1848139923
22. Kim, J. (1997). "Three Sources of Increasing Returns to Scale." Finance and Economics Discussion Series 1997-18, Federal Reserve Board.
23. Vladimir Matveenko (2014) Tropical support sets in analysis of weak links and complementarity *Contemporary Mathematics Volume 616*, 2014 <http://dx.doi.org/10.1090/conm/616/12315>
24. Matveenko, V. (2011) Anatomy of production function: a technological menu and a choice of the best technology, *Economics Bulletin*, vol. 30, 2011, pp. 1906-1913.
25. Matveenko, V., (2012) Powers of matrices with an idempotent operation and an application to dynamics of spatial agglomerations, In: *TROPICAL AND IDEMPOTENT MATHEMATICS*, G. L. Litvinov, V. P. Maslov, A. G. Kushner, S. N. Sergeev (Eds.), Moscow — 2012, P. 149-155.
26. Oi, W. Y. (1971). "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *The*
27. Posner R. A., (1993) Nobel Laureate: Ronald Coase and Methodology. 7 *J Econ Persp* 195,197 (1993).

28. Samuelson P. A. (1954). "The Pure Theory of Public Expenditure Paul A. Samuelson The Review of Economics and Statistics, Vol. 36, No. 4. (Nov., 1954), pp. 387–389.
<http://links.jstor.org/sici?sici=0034-6535%28195411%2936%3A4%3C387%3ATPTOPE%3E2.0.CO%3B2-A>
29. Stomper, R. (2015) The Marginal Cost Controversy Revisited (May 4, 2015). Available at SSRN: <https://ssrn.com/abstract=2602548> or <http://dx.doi.org/10.2139/ssrn.2602548>
30. Varian, H. R. (1984) Microeconomic Analysis, New York: W. W. Norton.
31. Varian, H. R. (1992) Microeconomic Analysis, new edn, New York: W. W. Norton Quarterly Journal of Economics, Oxford University Press, vol. 85(1), pages 77-96.
32. Williamson O. (1985) The economic institutions of capitalism: firms, markets, relational contracting New York : Free Press ; London : Collier Macmillan

References in Cyrillics

1. Ve`rian X. R. (1997) Mikroe`konomika. Promezhutochny`j uroven`. Sovremenny`j podxod. Ucheb-nik dlya vuzov/Per. s angl, pod red. N.L. Frolovoj. — M.: YuNITI, 1997. — 767 s.
2. Kozy`rev A.N. Obshhee ravnovesie v e`konomie s ry`nkami licenzij i produktov. Tezisy` do-kladov Vsesoyuznoj shkoly`-seminara "Social`no-e`konomicheskie processy", Kishinyov, 1989. S. 163—164.
3. Kozy`rev A.N. Algebraicheskie svoystva informacii i ry`nok. Nauchno-texnicheskaya infor-maciya. ser.1, 1999, № 5. S. 15–20
4. Kozy`rev A.N. Model` obshhego ravnovesiya s e`ndogenny`m texnologicheskim progressom i dvuxe`tapny`m cenoobrazovaniem Preprint CzE`MI RAN WP/2002/0133. — 2002, 58 s.
5. Kozy`rev A.N. Modelirovanie NTP, uporyadochennost` i cifrovaya e`konomika// E`konomika i ma-tematicheskie metody`, t. 47, № 4, 2011 g.
6. Litvinov G. L. (2005) Dekvantovanie Maslova, idempotentnaya i tropicheskaya matematika: kratkoe vvedenie Zap. nauchn. sem. POMI, 2005, tom 326, stranicy 145–182 (Mi znsI341)
7. Makarov V.L., (1973) Balans nauchny`x razrabotok i algoritm ego resheniya // Sb.st. Optimi-zaciya, Novosibirsk, 1973, vy`p.11(28), S.37 – 45
8. Maslov V.P., Kolokol`cev, (1994) Idempotentny`j analiz i ego primenenie v optimal`nom upravlenii. — M.: Fizmatlit, 1994. — 146 s.

Козырев Анатолий Николаевич, к.ф.-м.н., д.э.н.

Центральный экономико-математический институт РАН

ORCID 0000-0003-3879-5745,

kozyrevan@yandex.ru

Ключевые слова

Равновесие, возрастающая отдача от масштаба, предельные цены, двухкомпонентные цены

Anatoly Kozyrev, Optimal two-component prices in economies with increasing returns

Keywords

Equilibrium, increasing returns to scale, marginal prices, two-component prices,

DOI: 10.34706/DE-2023-01-07

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract

The possibilities of constructing and applying optimal two-component prices in economies with increasing returns to scale are shown. Optimality is understood as the optimal ratio between the components of the price in the form of a one-time payment and the price per unit of product. The necessity of softening the anti-monopoly policy regarding the use of price discrimination and price algorithms is substantiated. As the main interpretation of the price components, lump-sum (one-time) payment and royalties in the license agreement are considered (after bringing them to a single form, that is, to payments in cash). The interpretation is also considered with the choice of a tariff for connecting to the network and directly for traffic, where the traffic fee is lower, the higher the connection fee.

1.8. АРХИВ КАК ФАКТОР ЦИФРОВОЙ ТРАНСФОРМАЦИИ ОРГАНИЗАЦИИ

Попов Д.В.¹, Ральникова К.В.¹

¹Московский государственный технологический университет «СТАНКИН», Москва, Россия.

В статье рассматривается вопрос полной цифровой трансформации организации на основе формирования контуров обмена между архивом организации и муниципальным архивом на примере взаимодействия федерального государственного бюджетного образовательного учреждения высшего образования, находящегося на территории г. Москвы, и Главного архивного управления города Москвы на базе нормативных актов Федерального архивного агентства. Разработан классификатор оценки уровня цифровой трансформации организации на основе этапов цифровой трансформации и уровня внедрения электронного документооборота.

Введение

Начиная с конца 2000-х годов и по настоящее время разработка, внедрение и использование электронного документооборота в организациях является актуальным вопросом. В условиях быстроизменяющейся рыночной среды перед малым бизнесом, крупным бизнесом и государственными предприятиями и организациями стоят задачи по снижению издержек обеспечивающих процессов, к которым относится и документооборот. Опубликовано большое количество рекомендаций и экспертных мнений о преимуществах электронного документооборота в организациях, где авторы аргументированно обосновывают положительные стороны от внедрения СЭД в организациях. Так, например, со слов Русскевича С.Н. следует, что эффективное внедрение СЭД – важный элемент создания виртуальной рабочей среды и преобразования возможностей современной организации и ее персонала [Русскевич, Медникова, 2022]. Алифинова А.М. и Васильев В.П. считают, что для решения трудностей и проблем, возникающих из-за неправильного ведения бумажного документооборота, следует ввести СЭД и что его внедрение выведет на новейший уровень бухгалтерские услуги [Алифинова, Васильев, 2016].

Анализ рынка программного обеспечения в части электронного документооборота

Анализ экспертных мнений позволяет сделать вывод о том, что к ключевым преимуществам использования СЭД следует относить:

- Значительное уменьшение бумажного документооборота.
- Сокращение времени на обработку документов.
- Возможность удаленной работы с документами.
- Прозрачность процессов (видеть ход и статус каждого бизнес-процесса).
- Упрощение работы по контролю исполнения на протяжении всего жизненного цикла документа.
- Удобство хранения и эффективно организованная система поиска документа.
- Возможность интеграции с необходимыми информационными системами.
- Экономия средств на обустройство помещений, закупку бумаги и других канцтоваров.
- Оптимизация штатной численности сотрудников службы ДОУ.

Наряду с заявленными положительными эффектами от внедрения СЭД в организации, на российском рынке представлен большой ассортимент программных решений для организации документооборота, которые включены в Единый реестр российских программ для электронных вычислительных машин и баз данных, созданный в соответствии со статьей 12.1 Федерального закона «Об информации, информационных технологиях и о защите информации»¹.

В результате не остается никаких сомнений, что внедрение СЭД в организации приносит ощутимую пользу и решает ряд задач по оптимизации и сокращению издержек. Широта ассортимента программных решений на рынке и пособий для пользования позволяет осуществить внедрение прямо здесь и сейчас, ограничиваясь только финансовыми возможностями организации в зависимости от широты функционала программы.

¹ Официальный сайт оператора единого реестра российских программ для электронных вычислительных машин и баз данных в информационно-телекоммуникационной сети «Интернет» URL: <https://reestr.digital.gov.ru/> (дата обращения: 10.06.2022)

Таблица 1. Наиболее популярное программное обеспечение

Программное обеспечение	Примеры организаций клиентов, использующих программные решения
1С: Документооборот 8	ООО «МСТ Групп», РОССИЙСКОЕ ОБЩЕСТВО «ЗНАНИЕ», Министерство промышленности и торговли Российской Федерации и др. ²
Система Directum	Bonduelle, ВР, Аскона, Газпром ПАО, ПАО Роснефть, Финансовый Университет при правительстве РФ и др. ³
Docsvision Делопроизводство	Сбербанк РФ, АЛРОСА, предприятия «Роснефти» и др. ⁴
Контур.Диадок	Asus, Google AdWords, Hyundai, KIA, S7, Альфа-Банк, Ашан, ВКОНТАКТЕ, Лента, Л'ОРЕАЛЬ, Магнит, Нефть Магистраль, НПК, НТВ-ПЛЮС, ПАО «Группа Компаний ПИК», Перекресток, Пятерочка, Ростелеком Москва, Ситилинк, Яндекс.Касса и др. ⁵
Практика	Ленинградская область, Минцифра, Министерство финансов, Минобрнауки, Министерство транспорта, Московский метрополитен, Московская область, Правительство Москвы, Пенсионный фонд РФ, Росавиация, Роспотребнадзор, Росприроднадзор, Ространснадзор, Татнефть и др. ⁶
Система автоматизации делопроизводства и электронного документооборота («Дело»)	Центральный Банк Российской Федерации (Банк России), Арбитражные суды, Исполнительный комитет СНГ, ПАО «Мосэнерго», Совет Федерации Федерального Собрания Российской Федерации и др. ⁷
Система электронного документооборота и автоматизации бизнес-процессов «Е1 Евфрат»	ОАО «РЖД Логистика», ООО «Нефтеюганский научно-исследовательский и проектный институт», ООО СК «ВТБ Страхование», Учреждение Российской академии наук Институт системного анализа РАН, Министерство внутренних дел Российской Федерации (МВД России) и др. ⁸
Система управления документами и задачами «ТЕЗИС»	СПАО «Ингосстрах», ПАО «Промсвязьбанк», МИД России, Правительство Челябинской области, ПАО «Почта Банк», Российская академия наук, Nissan Россия, МТС-БАНК, ФГБОУ ВО Новосибирский государственный технический университет, ООО «РГС-Медицина» и др. ⁹

Использование СЭД для учреждений, работающих с Муниципальными архивами и другими органами местного самоуправления, сталкивается с рядом трудностей. В какой-то момент руководители организаций, внедрившие СЭД, обнаруживают, что по разным причинам не могут использовать функционал в полном объеме, например подписывать документы электронной цифровой подписью, хранить документы только в электронном виде без дублирования на бумажном носителе. Соколов Е.А. и Середа С.Н., описывая автоматизацию документооборота в Муромском институте (филиале) Владимирского государственного университета при разработке собственной СЭД, отмечают, что электронные документы, хранимые в системе, представляют собой скан – копии бумажных оригиналов приказов, распоряжений и т.д. [Соколов, Середа, 2012.] На официальном сайте Национального исследовательского Томского государственного университета говорится: «...Работающая система на базе платформы «1С-Битрикс24» обеспечивает автоматическую генерацию приказов и паспортов проектов, а также их последующее согласование в цифровом формате...»¹⁰, но информация про об утверждении и хранении вышеуказанных документов отсутствует. 29 апреля 2022 года в ФГУП «ЦНИИЧермет им. И.П. Бардина» состоялось заседание Комитета по развитию систем управления качеством ассоциации «Лига содействия оборонным предприятиям» на тему: «Импортозамещение, связанное с внедрением автоматизированных систем управления жизненным циклом, цифровизации моделей изделий, двойников процессов, эксплуатационной документации и каталогов запасных частей», на котором Директор по производству АО «НПО «Энергомаш» Марфин В.В., выступил с докладом на тему: «Цифровизация производства. Опыт АО «НПО «Энергомаш». Докладчик в первую очередь делал упор на достижениях его компании по ведению дел в электронном виде¹¹. Во время ответов на вопросы докладчик подчеркнул, что все виды документов дублируются, на бумажных носителях.

² ООО «1С» URL: <https://v8.1c.ru/applied-solutions/?nomenclature%5B0%5D=1938&nomenclature%5B1%5D=1939&branches%5B0%5D=652&city=&partner=&armFrom=&armTo=&dateSince=&dateTill=> (дата обращения: 10.06.2022)

³ Directum URL: <https://www.directum.ru/clients> (дата обращения: 10.06.2022)

⁴ «ДоксВижн» URL: <https://docsvision.com/o-kompanii/> (дата обращения: 10.06.2022)

⁵ СКБ Контур URL: <https://www.diadoc.ru/success> (дата обращения: 10.06.2022)

⁶ СЭД «Практика» URL: <https://sd-praktika.ru/> (дата обращения: 10.06.2022)

⁷ «Электронные Офисные Системы» URL: https://eos.ru/eos_about/eos_clients/ (дата обращения: 10.06.2022)

⁸ Cognitive Technologies URL: <https://evfrat.ru/primery-vnedreniy/> (дата обращения: 10.06.2022)

⁹ Haulmont URL: <https://www.tezis-doc.ru/customers/> (дата обращения: 10.06.2022)

¹⁰ НИ ТГУ URL: <https://news.tsu.ru/news/v-tgu-zapushchena-sistema-po-upravleniyu-proektami/> (дата обращения: 10.06.2022)

¹¹ НИКС – Компьютерный Супермаркет URL: https://www.nix.ru/computer_hardware_news/hardware_news_viewer.html?id=210370 (дата обращения: 10.06.2022)

Построение контуров обмена организации с Главархивом

Как показывает анализ практики, фирмы, организации, предприятия, особенно государственные, которые осуществляют взаимодействие с муниципальными Главархивами, сталкиваются с трудностями при переходе к полному документообороту и получают в итоге затраты на СЭД и удвоение функций в делопроизводстве на электронный и бумажный документооборот.

Это связано с низким уровнем системы управления, когда организации рассматривают документооборот не системно или не в рамках единых подсистем управления, а как самостоятельный выделенный процесс, требующий оптимизации. Систему управления с рассмотрением документооборота как неотъемлемой части подсистемы обеспечения должны рассматривать в рамках цифровой трансформации организации [Попов, Кутикова, 2022]. Таким образом, можно сказать, что процесс перехода должен быть комплексным и основываться в первую очередь на разработке нормативно-правовой базы организации.

Распространённой ошибкой при начале внедрения СЭД в организации является ориентация руководства только на оперативный документооборот, где и хотят произвести оптимизацию издержек, теряя из поля зрения архив организации и правила хранения архивных документов. Правительство Российской Федерации в лице Федерального архивного агентства уже совершенствует нормативно-правовую базу в стране. Так, например, действующим законодательством Российской Федерации, в том числе Федеральным законом Российской Федерации от 22.10.2004 № 125-ФЗ «Об архивном деле в Российской Федерации» и Федеральным законом от 22.11.2021 № 377-ФЗ «О внесении изменений в Трудовой кодекс Российской Федерации», Росархив предусмотрел возможность перевода внутренних документов в электронный вид.

Вместе с тем в государственных организациях возможны ситуации, когда в соответствии с действующей в организации сводной номенклатурой дел на определенный период, утвержденной в Центральной экспертно-проверочной комиссией Главархива, определены сроки и вид хранения документов только на бумажном носителе. Ведение дел в организации, включающее их хранение и уничтожение, строго регламентированы сводной номенклатурой дел, архивным паспортом организации и вышеуказанным законом. За нарушение правил хранения и уничтожения документации Государство предусмотрело персональную ответственность руководителя организации, в соответствии со статьей 13.20 КоАП РФ (штрафы: для организации – от 5 до 10 тыс. руб.). Таким образом, все утвержденные электронные документы организаций, взаимодействующих с Главархивами, должны храниться в соответствии с законодательством Российской Федерации, приказом Федерального архивного агентства от 15.06.2020 № 69 «Об утверждении Типовых функциональных требований к системам электронного документооборота и системам хранения электронных документов в архивах государственных органов» и приказом Федерального архивного агентства от 02.03.2020 № 24 «Об утверждении Правил организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, научных организациях и др.».

Также обстоят дела с кадровым делопроизводством. Например, трудовой договор не получится заключить в электронном виде, если у работника нет личной электронной подписи, с помощью которой он мог бы подписать документ. Кроме того, нужно обратить особое внимание на следующее: Федеральным законом от 22.11.2021 № 377 – ФЗ «О внесении изменений в Трудовой кодекс Российской Федерации» в кодекс добавлены статьи (22.1-22.3), посвященные электронному документообороту. Также следует обращать внимание на даты вступления в силу. В ст. 22.1 указано, что правила этих статей не применяются в отношении: трудовых книжек и формируемых в соответствии с трудовым законодательством в электронном виде сведений о трудовой деятельности работников, акта о несчастном случае на производстве по установленной форме, приказа (распоряжения) об увольнении работника, документов, подтверждающих прохождение работником инструктажей по охране труда, в том числе лично подписываемых работником.

При этом есть ряд документов, у которых срок хранения не постоянный, а, например, до 10 лет. Такие документы организация имеет право перевести полностью в электронный вид, т.к. они не входят в перечень дел, являющихся источником комплектования и имеющих историческую ценность для Главархива, но после переутверждения обновленной номенклатуры дел в Центральной экспертно-проверочной комиссии будет задокументировано, что хранится только в бумажном виде, а что в электронном.

Поэтому Архив организации играет особо важную роль при внедрении СЭД и обеспечении цифровой трансформации организации. На этапе планирования внедрения СЭД в организации должны предусматриваться вопросы хранения электронных документов, а, следовательно, это затронет внутренние документы организации, такие как Инструкция по делопроизводству, Номенклатура дел и др. Также изменятся способы предоставления в Главархив и уничтожения оригиналов документов в соответствии со сроками хранения.



Рис. 1. Взаимодействие Архива организации с Главархивом

На схеме представлены контуры обмена документами с Главархивом, по этим документам организация выстраивает взаимодействие в соответствии с нормативными правовыми актами правительства России и федеральных органов исполнительной власти.

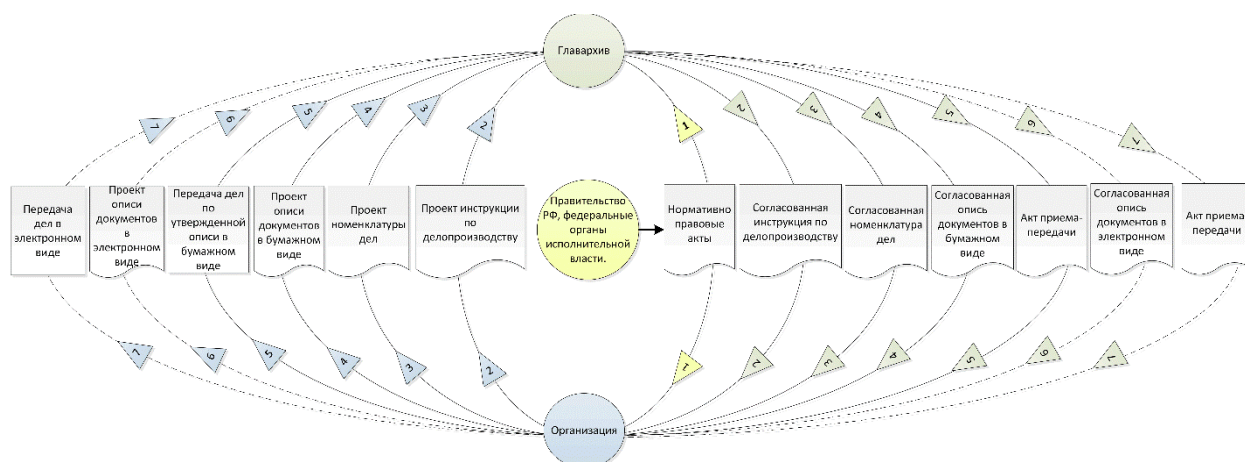


Рис. 2. Контуры взаимодействия Архива организации с Главархивом

Организация, на основе соглашения, заключенного с Главархивом, готовит и подает на утверждение описи дел постоянного хранения (контур № 4) и, согласно утвержденной описи, передает дела в Главархив (контур № 5), а также готовит и подает на согласование следующие документы:

- описи дел по личному составу (контур № 4);
- акты о выделении к уничтожению документов, не подлежащих хранению;
- акты об утрате документов постоянного хранения и по личному составу;
- акты об обнаружении дел, пути розыска которых исчерпаны;
- акты о неисправимых повреждениях документов;
- номенклатуру дел, образующихся в её деятельности (контур № 3);
- инструкции по делопроизводству (контур № 2);
- положения об архиве и экспертной комиссии Организации.

Все документы подготавливаются на основании нормативно-правовых документов и актов Российской Федерации или федеральных органов исполнительной власти (контур №1). Важно отметить, что организации в рамках цифровой трансформации и на этапе подготовки внедрения СЭД должны позаботиться об организации контуров № 6 и № 7 (Рис. 2).

Организация контуров № 6 и № 7 позволит использовать СЭД и электронные документы полностью в цифровом виде. Чтобы измерить процесс внедрения и охват всех документов по номенклатуре

дел и внутренних и внешних процессов организации, был разработан Классификатор цифровой трансформации на основе документооборота.

Описание Классификатора цифровой трансформации

Классификатор позволяет определить, на каком этапе в настоящий момент находится организация, и проследить тенденции развития по 2 критериям:

1. По охвату участников.
2. По уровню интеграции инфокоммуникационных технологий и делопроизводительных процессов.

Под критерием охвата участников подразумевается географические, административные и нормативные границы, в рамках которых подразделение осуществляет свою документационную деятельность.

Под критерием интеграции инфокоммуникационных технологий и делопроизводительных процессов подразумевается уровень интеграции инфокоммуникационных технологий делопроизводства и их использование в период перехода от оцифровки к непрерывному процессу цифровой трансформации, где документооборот выступает в качестве основания для трансформации. Другими словами, данный критерий представляет собой уровень замещения материальных потоков информационными.

Каждый критерий оценивается (измеряется) по 3 бальной шкале:

1. Минимальный уровень освоения/использования.
2. Средний уровень освоения/использования.
3. Высокий уровень освоения/использования.

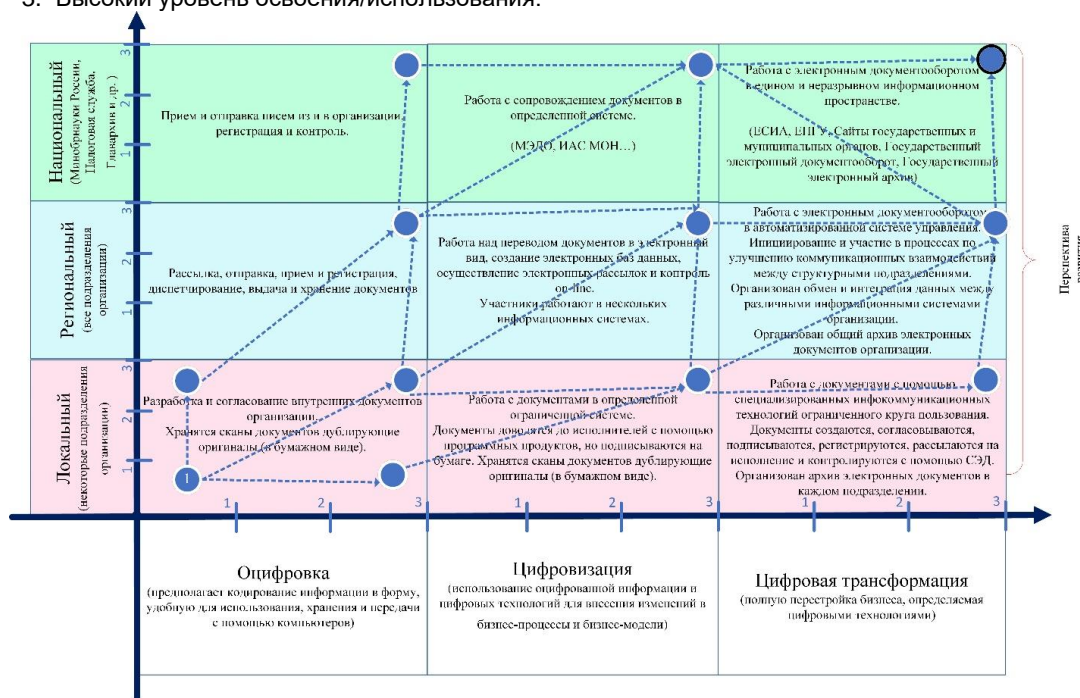


Рис. 3. Классификатор цифровой трансформации на основе документооборота

На схеме изображены тенденции дальнейшего развития. Но с учетом развития современного мира стоит отметить, что на этапе цифровой трансформации должно происходить непрерывное улучшение снова и снова, и организация должна быть готова к этим изменениям.

Возьмем в качестве примера организацию с низким уровнем взаимодействия, которая находится в точке 1. В точке один компания находится на стадии сканирования бумажных документов, т.е. их оцифровки. Ее движение в пределах первого «прямоугольника» может пойти вертикально, горизонтально или же по диагонали. Таким образом, поднимаясь по вертикали, организация увеличивает локальный уровень охвата оцифровки, пока в конечном итоге не будет использоваться полностью в некоторых подразделениях организации. Если двигаться по горизонтали, то будет увеличиваться уровень освоения и использования, пока в конечном счете организация не перейдет к стадии цифровизации. В случае движения по диагонали в целях организации должно быть представлено все вышеуказанное.

Ее движение по схеме также может пойти вертикально, горизонтально или же по диагонали.

Заключение

Для того чтобы перейти с оцифровки в цифровизацию организации необходимо внедрить СЭД, заключив дополнительное сотрудничество с пользователями СЭД и направив оцифрованные документы с помощью СЭД для внесения изменений в бизнес-процессы и бизнес-модели. Не стоит забывать

про увеличение количества пользователей в СЭД. После прохождения локального этапа, можно начать с обязательного использования СЭД у руководителей структурных подразделений, после чего обязать использование его ответственными за делопроизводство в структурных подразделениях, и т.д. Но делать это необходимо постепенно, начиная с порядка выдачи учетных записей в СЭД и заканчивая обучением сотрудников организации работе в СЭД.

При переходе из цифровизации в цифровую трансформацию помимо перестройки всех бизнес-процессов предприятия, организации будет необходимо проработать нормативно-правовое обеспечение данного вопроса, в особенности касательно использования ЭЦП и хранения документов в электронном виде. На основании проработки нормативно-правового обеспечения организация понимает, какие документы она может перевести полностью в электронный вид, а какие нет. После этого и при актуализации номенклатуры дел, организация может приступить к полному переводу документов в электронный вид.

На этапе цифровой трансформации происходит перестройка бизнеса, определяемая цифровыми технологиями [Кудрявцева, Кожина, 2021]. В итоге работа с документами будет осуществляться с помощью специализированных инфокоммуникационных технологий, документы будут создаваться, согласовываться, подписываться, регистрироваться, рассылаться на исполнение и контролироваться с помощью СЭД, также должен быть организован архив электронных документов. На региональном этапе это будет заключаться в интеграции системы СЭД с другими информационными ресурсами организации, использующими ЭЦП. А на национальном уровне будет происходить работа с электронным документооборотом в едином и неразрывном информационном пространстве. Но как именно это будет реализовано, пока сказать сложно.

Литература

1. Алифирова, А.М., Васильев В.П. К вопросу о ведении электронного документооборота в организации // Символ науки. № 6-1. 2016. С.133-135.
2. Кудрявцева Т. Ю., Кожина К. С. Основные понятия цифровизации // Вестник Академии знаний. №3 (44). 2021. С. 149–151.
3. Попов Д. В., Кутикова С. П. Формирование системы управления организации // Техническое творчество молодежи. № 3(133). 2022. С. 38-43. – EDN WTJYYD.
4. Русскевич С.Н., Медникова О.В. Система единого электронного документооборота // Вестник ММА. 2022. С. 89-91.
5. Соколов Е.А., Середа С.Н. Электронный документооборот ВУЗа // Перспективы развития информационных технологий. № 9. 2012. С. 39-44.

References in Cyrillics

1. Alifirova, A.M., Vasilyev V.P. K voprosu o vedenii elektronnoy dokumentooborota v organizatsii // Simvol nauki. № 6-1. 2016. S.133-135.
2. Kudryavtseva T. Yu., Kozhina K. S. Osnovnye ponyatiya tsifrovizatsii // Vestnik Akademii znaniy. №3 (44). 2021. S. 149–151.
3. Popov D. V., Kutikova S. P. Formirovanie sistemy upravleniya organizatsii // Tekhnicheskoe tvorchestvo molodezhi. № 3(133). 2022. S. 38-43. – EDN WTJYYD.
4. Russkevich S.N., Mednikova O.V. Sistema edinogo elektronnoy dokumentooborota // Vestnik MMA. 2022. S. 89-91.
5. Sokolov E.A., Sereda S.N. Elektronnyy dokumentooborot VUZa // Perspektivy razvitiya informatsionnykh tekhnologiy. № 9. 2012. S. 39-44.

Сведения об авторах

*Попов Дмитрий Владимирович, к.э.н., доцент
Московский государственный технологический университет «СТАНКИН», Москва, Россия,
Email: d.popov@stankin.ru*

*Ральникова Катрина Владиславовна, аспирант
Московский государственный технологический университет «СТАНКИН», Москва, Россия
Email: k.ralnikova@stankin.ru*

Ключевые слова

Архив, электронный архив, документооборот, электронный документооборот, цифровая трансформация, нормативно-правовое обеспечение, системы электронного документооборота, внедрение электронного документооборота, государственные учреждения, классификатор, взаимодействие государственных организаций, электронные документы.

Dmitrij Popov, Katrina Ralnikova, Archive as a factor of digital transformation of an organization

Keywords

Archive, electronic archive, document management, electronic document management, digital transformation, normative-legal support, system of electronic document management, introduction of electronic document management, state institution, classifier, cooperation of state organizations, electronic documents.

DOI: 10.34706/DE-2023-01-08

JEL classification: M10 – Деловое администрирование: Общее, M15 – Управление ИТ, L86 – информационные и интернет-услуги; Компьютерное программное обеспечение, K39 – Другие основные области права: Другое,

Abstract

The article deals with the issue of complete transition to electronic document management for organization within the framework of digital transformation based on the formation of exchange contours between the archive of the organization and the municipal archive. Interaction of the federal state budgetary educational institution of higher education located in Moscow and the Main Archive Department of the city of Moscow based on the basis of regulations of the Federal Archival Agency is used as an example.

2. МНЕНИЯ

2.1. ИНДОКТРИНАЦИЯ ЭКОНОМИЧЕСКИХ НАРРАТИВОВ ДЛЯ МОЛОДЕЖИ КАК ЦИФРОВОЙ ПОКОЛЕНЧЕСКОЙ ОБЩНОСТИ ПОСРЕДСТВОМ ФГОС ВО 3++

Апенько С.Н., Лукаш А.В.

Омский государственный университет путей сообщения Омск, Россия

Статья выполнена на основе интерпретации универсальных компетенций федеральных образовательных стандартов поколения 3++ по направлениям бакалавриата, магистратуры и специалитета, которые рассматриваются авторами в качестве формализованных в нормативные тексты нарративов, а также проанализированы повествовательные структуры ключевых молодежных форумов России с 2019 года. Цель работы – установить на основе интерпретации нарративов индоктринируемых государством через систему высшего образования ориентиры социально-экономического поведения для молодежи, реализуемых как в рамках их трудовой деятельности, так и выходящие за границы профессиональной сферы. В результате установлены следующие нарративы: сетевого взаимодействия в социальных и экономических практиках; личной ответственности за результат; проектного мышления и организации в профессиональной деятельности; непрерывного формального и неформального образования в профессиональной деятельности; ответственного финансового и экономического поведения; стигматизации коррумпированного поведения и исключительного использования современных информационных технологий для решения задач профессиональной деятельности. В ходе интерпретации нарративов образовательных стандартов и проектных сессий молодежных мероприятий установлено, что как формализованные, так и не формализованные повествовательные структуры направлены на укоренение среди молодёжи идей о развитии предпринимательской инициативы, независимо от профессионализации выпускника, координации индивидуальной и социальной ответственности, важности поддержания и развития бриджингового социального капитала. В статье показано, что государство через систему высшего образования и программы молодежной политики транслирует для современных студентов и выпускников нарративы экономического и трудового поведения, ориентированные на траекторию инновационного и цифрового развития, а также, на снятие противоречий между просоциальными и эгоцентричными стратегиями поведения.

Введение.

Дискуссии о том, являются ли поколенческие общности исключительно плодом маркетинговых и рекламно-коммуникационных стратегий продвижения или группой, которая может и должна быть осмыслена академическим сообществом, в фокусе научного исследования, идут не один десяток лет. Особый интерес традиционно представляет молодое поколение, вокруг поведенческих характеристик которого разворачивается наиболее интенсивное обсуждение и среди практиков маркетинговых услуг, и в академической среде. Причина очевидна, ведь именно оно является важным фактором инновационной модернизации общества, политической и экономической сфер, системы образования – профессионального и высшего. В публикациях, посвященных молодежи как поколенческой общности, часто встречаются попытки обосновать использования понятия, которое бы наиболее точно и содержательно передавало ее ключевые характеристики – «поколение Google», «цифровые аборигены», «поколение Z», «Digital Tribe» и т.д. Развитие понятийного аппарата на примере молодежи, говорит в целом о развитии поколенческой теории и важности проработки ее методологической основе, благодаря которой мы можем получить устойчивые характеристики поколения, выражающие его сущность в политическом, культурном и социально-экономическом дискурсах.

Нам представляется, что сегодня актуально рассматривать молодое поколение как продукт целенаправленного, интеллектуального производства системы образования, маркетинговой и рекламной практики, массовой культуры, общественных и государственных институтов, фондов грантодателей и т.д. Важно установить приписываемые через нарративы образовательной среды молодому поколению экономических, профессиональных и трудовых стратегий так как, во-первых, это позволит в дальнейшем выделить необходимые индикаторы для терминологического обозначения общности, а, во-вторых, формирует перспективы для представления своего рода идеального типа современной молодежи как социально-экономической группы, который индоктринируется самим обществом – через культурный и социальный контекст.

Под понятием «индоктринация» мы понимаем процесс целенаправленного воздействия, основанного на явном или латентном внушении и повторении определенных идей, принципов, правил и убеждений всему обществу или определенной его части, в том числе поколенческой группе. Индоктринация предполагает своей целью безусловное принятие транслируемых смыслов, стремится к их легитимации и институализации. Нарратив в этом контексте выступает в качестве одного из ключевых инструментов индоктринации, так как его часто иррациональная основа способствует снижению уровня критического восприятия инкорпорируемых смыслов.

В нашей работе мы опираемся на разделение нарративов на две группы: повествовательные структуры, формализованные в нормативные тексты и неформализованные истории акторов. Основное внимание в ходе исследования было сделано на нарративы первого типа – к ним были отнесены компетенции из Федеральных государственных образовательных стандартов, которые индоктринируют реперные точки социального и профессионального развития выпускника высшей школы. В них заключен запрос государства на подготовку молодых людей как идеальных акторов социально-экономических отношений. В прочтении компетенций как нарративных конструкций мы опирались на концепцию Й.Брокмейера и Р.Харре, где они рассматриваются в качестве действенного слова для обозначения специального набора инструкций и норм, предписывающих, что следует и чего не следует делать в жизни. Исследуя проблемы нарративной конвенции, они подробно изучали механизмы влияния повествования на человека, заключая, что «истории руководят жизнью», «мы (*слушатели*) просто привыкаем к широкому репертуару сюжетных линий...врастаем в культурный канон нарративных моделей» [1, с.39-42]. Нарратив есть не только рассказ, сколько формализованное или нет повествование ориентированное на принятие, понимание и усвоение определённых правил в разных сферах жизнедеятельности человека, в том числе социально-трудовых отношений, потребления, отдыха, планирования семьи и т.д. Нарратив, таким образом, важный структурный элемент в механизме формирования институтов, их легализации и легитимации.

Изучение нарративных интеракций представляется актуальным для понимания алгоритмизации повседневной экономической жизни молодых людей, столь восприимчивых и открытых в силу своего возраста к убедительным по своему эмоциональному звучанию историям и рассказам.

В целом нарративный подход к изучению экономического поведения молодёжи позволяет:

акцентировать внимание на идеях, оформленных в повествовательные структуры, преимущественно для этой поколенческой общности;

установить на основе соотнесения этих идей с реальными стратегиями экономического поведения молодёжи характер обратной связи с ними;

определить соотношение рационального и иррационального обоснования идейного ядра повествования;

прогнозировать развитие экономических стратегий молодежи на основе вычленения и анализа вновь генерируемых нарративов.

Обзор литературы.

Термин «нарратив» и нарративный подход давно и успешно укрепили свое место в методологическом базисе многих наук о человеке и обществе, но остаются во многом эксклюзивными в экономике, несмотря на рост интереса к нему в последние годы. Одним из главных исследователей проблематика нарратива в экономических процессах является R.J. Shiller [2,3]. В своих выступлениях и публикациях, в том числе в книге «Нарративная экономика: как истории становятся вирусными и определяют крупные экономические события», он предлагает рассматривать нарративы (повествования) как одну из самых главных движущих сил современных рынков. R.J. Shiller последовательно доказывает большую склонность рядового человека, независимо от его принадлежности к конкретной культуре, скорее игнорировать аналитические аргументы и доводы, в пользу историй и рассказов, оставляя в стороне их правдивость и реалистичность. Дальнейшие исследования в области нарративов должны, по его мнению, принципиально изменить науку и представление о хозяйствовании человека: «Экономика в течении следующего десятилетия коренным образом изменится, и эта привязанность к рациональному индивидуалистическому человеческому поведению будет существенно смещена» [2, с.84].

R.J. Shiller изучал вирусные возможности нарратива, которая проявляется в том, что та или иная экономическая идея, оформленная в историю, через коммуникативные практики способна захватить весь мир, став глобальным трендом в экономической повестке. Вирусная сила нарратива дает основание R.J. Shiller утверждать в своих работах, что увлеченность повествованием может вырастать до масштабов вирусных бедствий, видимым проявлением которых является расхождение реальных возможностей рынков и экономическим поведением людей. Несмотря на обилие примеров нарративов с негативной коннотацией, которые прямо пугали и озадачивали людей (пример с нарративом о «технологической безработицы» в годы Великой депрессии, который влиял на обывателей – резкое сокращение трат и пессимистичные оценки перспектив своего будущего), в своих трудах R.J. Shiller не стигматизирует нарратив, считая, что в целом рассказ способен к продуцированию продуктивных социально-экономических практик.

Akerlof G. раскрывает функциональную сторону повествовательных структур, которые играют важную роль в понимании окружающей среды; фокусировании внимания; прогнозировании событий;

мотивации действий и поступков; распределении социальных роли и идентичности, определении отношений власти, установлении и передаче социальных норм [4, с.70].

Место повествовательных конструкций в микро и макроэкономических процессах представлено сегодня в следующих исследованиях: анализ нарративов о макроэкономических мерах реагирования на сложные кризисы в экономике и окружающей среде (E.Urhammer, I. Rørke) [5]; влияние нарративов экономики замкнутого цикла на интерпретацию проблем и действий по борьбе с отходами в производственном секторе (S. Leipold, K. Weldner, M. Hohl) [6]; структура нарративов о деловых циклах за период с 1950 по 2019 годы (Ch. Bertsch, Is.Hull, X. Zhang) [7]; связь историй и повествований из социальных сетей на колебания настроений инвесторов и временные отклонения цен на рынках акций (S. Liu, P. Gaskell, Fr. McGroarty) [8]; управление нарративами в корпоративном секторе (L.D. Brown, A.C. Call, M.B. Clement, N.Y. Sharp) [9]; зависимость нарративов крупных менеджеров компаний от состояния финансовых дел в ней (H. S. Asay, R. Libby, K.Rennekamp) [10].

Сегодня активно идет развитие прикладных исследований о связи нарратив и экономических практик предпринимателей: A. Vossen, Ch. Ihl [11] исследуют вопросы влияния повествования и их ответственности потребителем ожиданиям на предпринимательские стратегии по продвижению и сохранению на рынках своих продуктов; S. Singh, P.D. Corner, K. Pavlovich [12] – качественный нарративный подход к исследованию личного опыта предпринимателей, испытывающих стигматизацию, связанную с неудачей предприятия; G. Fisher, E. Neubert, D. Burnell [13] – специфические нарративы о находчивости, как инструменты привлечения необходимых ресурсов для предпринимателей в условиях их ограниченности.

P.Andre, I. Haaland, Ch. Roth, J. Wohlfart [14] сфокусировали свое внимание на историях, которые люди рассказывают для объяснения макроэкономических явлений в контексте исторического всплеска инфляции. Исследовательский коллектив опросил более чем 8 000 американских домохозяйств и 100 академических экспертов. В результате установлено. Во-первых, по сравнению с экспертами, нарративы домохозяйств более грубые, в них меньше внимания уделяется стороне спроса и чаще встречаются политически нагруженные объяснения. Во-вторых, нарративы домохозяйств сильно влияют на их инфляционные ожидания. В-третьих, эксперимент по изменению потребления новостей показывает, что СМИ являются важным источником нарративов в американском обществе.

R. Venabou, A. Falk, J. Tirole [15] исследовали как интерпретируются и передаются нарративы дружим, в каких ситуациях они распространяются вирусно. Особое внимание в своей работе коллектив авторов уделил тому, как нарративы конкурируют с императивами (общими моральными правилами или предписаниями) в качестве альтернативных способов коммуникации для убеждения агентов вести себя желательным образом. Одно из фундаментальных наблюдений, которое удалось им установить – оправдательные нарративы могут позволить людям поддерживать положительный имидж, когда на самом деле они действуют сомнительным с моральной (т.е. императивной) точки зрения образом. И наоборот, ответственные нарративы могут способствовать поддержанию лучших социальных норм в обществе.

Серьезную работу по изучению повествовательных структур в экономике провели A. Hillenbrand, E. Verrin [16]. Они доказывают, что нарративы играют решающую роль в формировании наших управленческих решений. Авторы исследовали позитивные нарративы (истории в пользу просоциального действия) и негативные нарративы (истории в пользу аморального и эгоистичного поведения) влияют на просоциальное поведение, опираясь на прикладные эксперименты (использованы игровые технологии) с более чем 1500 респондентов.

Нарратив как действенную форму преимущественно представления содержания идеи рассматривает В.Л. Тамбовцев [17]. Критический анализ идейного и нарративного «поворотов» в экономических исследованиях построен им на то, что воспринимаются и становятся убеждениями и трансформируются в действия те идеи, которые отвечают интересам акторов и у него есть достаточный ресурсный потенциал для их воплощения в своей практике.

Нарратив как ценный источник качественных данных, опора на который открывает перспективы для понимания и моделирования поведения акторов в социально-экономических отношениях, рассматривают В.В. Вольчик [18, 19] и Е.В. Маслюкова [18]. В своей работе они рассматривают не только нарратив, но и мем, который является квинтэссенцией идеи и имеет большой потенциал для распространения, так как несет в себе более обобщенную и легкую для восприятия информацию. «Идеи циркулируют в группах (больших и малых) посредством рассказывания историй, что включенными и сторонними наблюдателями воспринимается через устные, письменные и визуальные источники информации, – пишут В.В. Вольчик и Е.В. Маслюкова – такие источники информации являются нарративами и мемами» [18, с.162]. Понимание нарратива как информации, которая не всегда должна быть оформлена в виде текста, но может представлять собой визуализированную форму, является важным для изучения социокультурного пространства, в том числе архитектуры, профессионального, городского, рекреационного пространства как транслятора идей и смыслов, направляющих и побуждающих людей к определенной экономической интеракции.

Влияние нарративов на поведение экономических агентов на примере современных США для поддержания социально-экономического качества жизни населения в стране исследует П.А. Патрон и

А.Ю. Шмелева. Главный тезис, доказываемый авторами – «нарратив не только интерпретирует институты, но также распространяет и изменяет их» [20 с.67].

В.А. Ермаков, Ю.В. Бажданова [21] рассматривают деятельность СМИ, через которые рассказываемые истории как задают тренды, так и стигматизируют определенные профессии, для молодых людей. Авторы делают акцент на искажении образа профессиональной структуры общества в современном медиапространстве страны посредством нарративов. Исследователи обращают внимание на основные факторы формирования профессиональных иллюзий студентов через повествование отечественных медиа.

В настоящее время идет активное экспериментальное изучение экономического поведения не только молодежи с 18 лет, но также подростков и детей. В частности M. Sutter, C. Zoller, D. Glätzle-Rützler [22] экспериментально рассмотрели проявление рациональности выбора детей и подростков, временных предпочтений, риска, социальных предпочтений, при этом были учтены детерминанты экономического поведения, такие как социально-экономический статус родителей и социальный контекст взаимодействия детей и подростков до 18 лет, выраженный в том числе в нарративной интеракции. Авторы обращают внимание на политические интервенции (под которыми они рассматривают правила и их интерпретацию через нарратив целенаправленно устанавливаемые государственными институтами), которые имеют целью повлиять на экономическое поведение детей и подростков. Один из выводов их исследования – политические вмешательства, направленные на поощрение одного типа поведения в большей степени, чем другого, проявляются в экономических практиках подростков и детей.

W.N. Lekfuangfu, R. Odermatt [23] изучая роль устремлений в мобильности и благосостоянии поколений, приходят к выводу, что стремления молодых людей в значительной степени зависят не только от социально-экономического статуса родителей, но от собственных стремлений последних в отношении своих детей, которая выражается через нарративную интеракцию.

Таким образом, исследовательский интерес в современной академической среде сфокусирован вокруг экономического нарратива вообще, без относительного того, какая социальная группа является его реципиентам. Возраст, пол, профессиональные отличия не исследуются ни как барьеры, корректирующие восприятие нарративов, ни как возможные характеристики для формирования групповых нарративов, что, на наш взгляд актуализирует важность изучения индоктринации экономических и трудовых нарративов для современной молодежи как цифровой поколенческой общности.

Гипотезы и методы исследования.

Современная молодежь как поколенческая общность является во многом результатом индоктринации определенных норм, установок и стереотипов поведения, относительно социальных ролей и экономического поведения, которые как целенаправленно, так и спонтанно, через нарратив создают такие образцы и стереотипы реагирования, на базе которых формируется жизненная позиция личности и группы, их экономические стратегии. Важную роль в этом процессе выполняет государственная система высшего образования.

Экономические установки личности – это не только знания и отношение к процессу хозяйствования, охватывающего производство, потребление, распределение, но и готовность действовать в отношении них определенным образом и по определенным правилам, согласно индоктринируемым нарративам. Несмотря на то, что сила и интенсивность индоктринации нарративов, смысловые единицы которых обращены к экономическим ценностям, нормам и установкам, во-первых, не предполагает автоматического укоренения предлагаемых интенций в социально-экономической практике среди всей молодежи, а, во-вторых, они могут быть значительно нивелированы и ослаблены альтернативными установками, трансляция и формирование которых реализуется конкурирующими субъектами общественного производства, они представляют собой своеобразную технологию формирования идеальной модели экономического поведения.

Гипотеза нашей работы сформулирована следующим образом – государство через систему высшего образования транслирует для молодого поколения нарративы экономического и трудового поведения ориентированные, во-первых, на траекторию инновационного и цифрового развития, а, во-вторых, на снятие противоречий между просоциальными и эгоцентричными стратегиями поведения.

Объектом исследования являются экономические и трудовые нарративы для молодежи. Предметная область исследования – федеральные государственные образовательные стандарты как нарративные конструкции об экономическом и трудовом поведении молодежи.

Мы предполагаем, что индоктринация экономических нарративов реализуется, как в результате работы вертикальных каналов влияния, например, через образовательные стандарты, национальные проекты, инициативы частных фондов и т.д., так и горизонтальных – через кооперацию и взаимодействие, в конкурсных программах, направленных на поддержку и развитие молодежных инициатив, профессиональную общность, корпоративную культуру, образовательную среду и т.д.

Индоктринация экономических нарративов не всегда является результатом целенаправленного влияния и воздействия, этот процесс может иметь спонтанный характер и быть результатом работы акторов, чья инициатива и деятельность посредством рассказывания историй была направлена на решение других задач.

Для выявления установок и норм экономического поведения молодежи, индоктринируемых в современном обществе посредством нарративов мы провели дискурсивный анализ повествований ФГОС ВО 3++ с опорой на идеи и принципы социального конструкционизма. В соответствии его логики вся социальная реальность конструируема и воспроизводится людьми в процессах её интерпретации ими и формулировки знаний о ней. Следуя идеи социального конструкционизма о том, что любое знание происходит и поддерживается за счёт социальных интеракций, мы рассматриваем экономическое поведение современного молодого поколения как результат социальных интеракций. Так как любые знания являются продуктом социальной конвенции, то любые системы ценностей, социальные образования, стереотипы поведения и реагирования воспринимаются людьми как объективная реальность.

Таким образом, как вся социальная реальность, так и ее часть – экономические установки и ценности, приписываемые молодому поколению, конструируется самими акторами социальных процессов. Нарративный подход в изучении экономических вопросов позволяет не только вычлнить единицу повествования, предписывающую должное, в смысле легитимное и легальное поведение, конгруэнтное ожиданию общества и его перспективным планам на будущее, но и, следуя обратному принципу, понять какие модели экономического поведения признаются не желательными, опасными, деструктивными и т.д. Данный подход позволяет также установить перечень тем и вопросов, которые находятся либо вне нарративного авангарда, либо вообще максимально стигматизированы социальными институтами.

Конкретным методом изучения экономических нарративов, используемым в работе, была выбрана интерпретация, так как она позволяет в соответствии с главной стратегией нашего исследования выявлять смыслы повествования, содержащихся в репрезентативных источниках. Интерпретация нарратива, как показано Э.Бетти, позволяет решать три задачи: понять смысл нарратива (распознающая интерпретация), в случае заинтересованности или обстоятельств обеспечить процедуру передачи его смысла (репрезентативная интерпретация) и выстроить свое поведение в соответствии с нормами, догмами и требованиями нарратива (нормативная интерпретация) [24, с.37]. Качественные исследования экономических нарративов открывают в дальнейшем хорошие перспективы для установления типов идей, упакованных в устный, письменный или визуализированный нарратив.

Результаты исследования.

В ходе работы над статьей была составлена нарративная рамка, на основе рассмотрения универсальных компетенций из ФГОС ВО 3++ по направлениям бакалавриата (188 стандартов) и магистратуры (188 стандартов), специалитета (103 стандарта), которые на наш взгляд прямо связаны с трансляцией смыслов должного, то есть ожидаемого со стороны государства, экономического поведения молодежи.

В своей профессиональной деятельности выпускник независимо от направления специальности подготовки должен быть способен:

Самостоятельно разрабатывать и реализовывать проекты (УК-2) – нарратив проектного мышления и организации в профессиональной деятельности. Анализ компетенции показывает, что государство ждет от молодого профессионала способностей действовать исходя из имеющихся ресурсов и ограничений. Нарратив предписывает выпускнику ориентироваться на оптимальные способы решения поставленных профессиональных задач с обязательным соблюдением правовых норм. Идея проектного мышления ориентирована шире границ трудовой и профессиональной сфер, она предполагает взгляд на любые вызовы как на проекты. Нарратив проектного мышления и действий отвечает имманентным принципам инновационного развития, на которое ориентированы современные программы экономического развития страны. Государству важно, чтобы молодой профессионал был способен к креативному генерированию идей и проектов, а также их последующему производственному воплощению, включая кастомизацию продукта. Данный нарратив направлен на стимулирование просоциального экономического поведения.

Проявлять лидерские качества и развивать навыки командной работы (УК-3) – нарратив сетевого взаимодействия в социальных и экономических практиках и нарратив личной ответственности за результат. Стандарты воспроизводят повествовательные структуры как о важности кооперации и координаций своей деятельности в трудовом коллективе, проектной группе, проектной команде и т.д. (т.е., просоциальный нарратив), так и о необходимости не бояться принимать на себя ответственность, демонстрировать уверенность в себе, заявлять о своих амбициях, осознавать, что успех в жизни зависит от себя и т.д. (эгоцентрический нарратив).

Первый нарратив ориентирован на социальное взаимодействие выпускников как важную составляющую их профессиональной и трудовой коммуникации. Через раскрытие этой компетенции мы видим понимание со стороны государства важности развития горизонтальных связей в современной экономике, где инновационные предприятия создают сетевую институционально-организационную среду, эффективную для диффузии новшеств и объединения малых фирм, нацеленных на решение единых производственных задач.

Второй нарратив – в большей степени на практике предпринимательской деятельности с присущей ей креативностью и инициативностью; осознанным риском и личной ответственностью за принимаемые решения и риск; регулярным поиском и реализацией новых экономических возможностей.

Выстраивать и следовать стратегии саморазвития и самоорганизация (УК-6) – нарратив непрерывного формального и неформального образования. Нарратив воспроизводит идею об образовании в течение всей жизни человека. Повествовательная структура акцентирует внимание на том, что время и здоровье это важные ресурсы, от эффективного управления которыми зависит успешность личности в социальном и профессиональном плане. Нарративы стандартов о саморазвитии воспроизводят повествовательные структуры, которые генерируют с начала 1970-ых международные структуры (например, ЮНЕСКО). Их прочтение показывает, что любое профессиональное образование не достаточно для того, чтобы быть конкретным на рынке труда в среднесрочной или долгосрочной перспективе. Технологические вызовы диктуют необходимость следовать стратегии постоянного образования, как формального, так и нет. В совокупности с неформализованными историями акторов медиакоммуникативных структур (например, блогерами) они индоктринируют идею об образовательном и физическом капитале личности, инвестиции в которые социально одобряемы и имеют прямую экономическую отдачу. Данный нарратив акцентирует внимание на эгоцентричном поведении – ориентирует личную активность на самоменджменте.

Анализ нарративов ФГОС ВО 3+, в виду того, что их утверждение и опубликование было растянуто во времени, позволяет проследить в каком направлении корректировался запрос государства на подготовку молодых людей как акторов социально-экономических отношений. Так, часть стандартов (112 по направлениям бакалавриата и 40 по специалитету), вышедших до ноября 2020 года были реализованы без двух универсальных компетенций «Экономическая культура, в том числе финансовая грамотность» и «Гражданская позиция». С целью выравнивания компетентной рамки молодых профессионалов во все ранее опубликованные стандарты с 1 сентября 2021 г. Приказом Минобрнауки России от 26 ноября 2020 г. N1456 они были включены.

В будущей профессиональной деятельности выпускник бакалавр и специалист в соответствии с ними должен быть способен (в магистерские программы эти нарративы не включены):

Принимать обоснованные экономические решения (УК-10) – нарратив ответственного финансового и экономического поведения. Повествовательная структура не ограничивается только рамками профессиональной сферы и трудовых отношений молодого специалиста, она индоктринирует идею об исключительной важности демонстрировать обоснованные экономические решения постоянно, в различных ситуациях социального взаимодействия, во всех областях жизнедеятельности. Идея о финансовой грамотности, заключенная в этом нарративе, связана со Стратегией повышения финансовой грамотности в Российской Федерации на 2017 – 2023 годы, утвержденной Распоряжением Правительства РФ от 25 сентября 2017 г. № 2039-р. Нарративы стандарта и Стратегии ориентируют молодых людей на возрастание индивидуальной ответственности в обеспечении личной финансовой безопасности, которая в свою очередь становится важным фактором как семейного экономического благополучия, так и всего общества. Нарратив не дает прямого ответа на вопрос – какая теория экономического поведения заложена в требованиях стандартов, но его анализ в совокупности с остальными компетенциями показывает, что государство демонстрирует понимание – молодому профессионалу предстоит действовать в ситуации, где экономическое пространство наполнено усложняющимися институтами, высока неопределенность и риски.

Быть готовым проявлять и защищать свою гражданскую позицию (УК-11) – нарратив стигматизации любого проявления коррумпированного поведения. Сама формулировка наименования универсальной компетенции указывает на нетерпимое отношение к коррупции как имманентного принципа экономического поведения молодых людей. Повествовательная структура стандарта индоктринирует идеи о том, что уровень и масштаб взяточничества, его мотивы и сферы применения не имеют никакого значения – любое его проявление разъедает общество, выступая тормозом развития экономики. Так же, как и нарратив о финансовой грамотности, повествовательная структура о нетерпимом коррумпированном поведении является частью государственной политики, в частности она согласуется с Национальным планом противодействия коррупции на 2021-2024 годы. Эти нарративы ориентированы на просоциальную модель поведения.

Только одна из универсальных компетенций так и осталась, представлена в части стандартов бакалавриата и специалитета, и не была добавлена во все ФГОС ВО 3+ приказом от 26 ноября 2020 г. N1456 – инклюзивная. Ее нарратив индоктринирует идею, которой государство последовательно с 2013 года уделяет большое внимание – создание материальной инфраструктуры и социальной среды, которая позволит включать в активные экономические и трудовые отношения людей с ограниченными возможностями здоровья, несмотря на типы их нозологий.

Положение о том, что государство через образовательный стандарт индоктринирует идею о современном молодом поколении как цифровом может быть подкреплена анализом общепрофессиональной компетенции, которая в разных ФГОС ВО 3+ отличается индексом и не значительно в содержательной части, но уже упомянутым приказом №1456 была существенно скорректирована – нарратив был сориентирован на область профессиональной деятельности выпускника. Стандарт устанавливает норму, предписывает как обязательное условие успешности молодого человека на рынке труда и в экономических отношениях, способность использовать современные информационные технологии для решения задач профессиональной деятельности, понимая принципы их работы. Данный формализо-

ванный нарратив связан с федеральным проектом «Кадры для цифровой экономики» национального проекта «Цифровая экономика» на период с 2019 по 2024 годы.

Анализ компетенций ФГОС ВО 3++, их универсальный характер, показывает, что, во-первых, не смотря на область трудовой деятельности, к которой готовится выпускник конкретной программы, государство активно использует возможности нарративов в системе высшего образования для целенаправленных институциональных изменений в стране и преодоления накопившихся институциональных разрывов. Молодежь, как основная часть выпускников высшей школы в нашей стране ориентируют на модели экономического поведения и трудовой деятельности в краткосрочной перспективе понятные и легитимные для их сверстников, так как их логика экономического мышления и навыки экономических отношений, формируются в относительно гомогенной университетской среде. Это в перспективе должно повысить доверие между участниками экономических отношений и снизить общий уровень транзакционных издержек.

Во-вторых, опираясь на исследования Г. Хофстеде, можно уверенно говорить, что через нарративы образования у государства есть возможность целенаправленно влиять на социокультурный контекст, делая его инструментом развития общества. В частности, нарратив сетевого, командного взаимодействия в социальных и экономических практиках (УК-3) направлен на формирование бриджингового социального капитала, который может способствовать избеганию неопределенности и снижению дистанции власти. Этот вид социального капитала, А.А. Аузан, возможно является в текущих условиях самым ценным продуктом, который можно создавать в России [25, с. 178].

В-третьих, повествовательная структура стандартов в части универсального компонента согласуется с базовыми идеями теории лидерства о характере современного мира, как принципиально нестабильного (volatility), неопределенного (uncertainty), сложного (complexity) и неоднозначного (ambiguity). VUCA-мир индоктринирует установки на важность постоянного самообучения и самообразования, ввиду увеличения трудоспособного возраста и неизбежного реагирования на результаты инновационного развития общества. Это трансляция нарратива о приоритете «мягких компетенций» (soft skills) в период новой промышленной революции и технологических вызовов перед владением «твердых компетенций» (hard skills). Идеи об исключительном значении навыка командной работы и критического мышления, как конкурентного преимущества в ситуации, когда ускоряется не только характер производственного процесса, но и социальных, профессиональных связей. Эти и другие идеи, которые включает в себе концепция VUCA-мир, в полной мере разрываются в нарративах образовательных стандартов высшей школы. Особенно в компетенциях магистров, где индоктринируется нарратив о необходимости выработать стратегии действий проблемных ситуациях, как фактор успеха в профессиональной деятельности (УК-1).

В-четвертых, характер повествовательных структур стандартов говорит о стремлении государства получить выпускников способных к осуществлению в стране реальной траектории экономического развития в логике цифровой трансформации, основанной на рыночном и инновационном поведении акторов – молодого поколения.

Пятое положение вытекает из сопоставления универсальных компетенций ФГОС ВО 3++ с нарративом общекультурных требований ФГОС ВО предыдущего поколения – новые стандарты в большей степени концентрируются на идеях эффективности проектного подхода, командных (сетевых) форм организации труда и самозанятости, то есть практиках которые составляют основу организации субъектов информационно-коммуникативного бизнеса. Государство ожидает от системы образования, что результатом подготовки выпускника будет его естественная интеграция и в социальную, и в профессиональную цифровую среду.

Повествовательные структуры, формализованные в нормативные тексты, такие как ФГОС ВО, очевидно, что ориентированы в первую очередь на всех участников образовательного процесса, кроме обучающихся и родителей. Редко кто из абитуриентов интересуется простым набором дисциплин и практик, что предстоит освоить за годы обучения. Поэтому важную роль в индоктринации посредством нарративов описанных выше идей и установок, выполняют государственные структуры, реализующие молодежную политику, общественные организации и аффилированные с ними акторы, активно привлекаемые для работы в их проектах.

В рамках исследования мы проанализировали программы молодежных форумов с 2019 года: «Ростов», «СелиАс», «Молодежная команда страны», «Выше крыши», «Сибирь 2020», «Восток», «Ладога», «Утро», «Байкал», «Территория смыслов», «Бирюса», «iВолга», «Евразия Global», форум молодых семей и образовательный форум «Машук».

На этих площадках, как и в университетской среде, преобладают стратегические нарративы – повествовательные структуры ориентированы на пролонгированные практики в контексте их жизненного выбора.

Анализ нарративов генерируемых на мероприятиях, проводимых при участии государства, например, молодежном форуме «Территория смыслов», международный форум «Слет успешных предпринимателей», Всероссийский молодежный образовательный форум «Территория инициативной молодежи «Бирюса», преобладают повествования о важности вовлечения молодежи в бизнес, необходимости постоянно совершенствовать свои компетенции, подчеркивается важность диалога не только во внутрикорпоративном цеху предпринимателей, но и с государственными структурами, которые заинтере-

ресованы в стимулировании молодых людей к предпринимательству. Транслируется нарратив об исключительной важности различных социальных проектов во всех областях социальной сферы – социального предпринимательства. Со стороны приглашенных акторов, представителей бизнеса, генерируете нарратив о том, что предпринимательство — это важная предпосылка для развития демократических институтов и гражданского общества.

Идея о социальной ответственности, как неотъемлемой части любого бизнеса с первых шагов молодежного предпринимательства, говорит о просоциальной направленности нарратив акторов.

Ключевыми повествовательными структурами последних лет на этих площадках являются нарративы о том, что «Россия – страна возможностей» и развитие, самореализация молодежи не возможна без эффективной системы социальных лифтов, создаваемых государством и частным бизнесом.

Нарративы молодежных форумов полностью согласуются с содержанием идей ФГОС ВО 3++: речь и о важности обучения, стажировок – непрерывное обучение и самообразование через всю жизнь, коммуникации с коллегами из других сфер, обмен опытом и знаниями как факторе образовательном и профессиональном росте, внутренняя мотивация как ключевой фактор успеха.

В медийной повестке форумов присутствуют нарративы, направленные на развенчание устоявшихся стереотипов о той или иной профессии, которые существуют сегодня в общественном сознании. Например, форум «Алтай территория развития» продвигает нарратив, что сегодня аграрные профессии в стране являются высокотехнологичными, инновационными и востребованными сферами самореализации, с растущим спросом в будущем.

Общая коннотация нарративов молодежных форумов страны – поддержка со стороны общества и государства молодежных инициатив обретает все более разнообразные формы и становится все более доступной и понятной.

Нарративные ряды экономического и трудового поведения, адресованные российской молодежи, корректируются с учетом текущих геополитических факторов. На Международном молодежном экономическом форуме в рамках ПМЭФ в 2022 году повествовательная структура акцентирует внимание на точках роста и развития молодежного предпринимательства. «Сейчас за счёт освободившихся ниш на рынке, можно выйти совсем на другой уровень бизнес-сообщества. В экосистеме молодежного предпринимательства России есть тысячи успешных действующих кейсов, и это шанс для них обрести партнёров среди гостей международного форума, получить инвестиции или же сделать рекламу своему товару или услуге».

Направленность нарративов грантов и целевых программ, оператором которых являются государственные структуры в рамках молодежного форума транслируют следующие просоциальные нарративы: поддержка инициатив по временной и сезонной занятости, студенческие стажировки и практики, развитие студенческого социального предпринимательства, поддержка трудовых проектов студенческих отрядов, поддержка проектов по развитию инициатив самозанятых студентов.

Нарративы, сопровождающие продвижение грантов – молодые люди обладают достаточным потенциалом, чтобы внести вклад в экономическое развитие своего региона, в процессе которого государство является надёжным, доверительным партнером.

Заключение.

Анализ повествовательных структур, формализованных в нормативные тексты, показывает, что государство индоктринирует молодому поколению экономическое поведение соответствующее рыночным институтам и механизмам, с выраженными стратегиями на инновационную деятельность, которой предстоит состояться в цифровой среде. Нарративы направлены на укоренение идей о развитии предпринимательской инициативы, независимо от профессионализации выпускника, координации индивидуальной и социальной ответственности, важности поддержания и развития горизонтальных социальных связей.

Нарративы ФГОС ВО 3++ показывают, что государство рассматривает систему высшего образования в качестве одной из важнейших в процессе выращивания рыночно ориентированных социально-экономических институтов, которые должны отвечать требованиям цифровой материальной инфраструктуры и социальной среде. В повествовательных структурах стандартов записан запрос государства на подготовку выпускников, которые в своей профессиональной и социальной практике будут действовать, от постановки целей и принятия управленческих решений, до их реализации, в соответствии с консолидированными правилами, понятными и легитимными со стороны других участников рынка, также «выращенных» на универсальной компетентностной рамке.

Особое место в индоктринации смыслов занимает идея о самообразовании и образовании через всю жизнь. Нарративы государства показывают, что институциональная среда образования – первоначальная ступень в профессионализации молодых людей. Повествовательные структуры ориентируют выпускника на модель профессионального поведения в условиях укоренения цифровой среды, которая предполагает постоянное развитие его навыков и умений, конгруэнтным вызовам времени, через трудовую деятельность. Этот нарратив соответствует инновационной направленности государственных инициатив в сфере образования.

Если рассматривать нарративы экономического поведения молодежи в стратегиях выбора: предпринимательская деятельность или наемный труд, то молодежные форумы акцентируют большее

внимание на предпринимательских компетенциях, индоктринируя идею об исключительной их важности в вопросе хозяйственного развития страны и устойчивого положения регионов. Разница в том, что в случае, когда трансляции смыслов идут через медиакоммуникационные системы аффилированные с государством, то акцент в повествовательных структурах делается на важности координации государства и бизнеса, в логике которой формируется, одобряемая этими актора, стратегия действий и принятия управленческих решений. В случае, когда актер не аффилирован – преобладают нарративы об имманентных признаках предпринимательства – самостоятельность, инициатива и неограниченные возможности в росте дохода, в совокупности усиливающий ее конкурентный статус в сознании молодежи, перед наемным трудом, несмотря на все риски.

Индоктринация идей через нарративы в равной степени как формализованных в нормативных текстах, так и неформализованных историях акторов, совсем не означают их укоризненность и последующее проявление в реальных действиях реципиентов. Компетентностная рамка выпускника, как нарративная структура, удалена от реальной детерминации поведения выпускников. Это обусловлено влиянием объективных условий функционирования университетской среды, которая наделена полномочиями самостоятельно устанавливать индикаторы достижения компетенций (индивидуальное прочтение нарратива) и академической свободой выбора педагогических средств и технологий обучения, которые в разной степени способны операционализировать содержание повествовательных структур стандарта в устойчивые паттерны поведения. В свою очередь и сами будущие выпускники, интегрированные в разнообразные социальные связи, обладающие разным социальным и культурным капиталом и находящиеся под влиянием медиакоммуникационных систем, являются объектом воздействия нарратив, характер и содержание идей, заключенных в которые может противоречить ожидаемому со стороны государства идеальному экономическому и трудовому поведению

Литература

1. Брокмейер Й., Харре Р. Нарратив: проблемы и обещания одной альтернативной парадигмы // Вопросы философии. – 2000. – №3. – С. 29-42.
2. Shiller R.J. Narrative Economics and Neuroeconomics // Finance: Theory and Practice. – 2018. – Vol.22, iss.1. – P. 64-91. – DOI: [10.26794/2587-5671-2018-22-1-64-91](https://doi.org/10.26794/2587-5671-2018-22-1-64-91)
3. Shiller R.J. Narrative Economics: How Stories Go Viral and Drive Major Economic Events. – Princeton: Princeton University Press, 2019. – 378 p. – DOI: [10.1017/S1053837221000110](https://doi.org/10.1017/S1053837221000110)
4. Akerlof G., Snower D. Bread and bullets // Journal of Economic Behavior & Organization. – 2016. – Vol. 126, iss. PB. – P. 58–71. – DOI : [10.1016/j.jebo.2015.10.021](https://doi.org/10.1016/j.jebo.2015.10.021)
5. Urhammer E., Røpke I. Macroeconomic narratives in a world of crises: An analysis of stories about solving the system crisis // Ecological Economics. – 2013. – Vol. 96. – P. 62-70. – DOI: [10.1016/j.ecolecon.2013.10.002](https://doi.org/10.1016/j.ecolecon.2013.10.002)
6. Leipold S., Weldner K., Hohl M. Do we need a 'circular society'? Competing narratives of the circular economy in the French food sector // Ecological Economics. – 2021. – Vol. 187. – 107086. – DOI : [10.1016/j.ecolecon.2021.107086](https://doi.org/10.1016/j.ecolecon.2021.107086)
7. Bertsch Ch., Hull Is., Zhang X. Narrative fragmentation and the business cycle // Economics Letters. – 2021. – Vol. 201. – 109783. – DOI : [10.1016/j.econlet.2021.109783](https://doi.org/10.1016/j.econlet.2021.109783)
8. Liu S., Gaskell P., McGroarty Fr. Where and about what? Price relevant narratives depend on topic and media type // Economics Letters. – 2022. – Vol. 213. – 110363. – DOI : [10.1016/j.econlet.2022.110363](https://doi.org/10.1016/j.econlet.2022.110363)
9. Brown L.D., Call A.C., Clement M.B., Sharp N.Y. Managing the narrative: Investor relations officers and corporate disclosure // Journal of Accounting and Economics. – 2019. – Vol. 67, iss. 1. – P. 58-79. – DOI: [10.1016/j.jacceco.2018.08.014](https://doi.org/10.1016/j.jacceco.2018.08.014)
10. Asay H.S., Libby R., Rennekamp K. Firm performance, reporting goals, and language choices in narrative disclosures // Journal of Accounting and Economics. – 2018. – Vol. 65, iss. 2-3. – P. 380-398. – DOI: [10.1016/j.jacceco.2018.02.002](https://doi.org/10.1016/j.jacceco.2018.02.002)
11. Vossen A., Ihl Ch. More than words! How narrative anchoring and enrichment help to balance differentiation and conformity of entrepreneurial products // Journal of Business Venturing. – 2020. – Vol. 35, iss. 6. – 106050. – DOI: [10.1016/j.jbusvent.2020.106050](https://doi.org/10.1016/j.jbusvent.2020.106050)
12. Singh S., Corner P.D., Pavlovich K. Failed, not finished: A narrative approach to understanding venture failure stigmatization // Journal of Business Venturing. – 2015. – Vol. 30, iss. 1. – P.150-166. – DOI: [10.1016/j.jbusvent.2014.07.005](https://doi.org/10.1016/j.jbusvent.2014.07.005)
13. Fisher G., Neubert E., Burnell D. Resourcefulness narratives: Transforming actions into stories to mobilize support // Journal of Business Venturing. – 2021. – Vol. 36, iss. 4. – 106122. – DOI: [10.1016/j.jbusvent.2021.106122](https://doi.org/10.1016/j.jbusvent.2021.106122)
14. Andre P., Haaland I., Roth Ch., Wohlfart J. Narratives About the Macroeconomy. [ECONtribute Discussion Papers Series](https://repec.cepr.org/repec/cpr/ceprdp/DP17305.pdf). – 2022. – Vol. 127. – 104 p. – Available at: <https://repec.cepr.org/repec/cpr/ceprdp/DP17305.pdf> (accessed date: July 12, 2022)
15. Benabou R., Falk A., Tirole J. Narratives, Imperatives, and Moral Reasoning. Collaborative Research Center Transregio 224. – 2019. – 54 p. – Available at: <https://www.crctr224.de/en/research->

- output/discussion-papers/archive/2019/narratives-imperatives-and-moral-reasoning-roland-benabou-armin-falk-jean-tirole (accessed date: July 25, 2022)
16. Hillenbrand A., Verrin E. The asymmetric effect of narratives on prosocial behavior // *Games and Economic Behavior*. – 2022. – Vol.135. – P. 241-270. – DOI: [10.1016/j.geb.2022.06.008](https://doi.org/10.1016/j.geb.2022.06.008)
 17. Тамбовцев В. Л. Идеи, нарративы и изменения в экономике // *Terra Economicus*. – 2019. – № 1. – С. 24-40. – DOI: [10.23683/2073-6606-2019-17-1-24-40](https://doi.org/10.23683/2073-6606-2019-17-1-24-40)
 18. Вольчик В.В., Маслюкова Е.В. Нарративы, идеи и институт // *Terra Economicus*. – 2018. – №2. – С. 150-168.
 19. Вольчик В.В. Нарративы и понимание экономических институтов // *Terra Economicus*. – 2020. – №2. – С. 49-69. – DOI: [10.18522/2073-6606-2020-18-2-49-69](https://doi.org/10.18522/2073-6606-2020-18-2-49-69)
 20. Патрон П.А., Шмелева А.Ю. Роль нарративов в макроэкономической политике // *США & Канада: экономика – политика – культура*. – 2019. – № 9. – С. 63-78.
 21. Ермаков В.А., Бажданов Ю.В. Влияние средств массовой информации на формирование профессиональных иллюзий студентов // *Интерактивная наука*. – 2016. – № 9. – С. 56-59. – DOI [10.21661/r-114447](https://doi.org/10.21661/r-114447)
 22. Sutter M., Zoller C., Glätzle-Rützler D. Economic behavior of children and adolescents – A first survey of experimental economics results // *European Economic Review*. – 2019. – Vol.111. – P. 98-121. – DOI: [10.1016/j.eurocorev.2018.09.004](https://doi.org/10.1016/j.eurocorev.2018.09.004)
 23. Lekfuangfu W.N., Odermatt R. All I have to do is dream? The role of aspirations in intergenerational mobility and well-being. Discussion Paper Series, 2020. – 40 p. – Available at: https://www.cream-migration.org/publ_uploads/CDP_16_20.pdf (accessed date: July 27, 2022)
 24. Россуис Ю.Г. Эмилио Бетти: от истории права к общей теории интерпретации / Ю.Г.Бетти // *Философская мысль*. – 2020. – № 11.– С. 30–44.
 25. Аузан А. А. Социокультурная экономика: как культура влияет на экономику, а экономика – на культуру: курс лекций. – М.: Экономический факультет МГУ имени М. В. Ломоносова, 2021. – 200 с.

References in Cyrillics

1. Brockmeyer J., Harre R. Narratio: Problemata et promissiones alterutrius paradigma. *Quaestiones Philosophiae*, 2000, no. 3, pp. 29-42. (in Russian).
2. Shiller R.J. Narrative Economics and Neuroeconomics. *Finance: Theory and Practice*, 2018, Vol.22, iss.1, pp. 64-91. DOI: [10.26794/2587-5671-2018-22-1-64-91](https://doi.org/10.26794/2587-5671-2018-22-1-64-91)
3. Shiller R.J. *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton, Princeton University Press publ., 2019. 378 p. DOI :[10.1017/S1053837221000110](https://doi.org/10.1017/S1053837221000110)
4. Akerlof G., Snower D. Bread and bullets. *Journal of Economic Behavior & Organization*, 2016. Vol. 126, iss. PB, pp. 58–71. DOI : [10.1016/j.jebo.2015.10.021](https://doi.org/10.1016/j.jebo.2015.10.021)
5. Urhammer E., Røpke I. Macroeconomic narratives in a world of crises: An analysis of stories about solving the system crisis. *Ecological Economics*, 2013. Vol. 96. pp. 62-70. DOI: [10.1016/j.ecolecon.2013.10.002](https://doi.org/10.1016/j.ecolecon.2013.10.002)
6. Leipold S., Weldner K., Hohl M. Do we need a 'circular society'? Competing narratives of the circular economy in the French food sector. *Ecological Economics*. 2021. Vol. 187. 107086. DOI : [10.1016/j.ecolecon.2021.107086](https://doi.org/10.1016/j.ecolecon.2021.107086)
7. Bertsch Ch., Hull Is., Zhang X. Narrative fragmentation and the business cycle. *Economics Letters*, 2021. Vol. 201. 109783. DOI : [10.1016/j.econlet.2021.109783](https://doi.org/10.1016/j.econlet.2021.109783)
8. Liu S., Gaskell P., McGroarty Fr. Where and about what? Price relevant narratives depend on topic and media type. *Economics Letters*, 2022. Vol. 213. 110363. DOI : [10.1016/j.econlet.2022.110363](https://doi.org/10.1016/j.econlet.2022.110363)
9. Brown L.D., Call A.C., Clement M.B., Sharp N.Y. Managing the narrative: Investor relations officers and corporate disclosure. *Journal of Accounting and Economics*, 2019. Vol. 67, iss. 1, pp. 58-79. – DOI: [10.1016/j.jacceco.2018.08.014](https://doi.org/10.1016/j.jacceco.2018.08.014)
10. Asay H.S., Libby R., Rennekamp K. Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, 2018. Vol. 65, iss. 2-3, pp. 380-398. DOI: [10.1016/j.jacceco.2018.02.002](https://doi.org/10.1016/j.jacceco.2018.02.002)
11. Vossen A., Ihl Ch. More than words! How narrative anchoring and enrichment help to balance differentiation and conformity of entrepreneurial products. *Journal of Business Venturing*, 2020. Vol. 35, iss. 6. 106050. DOI: [10.1016/j.jbusvent.2020.106050](https://doi.org/10.1016/j.jbusvent.2020.106050)
12. Singh S., Corner P.D., Pavlovich K. Failed, not finished: A narrative approach to understanding venture failure stigmatization. *Journal of Business Venturing*, 2015. Vol. 30, iss. 1, pp.150-166. DOI: [10.1016/j.jbusvent.2014.07.005](https://doi.org/10.1016/j.jbusvent.2014.07.005)
13. Fisher G., Neubert E., Burnell D. Resourcefulness narratives: Transforming actions into stories to mobilize support. *Journal of Business Venturing*, 2021. Vol. 36, iss. 4. 106122. DOI: [10.1016/j.jbusvent.2021.106122](https://doi.org/10.1016/j.jbusvent.2021.106122)
14. Andre P., Haaland I., Roth Ch., Wohlfart J. Narratives About the Macroeconomy. *ECONtribute Discussion Papers Series*, 2022. Vol. 127. 104 p., available at: <https://repec.cepr.org/repec/cpr/ceprdp/DP17305.pdf> (accessed date: July 12, 2022)

15. Benabou R., Falk A., Tirole J. Narratives, Imperatives, and Moral Reasoning. *Collaborative Research Center Transregio 224*, 2019. 54 p., available at: <https://www.crctr224.de/en/research-output/discussion-papers/archive/2019/narratives-imperatives-and-moral-reasoning-roland-benabou-armin-falk-jean-tirole> (accessed date: July 25, 2022)
16. Hillenbrand A., Verrin E. The asymmetric effect of narratives on prosocial behavior. *Games and Economic Behavior*, 2022. Vol.135, pp. 241-270. DOI: [10.1016/j.geb.2022.06.008](https://doi.org/10.1016/j.geb.2022.06.008)
17. Tambovtsev V.L. Ideae, narrationes et mutationes in economia. *Terra Economicus*, 2019, no. 1, pp. 24-40. DOI: 10.23683/2073-6606-2019-17-1-24-40. (in Russian).
18. Volchik V.V., Maslyukova E.V., Narrationes, notiones et instituta. *Terra Economicus*, 2018, no. 2, pp. 150-168. (in Russian).
19. Volchik V.V. Narrationes et intellectus institutionum oeconomicarum. *Terra Economicus*, 2020, no. 2, pp. 49-69. DOI: 10.18522/2073-6606-2020-18-2-49-69 (in Russian).
20. Patronus P.A., Shmeleva A.Yu. Partes narrationes in consilio oeconomico. *USA & Canada: Oeconomica – Politica – Cultura*, 2019, no. 9, pp. 63-78. (in Russian).
21. Ermakov V.A., Bazhdanov Yu.V. Influxus instrumentorum communicationis in formatione studentium fallaciarum professionalium. *Scientia Interactive*, 2016, no. 9, pp. 56-59. DOI 10.21661/r-114447
22. Sutter M., Zoller C., Glätzle-Rützler D. Economic behavior of children and adolescents – A first survey of experimental economics results. *European Economic Review*, 2019. Vol.111, pp. 98-121. DOI: [10.1016/j.euroecorev.2018.09.004](https://doi.org/10.1016/j.euroecorev.2018.09.004)
23. Lekfuangfu W.N., Odermatt R. All I have to do is dream? The role of aspirations in intergenerational mobility and well-being. *Discussion Paper Series*, 2020. 40 p. available at: https://www.cream-migration.org/publ_uploads/CDP_16_20.pdf (accessed date: July 27, 2022)
24. Rossius Yu.G. Emilio Betti: Historiae iuris generalis interpretandi theoria. *Philosophica Cogitatio*, 2020, no.11, pp. 30 – 44. (in Russian).
25. Auzan A. A. *Economica Socio-culturalis: quomodo cultura afficit oeconomiam et oeconomiam culturam: cursus lectionum*, Moscow, Facultas Oeconomicorum Universitatis Civitatis Moscuensis nuncupata ab M.V. Lomonosov publ., 2021, 200 p. (in Russian).

Сведения об авторах

Апенко Светлана Николаевна – д-р экон. наук, профессор,
заведующий кафедрой менеджмента и маркетинга.
Омский государственный университет им. Ф.М. Достоевского (Омск, Россия)
Адрес для корреспонденции: 644077, Россия, Омск, пр. Мира, 55а
E-mail: apenkosn@yandex.ru
ORCID: 0000-0002-7618-3961,
Scopus AuthorID: 57189324004
Web of Science ResearcherID: D-1661-2015
ПИНЦ AuthorID: 261581

Лукаш Александр Викторович – кандидат философских наук,
доцент кафедры связи с общественностью, сервис и туризм»
Омский государственный университет путей сообщения (ОмГУПС).
Адрес для корреспонденции: Маркса пр., д. 35, г. Омск, 644046, Российская Федерация.
Тел.: +7 (3812) 31-18-33
E-mail: Lukashs2017@bk.ru
ORCID: 0000-0002-3468-141X
Web of Science ResearcherID: GLU-5137-2022
ПИНЦ AuthorID: 525231

About the authors

Svetlana N. Apenko – Doctor of Economic Sciences,
Professor, Head of the Department of Management
and Marketing.
Dostoevsky Omsk State University (Omsk, Russia)
Postal address: 55a, Mira pr., Omsk, 644077, Russia
E-mail: apenkosn@yandex.ru
ORCID: 0000-0002-7618-3961
Scopus AuthorID: 57189324004
Web of Science ResearcherID: D-1661-2015
RSCI AuthorID: 261581

Lukash Alexander Viktorovich –
PhD, Associate professor of the Department «Public relations, service and tourism»
Omsk State Transport University (OSTU).

Postal address: 35, Marx st., Omsk, 644046, the Russian Federation.

Phone: +7 (3812) 31-18-33

E-mail: Lukashs2017@bk.ru

ORCID: 0000-0002-3468-141X

Web of Science ResearcherID: GLU-5137-2022

RSCI AuthorID: 525231

Ключевые слова

Нарратив, молодежь, цифровое поколение, индоктринация, экономическое поведение.

, Svetlana Apenko, Alexander Lukash, Indoctrination of economic narratives for young people as a digital general society by federal statutory standards of higher education 3++

Keywords

Narrative, youth, digital generation, indoctrination, economic behavior

DOI: 10.34706/DE-2023-01-09

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract

The article is based on the interpretation of universal competences of the federal educational standards of generation 3++ in the areas of bachelor, master and specialist degrees, which are considered by the authors as narratives formalized in normative texts, and the narrative structures of the key youth forums in Russia since 2019 are analyzed. The aim of the work is to establish, on the basis of interpretation of narratives indoctrinated by the state through the system of higher education, benchmarks of socio-economic behavior for young people, implemented both within their work activities and beyond the boundaries of the professional sphere. As a result, the following narratives were established: networking in social and economic practices; personal responsibility for the result; project thinking and organization in professional activities; continuous formal and informal education in professional activities; responsible financial and economic behavior; stigmatization of corrupt behavior and exclusive use of modern information technologies to solve problems of professional activities. The interpretation of narratives of educational standards and project sessions of youth activities has established that both formalized and non-formalized narrative structures are aimed at entrenching among young people ideas about the development of entrepreneurial initiative, regardless of professionalization of the graduate, the coordination of individual and social responsibility, the importance of maintaining and developing bridging social capital. The article shows that the state through the system of higher education and youth policy programs translates for modern students and graduates the narratives of economic and labor behavior focused on the trajectory of innovation and digital development, as well as, to remove the contradictions between prosocial and egocentric strategies of behavior.

2.2. БЛОКЧЕЙН И БОЛЬШИЕ ДАННЫЕ В ФИНАНСОВОЙ СФЕРЕ

Родионов А. А.¹

¹ООО «СКМ», директор по информационным технологиям
Москва, Россия,

Финансовый сектор сегодня – это активное использование информационных технологий при выполнении самых различных операций. Это обусловлено стремлением к повышению качества выполняемых процедур, обеспечением безопасности осуществляемых операций, снижению затрат и рядом иных факторов, конечный выбор которых зависит от конкретной ситуации и требований организации. Наиболее обсуждаемыми технологиями в последнее время стали блокчейн и большие данные, которые всё активнее применяются в самых различных отраслях, включая и сферу финансов. Новизна исследования заключается в рассмотрении обобщенных аспектов использования двух различных технологий в финансовой сфере как по отдельности, так и совместно. Практическая значимость исследования заключается в выполнении актуального анализа совокупного использования технологий «Блокчейн» и «Большие данные» в рамках функционирования отечественных организаций и предприятий финансовой сферы.

Введение

Деятельность финансового сектора подразумевает активное использование современных Fintech-технологий, позволяющих совершать онлайн-платежи без открытия банковского счета с помощью платежной системы. Продукты в рамках Fintech технологий чаще всего реализуются организацией, которая впоследствии будет их использовать в собственной деятельности с целью упрощения реализации рабочих процедур. Актуальность выбранной темы исследования в современном мире может быть описана как анализ потенциальных возможностей совместного использования данных технологий.

Методология исследования

Цель исследования – рассмотреть основные аспекты использования технологий «Большие данные» и «Блокчейн» в деятельности организаций и предприятий финансовой сферы. Объект исследования – современные информационные технологии в деятельности финансовых организаций. При написании статьи использовался метод анализа научных источников и публикаций. Методология исследования заключается в анализе возможностей совместного применения технологий «Блокчейн» и «Большие данные», с учетом их особенностей реализации и работы. В частности, в большинстве исследований отмечают факт наличия практически противоположных требований по обеспечению работы данных технологий. Планируется рассмотреть особенности работы каждой технологии по отдельности и представить возможные пути их совместного использования.

Результаты и обсуждение

Активное развитие Fintech-платформ произошло в 2008 году, и было спровоцировано двумя факторами – это стремление сократить расходы организациями на фоне финансового кризиса, и стремительный рост мобильных технологий, что предоставило постоянный доступ к мобильному интернету большому числу людей. Сейчас современные финансовые технологии являются одними из наиболее быстрорастущих рынков в мире. К числу популярных технологий в данной сфере относятся «Блокчейн» и «Большие данные», речь о которых в разрезе финансовой отрасли будет идти в рамках данной статьи.

Блокчейн представляет собой технологию распределенного общего реестра, в рамках которой наблюдается реализация открытого контроля проводимых транзакций. За счет данной организации любой из участников бизнес-сети может просмотреть и выполнить проверку корректности и действительности выполнения транзакции. Даже в случае осуществления одновременной записи одной транзакции в нескольких распределенных базах данных, конструкция технологии обеспечит её безопасность. Технология блокчейн представляет собой книгу с децентрализованными данными, информационный обмен которыми происходит между участниками сети с применением безопасных каналов [7]. Для этого передаваемая информация дробится на общие блоки, объединение которых формируется посредством уникальных идентификаторов, выполненных в виде криптографических хеш-функций. Технология блокчейна позволяет обеспечить целостность данных с использованием единого источника достоверной информации, за счет чего попросту сводится на нет возможность дублирования данных. Все вышеперечисленное делает невозможным осуществление каких-либо мошеннических действий, так как любые изменения записей возможны только на основании кворума сторон, а при попытке выполнения их несанкционированно, всем участникам будет отправлено соответствующее уведомление. Таким образом, блокчейн представляет собой технологию, обеспечивающую гарантированно безопасный обмен информацией [7].

Применение данной технологии в сфере финансов чаще всего осуществляется для того, чтобы добиться дешевизны, высокой скорости и безопасности выполняемых транзакций. В качестве примера можно назвать операцию перевода денежных средств между двумя клиентами. Как правило, это довольно простая и быстрая процедура, однако в ряде ситуаций на её осуществление может потребоваться несколько дней. Использование при её осуществлении технологии блокчейн позволит добиться существенного роста скорости выполняемых операций, а также минимизировать размер комиссии за её выполнение. Еще одним примером следует назвать обеспечение совместного хранения и использования данных. В большинстве случаев каждый банк формирует и использует в работе собственное внутреннее хранилище данных, работающее с другими банками лишь поверхностно. Это требует значительных финансовых затрат на его реализацию и обслуживание. Применение технологии блокчейн может позволить реализовать единое или совместное хранилище, которое будет использоваться несколькими финансовыми организациями. Результатом такого действия станет существенная экономия на расходах по обслуживанию хранилища, обеспечение достоверности данных, а также обеспечение гарантированной их безопасности. Также технология блокчейн может быть использована для организации процедуры идентификации личности клиента, что упростит жизнь как клиентам, так и финансовым организациям [9].

Блокчейн довольно молодая технология, которая большинству рядовых обывателей известна лишь на примере рынка криптовалюты, однако это далеко не так. Он может быть использован в различных отраслях и сферах деятельности, хотя несомненную популярность получил именно в финансовой сфере. Это отлично демонстрирует график, построенный на основании результатов опроса компании Deloitte.

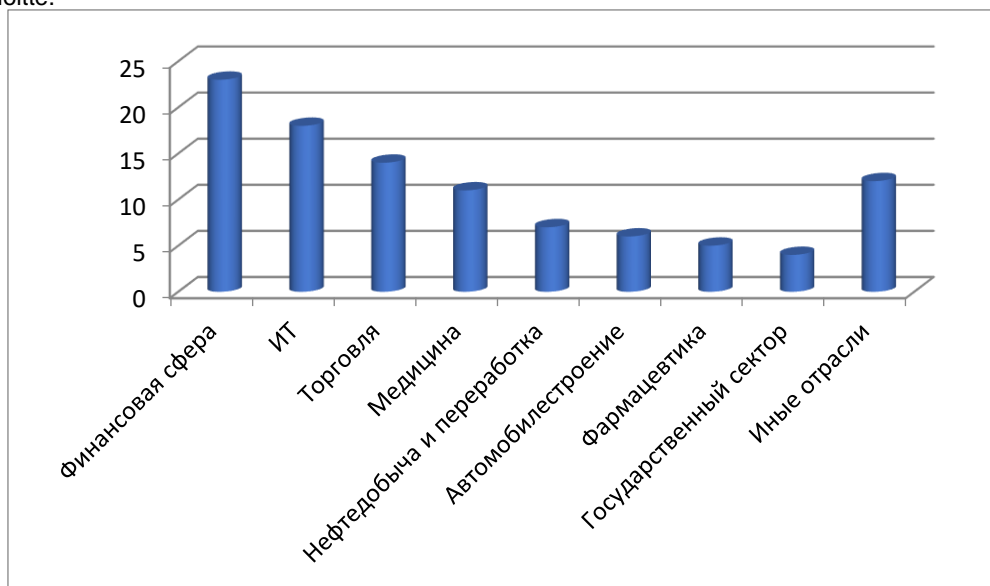


Рис. 1. Статистика использования технологии блокчейн в различных отраслях

Именно по причине популярности технологии блокчейн в сфере финансов, большинством экономик активно рассматривается вопрос нормативно-правового регулирования и развития данной технологии. В РФ, например, 1 июля 2016 года был создан первый консорциум по развитию технологии блокчейн, в состав которого вошли крупные финансовые организации.

Наиболее ярким примером популяризации технологии блокчейн следует назвать представление Сбербанком 20 мая 2021 года собственной блокчейн-платформы, предоставляющей в общедоступном формате разработчикам полнофункциональный API, а также библиотеки для организации работы с токенами. Функциональным назначением сервиса является реализация легких клиентских приложений, которые смогут напрямую функционировать с блокчейн-платформой, используя в операциях специальную расчетную единицу, интегрированную с банком. Это позволит реализовывать расчетные операции на основе рублевых смарт-контрактов. Обеспечение распределенной структуры платформы стало возможным за счет привлечения независимых участников-валидаторов, которые организуют проверки корректности транзакций, не имея доступа к конфиденциальным сведениям в данных транзакциях [1].

Переходя к технологии больших данных необходимо отметить, что она представляет собой инструментальный по обработке больших объемов информации, которая поступает с высокой скоростью и постоянно растет, при этом еще и выражается в самой различной форме. Объем обрабатываемых данных является настолько большим, что их невозможно обработать стандартным программным обеспечением на обычных персональных компьютерах, однако результат их обработки могут быть применены для успешного решения бизнес-задач, которые ранее решить столь быстро и просто не представлялось возможным. Сейчас большие данные стали своего рода разновидностью капитала, так как

позволяют повышать эффективность работы компаний и корпораций за счет объективной оценки и предоставлению результатов, нацеленных на качественное решение задач. Все это, в совокупности с активным ростом достижений в сфере технологий делает использование больших данных более простым и не столь затратным, по причине чего крупные организации, в том числе и в финансовой сфере, все активнее прибегают к их использованию [4].

Использование больших данных все активнее наблюдается в таких операциях, как риск-моделирование, поведенческий анализ клиентов, также обеспечение задач информационной безопасности и противодействия мошенникам. К отдельной категории вариантов использования больших данных следует отнести автоматизацию работы с запросами клиентов при реализации чат-ботов, разработку активных каналов коммуникации с клиентами на основании систем класса CRM, а также разработку моделей кросс продаж и оценки тарифов для клиентов. Некоторые финансовые организации прибегают к большим данным при выполнении анализа транзакций их клиентов с целью формирования кредитных рейтингов, а также осуществления мониторинга рисков контрагентов. За счет реализации механизмов новостной аналитики происходит выявление взаимосвязей между заемщиками.

Банком России в 2021 году был проведен опрос, в котором крупные российские финансовые институты отвечали на вопросы в отношении применения больших данных в их деятельности. На рисунке 2 представлена гистограмма, демонстрирующая оцененные преимущества в результате использования данной технологии.



Рис. 2. – Гистограмма оценки преимуществ от использования больших данных в финансовом секторе

Таким образом, необходимо отметить, что большинством опрошенных технология больших данных была оценена положительно, а её применение в деятельности организации связано с рядом существенных преимуществ.

Процессы в отношении обработки и хранения больших данных организованы в большинстве финансовых организации на основании единого хранилища данных, работу с которым выполняет отдельное подразделение. Далее в каждом бизнес-блоке или секторе назначаются специалисты либо подразделения, ответственные за работу с данными.

Технология блокчейн и технология больших данных предоставляют финансовым организациям несомненные преимущества в отношении организации рабочих процессов различного уровня. При этом важно отметить факт принципиального различия данных технологий. Первая организует хранение информации на множестве различных компьютеров, вторая же наоборот, требует размещения данных в едином информационном центре. Это приводит и к различным характеристикам работы технологий. В рамках блокчейна обработка информации осуществляется медленнее, чем при работе с большими данными, но при этом обеспечивается более высокий уровень защиты. Большие данные же в своей концептуальной основе направлены на быструю обработку информации.

Информация, поступающая в систему на основании блокчейн-технологии, остается в ней навсегда. По этой причине применение данной технологии имеет смысл в решении задач, требующих постоянного хранения неизменяемой информации. А специфика организации информационно обмена делает эту технологию оптимальной в решении задач, где необходимо обеспечивать высокий уровень достоверности и неизменности информации. И яркий пример тому – решение вопросов информационной безопасности, при которых использование блокчейна может обеспечить решение проблем доверия к данным и организации достоверного обмена ими [5].

Получается, что данные технологии обладают разительными отличиями, и в совокупности их использовать очень сложно, однако по раздельности они предлагают финансовым организациям целый

ряд конкурентных преимуществ. Несмотря на данный факт, уже имеют место быть отдельные наработки, связанные с попытками объединения технологий блокчейна и больших данных для работы в данной отрасли.

С целью обеспечения должного уровня конкурентоспособности финансовыми организациями осуществляется разработка новых моделей, с применением которых осуществляется анализ различного рода рисков, привлекательности проектов и т.д. Содержание технологии, используемых для решения подобного рода задач довольно обширно, однако сейчас все активнее для этого используется как раз технология больших данных. Обеспечение представленных задач реализуется с целью организации сервисного обслуживания в соответствии с запросами клиента, а также обеспечения конкурентоспособности организации.

В зависимости от уровня внедрения технологий больших данных следует отметить, что различные направления деятельности финансовых организаций демонстрируют различный уровень работы в том или ином направлении. Так, в рамках риск-менеджмента и работы с клиентами необходимо отметить широкое внедрение подобного рода технологий больших данных. Сюда же следует отнести вопросы ценообразования. Рынок трейдинга, управления активами и проведения платежей в свою очередь вопросы применения больших данных держит на стадии обработки, и больше направлен на использование технологии блокчейна для обеспечения безопасности транзакций.

Современные технологии, к которым относятся распределенные реестры и большие данные предоставляют большое число различного рода преимуществ организациям, их использующим. Это демонстрирует активное их использование в различных сферах, в том числе и в сфере финансов, занимающей лидирующую позицию по объемам внедрения данных технологий в свою деятельность. Недостатком использования данных технологий является их новизна – в большинстве случаев невозможно реализовать ожидаемый функционал, так как чаще всего он описан сугубо в теории, и никто не имеет представления, как фактически реализовать его на практике.

Литература

1. Винья П. Машина правды. Блокчейн и будущее человечества. – М.: МИФ, 2018. -320 с.
2. Генкин А.С. Блокчейн. Как это работает и что ждет нас завтра. – М.: Альпина Паблишер, 2018. – 592 с.
3. Использование больших данных в финансовом секторе и риски финансовой стабильности: Доклад ЦБ РФ. – Москва, 2021.
4. Особенности использования Blockchain и Big Data. – URL: <https://maddata.agency/blog/blockchain-i-big-data-cto-neobkhodimo-znat> (дата обращения: 02.02.2023).
5. Потенциал больших данных для финансовой сферы. – URL: https://www.cfin.ru/itm/olap/opportunities_in_big_data.shtml (дата обращения: 02.02.2023).
6. Применение технологий Big Data в банковской сфере. – URL: http://secuteck.ru/articles2/sys_ogr_dost/primenenie-tehnologiy-big-data-v-bankovskoy-sfere (дата обращения: 02.02.2023).
7. Арефьев П. В., Восканов А. С., Гришин М. С. Технология блокчейн в финансовом секторе экономики // Стратегии бизнеса. 2019. №10 (66). URL: <https://cyberleninka.ru/article/n/tehnologiya-blokcheyn-v-finansovom-sektore-ekonomiki> (дата обращения: 07.02.2023).
8. Что такое блокчейн: все, что нужно знать о технологии. – URL: <https://www.forbes.ru/mneniya/456381-cto-takoe-blokcejn-vse-cto-nuzno-znat-o-tehnologii> (дата обращения: 02.02.2023).
9. Что такое блокчейн? – URL: <https://www.oracle.com/cis/blockchain/what-is-blockchain/> (дата обращения: 02.02.2023).

References in Cyrillics

1. Vin'ya P. Mashina pravdy. Blokcheyn i budushchee chelovechestva. – М.: MIF, 2018. -320 s.
2. Genkin A.S. Blokcheyn. Kak eto rabotaet i chto zhdet nas zavtra. – М.: Al'pina Pablisher, 2018. – 592 s.
3. Ispol'zovanie bol'shih dannyh v finansovom sektore i riski finansovoj stabil'nosti: Doklad CB RF. – Moskva, 2021.
4. Osobennosti ispol'zovaniya Blockchain i Big Data. – URL: <https://maddata.agency/blog/blockchain-i-big-data-cto-neobkhodimo-znat> (data obrashcheniya: 02.02.2023).
5. Potencial bol'shih dannyh dlya finansovoj sfery. – URL: https://www.cfin.ru/itm/olap/opportunities_in_big_data.shtml (data obrashcheniya: 02.02.2023).
6. Primenenie tekhnologij Big Data v bankovskoy sfere. – URL: http://secuteck.ru/articles2/sys_ogr_dost/primenenie-tehnologiy-big-data-v-bankovskoy-sfere (data obrashcheniya: 02.02.2023).

7. Aref'ev P. V., Voskanov A. S., Grishin M. S. Tekhnologiya blokchejn v finansovom sektore ekonomiki // Strategii biznesa. 2019. №10 (66). URL: <https://cyberleninka.ru/article/n/tehnologiya-blokchejn-v-finansovom-sektore-ekonomiki> (data obrashcheniya: 07.02.2023).
8. Chto takoe blokchejn: vse, chto nuzhno znat' o tekhnologii. – URL: <https://www.forbes.ru/mneniya/456381-cto-takoe-blokcejn-vse-cto-nuzhno-znat-o-tehnologii> (data obrashcheniya: 02.02.2023).
9. Chto takoe blokchejn? – URL: <https://www.oracle.com/cis/blockchain/what-is-blockchain/> (data obrashcheniya: 02.02.2023).

Ключевые слова

Блокчейн, большие данные, Fintech, финансовая сфера, информационные технологии

Родионов Андрей Анатольевич, директор по информационным технологиям,
ООО «СКМ», Москва, Россия,
andrey.mailmsk@gmail.com

Keywords

Blockchain, big data, Fintech, financial sector, information technology

Andrey A. Rodionov¹

¹ LLC SCM, Director of Information Technology
Moscow, Russia
andrey.mailmsk@gmail.com

Andrey Rodionov, Blockchain and Big data in the financial sphere

DOI: 10.34706/DE-2023-01-10

JEL classification C02 – Математические методы; M15 Управление информационными

Abstract

The financial sector today is the active use of information technology in a variety of operations. This is due to the desire to improve the quality of procedures performed, ensure the safety of ongoing operations, reduce costs and a number of other factors, the final choice of which depends on the specific situation and the requirements of the organization. The most discussed technologies in recent years have been blockchain and big data, which are increasingly being used in a wide variety of industries, including finance. The novelty of the study lies in the consideration of generalized aspects of the use of two different technologies in the financial sector, both separately and jointly. The practical significance of the study lies in the implementation of an up-to-date analysis of the combined use of Blockchain and Big Data technologies in the framework of the functioning of domestic organizations and enterprises in the financial sector.

2.3. РЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ И ИХ ОСОБЕННОСТИ

Калинина Е.С.¹, Манохина Т. В.², Ступаков С. А.³

¹²³ФГБОУ ВО «Омский государственный университет путей сообщения», Омск, Россия

В статье рассмотрена проблематика изучения реляционных баз данных и их особенностей. Установлено, что на данном этапе, концепцию баз данных NoSQL поддерживают следующие типы баз данных: хранилища ключ / значение (key-value stores), колонка-ориентированные хранилища (wide column stores), документарные хранилища (document stores) и графы базы данных (graph databases). Исследованы особенности применения технологии реляционных баз данных. Во-первых, они позволяют и систематизировать, и обобщать большие массивы данных в рамках проведения научных исследований, поиска научной информации, статей, метаданных и тому подобное. Во-вторых, технология базы данных позволяют сегментировать и структурировать информацию по определенным критериям

Введение

С каждым годом развитие информационных технологий во всех сферах деятельности человечества вызывает необходимость поиска и разработки новых способов их использования для анализа и и обработки информации [1]. Развитие интернета, появление языка разметки XML в 90-х годах XX в. и массовый переход пользователей на работу с веб-технологиями привели к накоплению в информационных системах и хранилищах данных полуструктурированных и неструктурированных данных, что значительно затрудняло работу реляционных баз данных. В это время появился новый вид баз данных – базы данных XML (NXD), которые поддерживали структуру и язык запросов XML – как базовую. Примерами этого вида баз данных можно считать MarkLogic, Virtuoso, Sedna, searchxml, eXist-db [2; 3]. Но, как и объектно-ориентированные базы данных, этот вид не получил широкого распространения и заменить собой базы данных SQL. Тем не менее, существенная доля современных реляционных баз данных используют и поддерживают язык XML. Тем не менее, существенная доля современных реляционных баз данных используют и поддерживают язык XML. В частности, в пятой версии стандарта SQL: 2003 отдельный раздел посвящен спецификации SQL / XML, определяющий правила взаимодействия между схемами XML и структурами SQL и функции обработки XML-данных через SQL-запросы [5].

Цель статьи – исследование особенностей реляционные базы данных.

Основная часть.

Начиная с 50-х годов и до конца XX в. концепции организации баз данных эволюционировали от файловых систем хранения данных на основе использования перфокарт к объектно-ориентированным и традиционным реляционным баз данных. Наиболее стабильные позиции и высокий уровень востребованности среди пользователей и разработчиков в конце XX в. получили традиционные базы данных, через поддержку языка запросов SQL и реляционной модели представления данных получили название баз данных SQL. Это целый класс баз данных и систем управления базами данных, которые активно используются и сейчас (Microsoft SQL, Oracle, MySQL, Postgres, IBM Db2 т.д.) [4; 5].

Большинство имеющихся на рынке баз данных SQL относится к типу проприетарного (закрытого, защищенного коммерческой лицензией) программного обеспечения, то есть для его использования необходимо приобрести у вендора лицензию на использование базы данных. Среди самых распространенных можно определить Oracle, Microsoft SQL Server, IBM Db2, Microsoft Access.

Доля открытых (open source databases) баз данных SQL значительно меньше, чем проприетарных на рынке, однако в последнее время их количество имеет тенденцию к росту. Наиболее распространенными считаются MySQL, SQLite и PostgreSQL [4].

В начале XXI в. состоялось переосмысление концепции реляционных баз данных на фоне появления новых методов добычи данных, развития концепций хранилищ данных, а также из-за проблем баз данных SQL по работе с неструктурированными данными, сложностью агрегирования, невысокой производительностью работы и др. Альтернативной концепцией баз данных стала концепция нереляционных или NoSQL баз данных. В противоположность традиционным реляционным базам данных, базы данных NoSQL способны обеспечивать высокую производительность, горизонтальную масштабируемость и гибкость в работе с неструктурированными и значительными по объемам данными.

На данном этапе, концепцию баз данных NoSQL поддерживают следующие типы баз данных [6]:

- хранилища ключ / значение (key-value stores),
- колонка-ориентированные хранилища (wide column stores),
- документарные хранилища (document stores),
- графы базы данных (graph databases).

Как подвид графов баз данных выделяют базы данных RDF, которые используют для хранения метаданных и представления их в виде семантической паутины. К числу типичных представителей этого подвида графов базы данных относятся: Apache Jena-TDB, Redland, RDF4J, CubicWeb.

Некоторые ученые склонны выделять еще один тип – поисковые базы данных (search engines database), хотя их отношение к типам NoSQL является условным. С одной стороны, поисковые базы данных не имеют жестких структурных требований к сохранению и представлению данных, что характерно для NoSQL, с другой – в них используется индексирование и язык запросов, восточнее на SQL. Почти все базы данных NoSQL имеют открытую лицензию и относятся к open source databases. Кроме того, базы данных NoSQL принято разделять по типу и среде обработки данных на [7]:

- базы данных, которые операции выполняют в оперативной памяти системы (in-memory),
- базы данных, работающие с использованием SSD или жесткими дисками (HDD).

Стоит добавить, что значительная часть баз данных NoSQL активно используются в своей работе облачные веб-технологии, работающие в режиме real-time и поддерживающие обмен данными через XML и JSON. К наиболее распространенным баз данных NoSQL можно отнести: MongoDB, Riak, Redis, Splunk, Cassandra, Hbase, Neo4j, Elasticsearch [8; 9].

Актуальными проблемами современных баз данных NoSQL считаются проблемы согласованности данных, функциональной ограниченности в реализации сложных запросов, точность результатов обработки. Появление концепции NoSQL и востребованность нереляционных баз данных на рынке привела к тому, что разработчики реляционных баз данных начали использовать элементы концепции и подходы по работе с крупными неструктурированными данными в свои базах данных, которые получили название NewSQL или новые реляционные базы данных. Они лучше масштабируются, чем традиционные реляционные базы, имеют высокую производительность работы, обеспечивают высокое качество в работе с данными за счет поддержки онлайн-транзакций.

Базы данных NewSQL поддерживают обмен различными видами данных, работу с облачными веб-технологиями, работающих в режимах real-time и streaming, обладают способностью фиксировать текущие и пространственные данные, поддерживают обмен данными через XML, JSON и др. Примерами баз данных типа NewSQL можно считать: ClustrixDB, VoltDB, SQL Server, NuoDB, CockroachDB. Сейчас на рынке программного обеспечения и информационных систем существует значительная конкуренция между реляционной, NoSQL и NewSQL концепциями баз данных. В результате этой конкуренции начали появляться базы данных, в которых реализуется поддержка как реляционных, так и нереляционных моделей данных [10]. Причем, такое сочетание происходит не на уровне технологий работы с данными, хотя и это имеет место, а на уровне разработки так называемого «мультихранилища» – надстройки, которая может управлять одновременно базами данных с реляционными и нереляционными моделями данных. Таким образом, современными подходами к дальнейшей эволюции концепции баз данных являются:

- Многомодельные базы данных (multi-model databases), которые реализуются на базе использования одного единственного интегрированного сервера, который обеспечивает поддержку нескольких типов моделей данных. Например, Oracle Database – система управления базами данных, основанный на многомодельном подходе, который поддерживает работу с реляционными, графовые и колонки-ориентированными моделями данных, ArangoDB – обеспечивает поддержку моделей ключ / значение, документо-ориентированные и графовые модели данных.

Можно выделить два основных вида многомодельных баз данных:

1) базы данных построены на основе поддержки конкурентной модели данных SQL (или NoSQL, или NewSQL) и дополненные отдельными типами поддержки других моделей данных;

2) базы данных построены на основе поддержки различных моделей данных. К этому виду относятся такие базы данных, имеющие равноценный функционал, характерный для как для NewSQL, так и для NoSQL.

В общем, многомодельность значительно упрощает интеграцию данных в середине хранилища;

Мультихранилища (polystore databases) – представляют собой так называемую «витрину данных», которая обеспечивает доступ интегрированных между собой баз данных на основе прямого доступа и использовании локальных языков запросов, что характерных для конкретной модели данных. Концепция организации мультихранилищ заключается в использовании различных типов баз данных для обработки различных потребностей на основе промежуточного программного обеспечения – интерпретаторов.

Выводы

Сбор и анализ увеличивающихся по объему и разнообразию данных превращают процесс их традиционного анализа в технологии обработки этих данных. Изменение и направленность использования больших бах данных зависят от целей организации и решаемых задач.

Исследованы особенности применения технологии реляционных баз данных. Во-первых, они позволяют и систематизировать, и обобщать большие массивы данных в рамках проведения научных исследований, поиска научной информации, статей, метаданных и тому подобное. Во-вторых, технология базы данных позволяют сегментировать и структурировать информацию по определенным критериям.

Литература

1. Васильева К. Н., Хусаинова Г. Я. Реляционные базы данных // Colloquium-journal. – 2020. – №2 (54). – URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh> (дата обращения: 03.02.2023).

2. Глумов А. А. Теоретический аспект технологической составляющей новой индустриализации: промышленный интернет / А. А. Глумов // Вестник НГИЭИ. – 2018. – № 5.- С. 147-156.
3. Дейт, К. Дж. Введение в системы баз данных, 8-е издание: пер. с англ. / Дж. К. Дейт. – М.: Издательский дом «Вильямс», 2005. – 1328 с.
4. Жук М.М. РЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ, ЯЗЫК SQL // StudNet. – 2022. – №6. – URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh-yazyk-sql> (дата обращения: 03.02.2023).
5. Иванов В.И., Новиков С.П. Программное средство исследования различных алгоритмов кэширования в реляционных СУБД // Молодой исследователь Дона. – 2018. – №3 (12). – URL: <https://cyberleninka.ru/article/n/programmnoe-sredstvo-issledovaniya-razlichnyh-algoritmov-keshirovaniya-v-relyatsionnyh-subd> (дата обращения: 03.02.2023).
6. Реляционная база данных и ее структура / Научная библиотека избранных естественно-научных изданий. – URL: http://sernam.ru/book_cbd.php?id=2 (дата обращения 04.02.2023).
7. Савоськин И.В., Фирсов А.О. Исследование способов применения NoSQL и реляционных баз данных // E-Scio. – 2019. – №6 (33). – URL: <https://cyberleninka.ru/article/n/issledovanie-sposobov-primeneniya-nosql-i-relyatsionnyh-baz-dannyh> (дата обращения: 03.02.2023).
8. Федорова Г. Н. Разработка и администрирование баз данных / Г. Н. Федорова. – М.: Академия, 2015. – 58с.
9. Хлопотов Роман Сергеевич ОСОБЕННОСТИ ПРОЕКТИРОВАНИЯ БАЗ ДАННЫХ ДЛЯ АВТОМАТИЗИРОВАННОГО РАБОЧЕГО МЕСТА ВРАЧА-НУТРИЦИОЛОГА // Известия ТулГУ. Технические науки. – 2022. – №9. – URL: <https://cyberleninka.ru/article/n/osobennosti-proektirovaniya-baz-dannyh-dlya-avtomatizirovannogo-rabochego-mesta-vracha-nutritsiologa> (дата обращения: 03.02.2023).
10. Brewer Eric A. Towards robust distributed systems // Proceedings of the XIX annual ACM symposium on Principles of distributed computing. – Portland, OR: ACM, 2000. – Vol. 19.

References in Cyrillics

1. Vasil'eva K. N., Husainova G. YA. Relyacionnye bazy dannyh // Colloquium-journal. – 2020. – №2 (54). – URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh> (data obrashcheniya: 03.02.2023).
2. Glumov A. A. Teoreticheskij aspekt tekhnologicheskoy sostavlyayushchej novoj industrializacii: promyshlennyj internet / A. A. Glumov // Vestnik NGIEI. – 2018. – № 5.- S. 147-156.
3. Dejт, K. Dzh. Vvedenie v sistemy baz dannyh, 8-e izdanie: per. s angl. / Dzh. K. Dejт. – М.: Izdatel'skij dom «Vil'yame», 2005. – 1328 s.
4. ZHuk M.M. RELYACIONNYYE BAZY DANNYYH, YAZYK SQL // StudNet. – 2022. – №6. – URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh-yazyk-sql> (data obrashcheniya: 03.02.2023).
5. Ivanov V.I., Novikov S.P. Programmnoe sredstvo issledovaniya razlichnyh algoritmov keshirovaniya v relyacionnyh SUBD // Molodoy issledovatel' Dona. – 2018. – №3 (12). – URL: <https://cyberleninka.ru/article/n/programmnoe-sredstvo-issledovaniya-razlichnyh-algoritmov-keshirovaniya-v-relyatsionnyh-subd> (data obrashcheniya: 03.02.2023).
6. Relyacionnaya baza dannyh i ee struktura / Nauchnaya biblioteka izbrannyh estestvenno-nauchnyh izdaniy. – URL: http://sernam.ru/book_cbd.php?id=2 (data obrashcheniya 04.02.2023).
7. Savos'kin I.V., Firsov A.O. Issledovanie sposobov primeneniya NoSQL i relyacionnyh baz dannyh // E-Scio. – 2019. – №6 (33). – URL: <https://cyberleninka.ru/article/n/issledovanie-sposobov-primeneniya-nosql-i-relyatsionnyh-baz-dannyh> (data obrashcheniya: 03.02.2023).
8. Fedorova G. N. Razrabotka i administrirovanie baz dannyh / G. N. Fedorova. – М.: Akademiya, 2015. – 58s.
9. Hlopotov Roman Sergeevich OSOБENNOSTI PROEKTIROVANIYA BAZ DANNYYH DLYA AVTOMATIZIROVANNOGO RABOCHEGO MESTA VRACHA-NUTRICIOLOGA // Izvestiya TulGU. Tekhnicheskie nauki. – 2022. – №9. – URL: <https://cyberleninka.ru/article/n/osobennosti-proektirovaniya-baz-dannyh-dlya-avtomatizirovannogo-rabochego-mesta-vracha-nutritsiologa> (data obrashcheniya: 03.02.2023).
10. Brewer Eric A. Towards robust distributed systems // Proceedings of the XIX annual ACM symposium on Principles of distributed computing. – Portland, OR: ACM, 2000. – Vol. 19.

Ключевые слова

Базы данных, реляционные, особенности, изучение, применение

Калинина Екатерина Сергеевна, к. т. н, доцент кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения», Омск, Россия, ekkalinina@mail.ru

Манохина Татьяна Витальевна, старший преподаватель кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения»,
mtv-gups@mail.ru

Ступаков Сергей Анатольевич, к. т. н, доцент кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения», Омск, Россия,
stupakov1@yandex.ru

Keywords

Databases, relational, features, study, application

Kalinina Ekaterina Sergeevna, Candidate of Technical Sciences, Associate Professor of the Department of Informatics and Computer Graphics, Omsk State Transport University, Omsk, Russia,
ekkalinina@mail.ru

Manokhina Tatyana Vitalievna, Senior Lecturer of the Department of Informatics and Computer Graphics, Omsk State Transport University,
mtv-gups@mail.ru

Stupakov Sergey Anatolyevich, Candidate of Technical Sciences, Associate Professor of the Department of Informatics and Computer Graphics, Omsk State Transport University, Omsk, Russia,
stupakov1@yandex.ru

DOI: 10.34706/DE-2023-01-11

JEL classification C02 – Математические методы; M15 Управление информационными

RELATIONAL DATABASES AND THEIR FEATURES

The article deals with the problems of studying relational databases and their features. It has been established that at this stage, the concept of NoSQL databases is supported by the following types of databases: key / value stores (key-value stores), column-oriented stores (wide column stores), document stores (document stores) and database graphs (graph databases). The features of application of technology of relational databases are investigated. Firstly, they allow you to systematize and generalize large amounts of data in the framework of scientific research, search for scientific information, articles, metadata, and the like. Secondly, database technology allows you to segment and structure information according to certain criteria.

2.4. ВИЗУАЛИЗАЦИЯ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ

Калинина Е.С.¹, Манохина Т. В.², Ступаков С. А.³
¹²³ФГБОУ ВО «Омский государственный университет путей сообщения»,
Омск, Россия

В статье рассмотрена проблематика изучения визуализации баз данных реляционно-го типа. В статье выделены преимущества и недостатки документо-ориентированной СУБД MongoDB на примере создания системы визуализации данных реального времени. Создана система визуализации данных реального времени, предназначенная для консолидации и обработки информации о событиях в высоконагруженных веб-системах. Система использует документо-ориентированную СУБД MongoDB для хранения полученных данных и JavaScript для обработки событий на стороне сервера, запускаемого с применением платформы Node.js.

Введение

Рост объема данных, ориентированных на пользователей, вызвал быстрое увеличение объема и типов данных, которые создаются, отображаются, анализируются и архивируются. Кроме того, изменяется количество новых наборов источников данных, включая датчики, системы глобального позиционирования (GPS), автоматизированные трекеры и системы мониторинга, создающие большие массивы данных. Эти большие объемы наборов данных, часто называемых «большими данными» (англ. Big Data), создают новые вызовы и возможности вокруг хранения, анализа и архивирования [1].

Параллельно с быстрым ростом объемов данных данные также становятся все более структурированными и рассеянными. Это означает, что традиционные методы управления данными, основанные на предварительном определении схемы данных и ссылок между данными, уже не отвечают современным требованиям обработки и анализа больших разнотипных слабо структурированных данных [2].

Документо-ориентированная система управления базами данных (ДОСУБД) – это система, которая сохраняет коллекцию текстовых данных и позволяет создавать запросы и обновления объектов. Обычно база данных включает в себя множество документов, соотносящихся по предмету, происхождению или применению [3]. Содержание каждого документа может быть свободным текстом, частично структурированным текстом, включающим несколько четко определенных атрибутов (например, название, автор, дата), или хорошо структурированным текстом, который, например, может быть закодирован с использованием языка XML. Иногда документы могут содержать мультимедийные компоненты [4].

ДОСУБД могут применяться для сохранения большого объема данных: изменений курсов валют, акций и ценных бумаг, трекинга использования сайта (анализа журналов Интернет-серверов с большим количеством пользователей), которые могут использоваться для разработки системы мониторинга (визуализации данных) реального времени. Таким образом, ДОСУБД предоставляет новые информационные технологии для издательств, цифровых библиотек, электронного правительства и электронного бизнеса в целом [5]. Поскольку новые технологии обработки больших массивов слабо структурированной информации еще не исследованы, выбранная тема данной статьи представляется актуальной.

Методология исследования

Цель статьи – исследование визуализации баз данных реляционного типа.

Основная часть

Существует много причин использовать нереляционные СУБД, но можно выделить две основные причины:

– Производительность разработки приложений. Много усилий в области разработки программного обеспечения уходит на отображение модели и структуры данных, находящихся в оперативной памяти компьютера, в реляционных базах данных. NoSQL хранилища данных могут обеспечить модель данных, которая будет лучше отвечать потребностям разработчиков приложений, тем самым упрощая взаимодействие с СУБД и, в результате чего, уменьшая количество написанного кода, время на отладку и внедрение.

– Большие масштабы данных. Организации считают целесообразным хранить большие объемы данных и обрабатывать их быстрее. Сделать это оказывается достаточно дорого, если вообще возможно, с использованием реляционных СУБД. Основной причиной является то, что реляционная база данных предназначена для запуска на одном компьютере, но, как правило, более экономно сохранять большие объемы данных и запускать вычисления на кластерах, состоящих из гораздо меньших и более дешевых машин. Многие нереляционные СУБД специально спроектированы и разработаны для работы на кластерах, поэтому они лучше подходят для сохранения больших объемов данных.

В литературе чаще всего выделяют следующие виды не реляционных СУБД [6]:

– Столбико-ориентированные СУБД. Структура данных в них контрастирует с кортежным (строчным) форматом реляционных СУБД. Это позволяет избегать использования места при хранении null-данных, просто не сохраняя столбцы, если нет данных для этого столбца. Столбцы не требуют никакого априорного определения. Кроме того, они способны хранить любые типы данных, поскольку данные могут быть сохранены в виде массива байтов. Такая структура данных, называемая Bigtable, используется, в частности, компанией Google.

– СУБД типа «ключ-значение». Они являются простой хэш-таблицей, поэтому весь доступ к данным осуществляется через первичный ключ. В качестве значений данных используются blob-объекты (англ. Binary Large Objects – большие объекты в бинарном формате). Поскольку СУБД типа «ключ-значение» всегда используют первичный ключ для доступа к данным, они, как правило, имеют высокую производительность и легко масштабируются.

– Базы данных на основе графов. Они позволяют сохранять объекты предметной области и связи между этими объектами. Объекты также называют узлами, обладающими свойствами. Узел представляет собой представление экземпляра объекта в применении. Связи между объектами известны как ребра, которые могут обладать свойствами. Ребра имеют направление; узлы организованы связями, позволяющими найти определенные закономерности между узлами. Организация графа позволяет сохранить данные один раз и затем интерпретировать по-разному на основе связей между объектами. Такая структура данных имеет много общего с представлением знаний на основе семантических сетей.

– Многомерные базы данных. Используются для интерактивной аналитической обработки агрегированных данных, при этом агрегация может быть по нескольким уровням иерархии, например год, квартал, месяц. Гиперкуб данных можно рассматривать как множество отношений реляционной базы данных по значениям каждого из измерений. Следовательно, носителем многомерной модели данных является отношение реляционной базы данных, изображенное как зафиксированное измерение. Следовательно, между реляционной и многомерной моделью существует взаимное отображение, а значит, можно данные, представленные в одной модели, превратить в другую модель. Поэтому построенные на многомерной модели технологии OLAP (On Line Analysis processing) [4] реализованы в распространенных СУБД реляционного типа MS SQL Server, Oracle и других. Многомерная модель расширяет реляционную: множество операций, кроме традиционных реляционных, содержит операции среза, свертки, детализации, вращения.

– Объектные СУБД. В них данные моделируются в виде объектов, их атрибутов, методов и классов. Одним из известных представителей этого типа систем хранения и обработки данных является СУБД CACHÉ [5]. Помимо объектной модели она поддерживает также реляционную и многомерную, при этом представление данных в одной из моделей автоматически преобразуется в другие модели.

– документо-ориентированные СУБД. Документы являются основной концепцией в документо-ориентированных СУБД.

Была опубликована и представлена серия работ, объясняющая некоторые из ключевых элементов инфраструктуры созданных систем. Наиболее важные из этих публикаций представлены в [6–9].

После одобрения NoSQL двумя ведущими веб-гигантами – Google и Amazon – появилось несколько новых продуктов. Многие разработчики стали использовать методы и идеи NoSQL в своих приложениях. Меньше чем за 5 лет NoSQL и связанные с ним понятия для управления большими объемами данных получили широкое распространение и использование во многих известных компаниях, включая Facebook, Netflix, Yahoo, eBay, Hulu, IBM и т.д.

Проблемы параллелизма существуют и в реляционных базах данных (потеря результатов обновления, зависимость от незафиксированных результатов, несогласованная обработка результатов) [10], но у неструктурированных БД они более сложны.

2. Представление данных в виде агрегатов. С этой точки зрения они ближе к объектным базам данных. Например, в объектной СУБД CACHÉ [5] такие агрегаты называются глобалами. Поскольку в агрегате хранится вся информация для определенной задачи, чтобы свести к минимуму операции join между объектами, то следствием этого может быть дублирование информации в других агрегатах.

Исходя из вышеприведенного, выполнен анализ существующих документо-ориентированных (К) СУБД и выбор базовой для системы визуализации данных. На сегодняшний день существует достаточно много ДОСУБД. Они выпускаются под разными лицензиями и реализованы для использования на разных платформах, хотя некоторые могут использоваться на нескольких платформах одновременно. Несмотря на некоторые отличия, следует отметить, что каждая реализация документной базы данных базируется на основном понятии – «документе», описывающем одну запись в базе данных. При этом допускаются разные способы организации хранения списков документов.

Была реализована визуализация данных реального времени, предназначенная для консолидации и обработки информации о событиях в высоконагруженных веб-системах – рис. 1.

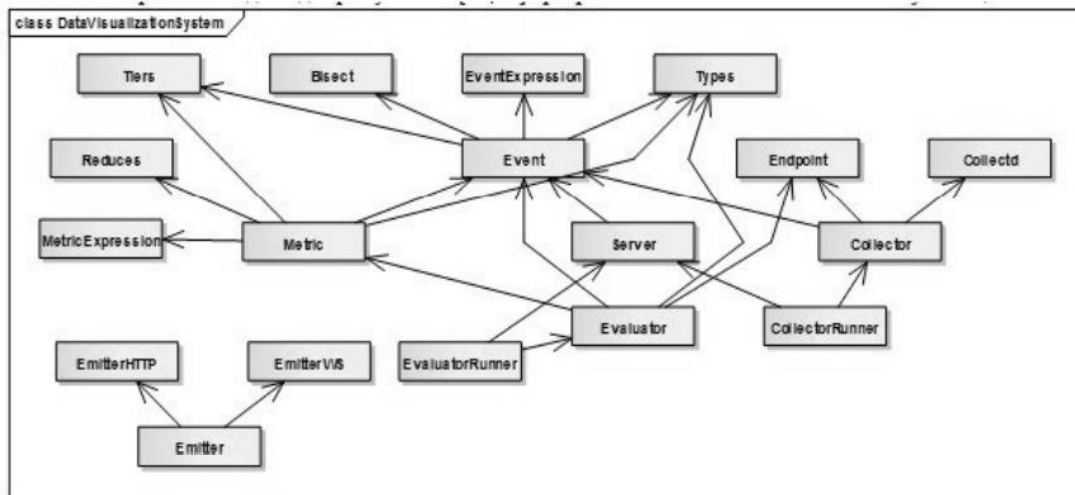


Рис. 1 – Диаграмма классов разработанной системы

Система использует документо-ориентированную СУБД MongoDB для хранения полученных данных и JavaScript для обработки событий на стороне сервера, запускаемого с применением платформы Node.js.

Выводы

Практическим результатом проделанной работы является созданная система визуализации данных реального времени, предназначенная для консолидации и обработки информации о событиях в высоконагруженных веб-системах. Система использует документо-ориентированную СУБД MongoDB для хранения полученных данных и JavaScript для обработки событий на стороне сервера, запускаемого с применением платформы Node.js.

Преимущество системы состоит в формировании визуальной картины работы системы в режиме реального времени. К сожалению, черно-белый формат представления графического материала не позволяет наглядно представить результаты примера работы системы визуализации данных.

Список литературы

1. Pedro Teixeira. Professional Node.js: Building Javascript Based Scalable Software. Indianapolis, IN. John Wiley & Sons, Inc. – 2012. – 412 p.
2. Абрамский М.М., Тимерханов Т.И. Сравнительный анализ использования реляционных и графовых баз данных в разработке цифровых образовательных систем // Вестник НГУ. Серия: Информационные технологии. 2018. №4. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-ispolzovaniya-relyatsionnyh-i-grafovyh-baz-dannyh-v-razrabotke-tsifrovyyh-obrazovatelnyh-sistem> (дата обращения: 03.02.2023).
3. Васильева К. Н., Хусаинова Г. Я. Реляционные базы данных // Colloquium-journal. 2020. №2 (54). URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh> (дата обращения: 03.02.2023).
4. Иванов В.И., Новиков С.П. Программное средство исследования различных алгоритмов кэширования в реляционных СУБД // Молодой исследователь Дона. 2018. №3 (12). URL: <https://cyberleninka.ru/article/n/programmnoe-sredstvo-issledovaniya-razlichnyh-algoritmov-keshirovaniya-v-relyatsionnyh-subd> (дата обращения: 03.02.2023).
5. Крэг Ларман. Применение UML 2.0 и шаблонов проектирования. – 3-е изд. – М.: Вильямс, 2006. – 736 с.
6. Применение CASE-средства VPwin 4.0 для информационного моделирования в системах обработки данных / С.В. Горин; А.Ю. Тандоев // СУБД. – М., 1995. – №3. – С. 125–127.
7. Реляционная база данных и её структура / Научная библиотека избранных естественно-научных изданий. — URL: http://sernam.ru/book_cbd.php?id=2 (дата обращения: 03.02.2023).
8. Савоськин И.В., Фирсов А.О. Исследование способов применения NoSQL и реляционных баз данных // E-Scio. 2019. №6 (33). URL: <https://cyberleninka.ru/article/n/issledovanie-sposobov-primeneniya-nosql-i-relyatsionnyh-baz-dannyh> (дата обращения: 03.02.2023).
9. Свиридов А.А. Хранение топологии сети в реляционных базах данных // JSRP. 2013. №1. URL: <https://cyberleninka.ru/article/n/hranenie-topologii-seti-v-relyatsionnyh-bazah-dannyh> (дата обращения: 03.02.2023).
10. Федорова Г. Н. Разработка и администрирование баз данных / Г. Н. Федорова. – М.: Академия, 2015. – 58 с.

References in Cyrillics

1. Pedro Teixeira. Professional Node.js: Building Javascript Based Scalable Software. Indianapolis, IN. John Wiley & Sons, Inc. – 2012. – 412 p.
2. Abramskij M.M., Timerhanov T.I. Sravnitel'nyj analiz ispol'zovaniya relyacionnyh i grafovyyh baz dannyh v razrabotke cifrovyyh obrazovatel'nyh sistem // Vestnik NGU. Seriya: Informacionnye tekhnologii. 2018. №4. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-ispolzovaniya-relyatsionnyh-i-grafovyh-baz-dannyh-v-razrabotke-tsifrovyyh-obrazovatelnyh-sistem> (data obrashcheniya: 03.02.2023).
3. Vasil'eva K. N., Husainova G. YA. Relyacionnye bazy dannyh // Colloquium-journal. 2020. №2 (54). URL: <https://cyberleninka.ru/article/n/relyatsionnye-bazy-dannyh> (data obrashcheniya: 03.02.2023).
4. Ivanov V.I., Novikov S.P. Programmnoe sredstvo issledovaniya razlichnyh algoritmov keshirovaniya v relyacionnyh SUBD // Molodoj issledovatel' Dona. 2018. №3 (12). URL: <https://cyberleninka.ru/article/n/programmnoe-sredstvo-issledovaniya-razlichnyh-algoritmov-keshirovaniya-v-relyatsionnyh-subd> (data obrashcheniya: 03.02.2023).
5. Kreg Larman. Primenenie UML 2.0 i shablonov proektirovaniya. – 3–e izd. – M.: Vil'yams, 2006. – 736 s.
6. Primenenie CASE–sredstva BPwin 4.0 dlya informacionnogo modelirovaniya v sistemah obrabotki dannyh / S.V. Gorin; A.YU. Tandoev // SUBD. – M., 1995. – №3. – S. 125–127.
7. Relyacionnaya baza dannyh i eyo struktura / Nauchnaya biblioteka izbrannyh estestvenno-nauchnyh izdanij. — URL.: http://sernam.ru/book_cbd.php?id=2 (data obrashcheniya: 03.02.2023).
8. Savos'kin I.V., Firsov A.O. Issledovanie sposobov primeneniya NoSQL i relyacionnyh baz dannyh // E-Scio. 2019. №6 (33). URL: <https://cyberleninka.ru/article/n/issledovanie-sposobov-primeneniya-nosql-i-relyatsionnyh-baz-dannyh> (data obrashcheniya: 03.02.2023).
9. Sviridov A.A. Hranenie topologii seti v relyacionnyh bazah dannyh // JSRP. 2013. №1. URL: <https://cyberleninka.ru/article/n/hranenie-topologii-seti-v-relyatsionnyh-bazah-dannyh> (data obrashcheniya: 03.02.2023).
10. Fedorova G. N. Razrabotka i administrirovanie baz dannyh / G. N. Fedorova. – M.: Akademiya, 2015. – 58 s.

Ключевые слова

Визуализация, базы данных, реляционные, особенности, изучение, применение

Калинина Екатерина Сергеевна, к. т. н, доцент кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения», Омск, Россия, ekkalinina@mail.ru

Манохина Татьяна Витальевна, старший преподаватель кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения», mtv-gups@mail.ru

Ступаков Сергей Анатольевич, к. т. н, доцент кафедры «Информатика и компьютерная графика» ФГБОУ ВО «Омский государственный университет путей сообщения», Омск, Россия, stupakov1@yandex.ru

Keywords

Visualization, databases, relational, features, study, application

Kalinina Ekaterina Sergeevna, Candidate of Technical Sciences, Associate Professor of the Department of Informatics and Computer Graphics, Omsk State Transport University, Omsk, Russia, ekkalinina@mail.ru

Manokhina Tatyana Vitalievna, Senior Lecturer of the Department of Informatics and Computer Graphics, Omsk State Transport University, mtv-gups@mail.ru

Stupakov Sergey Anatolyevich, Candidate of Technical Sciences, Associate Professor of the Department of Informatics and Computer Graphics, Omsk State Transport University, Omsk, Russia, stupakov1@yandex.ru

DOI: 10.34706/DE-2023-01-12

JEL classification C02 – Математические методы; M15 Управление информационными

The article deals with the problems of studying the visualization of databases of relational type. The article highlights the advantages and disadvantages of the MongoDB document-oriented DBMS using the example of creating a real-time data visualization system. A real-time data visualization system has been created, designed to consolidate and process information about events in highly loaded web systems. The system uses a document-oriented MongoDB DBMS to store the received data and JavaScript to process events on the server side, launched using the Node.js platform.

Памятка для авторов публикаций в журнале «Цифровая экономика»

В нашем журнале выполняются все требования Diamond-OA, включая отсутствие платы как со стороны авторов, так и со стороны читателей, рецензирование, а также проверка на плагиат и избыточное самоцитирование. Авторские права на опубликованные статьи остаются за авторами.

В журнале нет штатных сотрудников, все работы, включая проверку на плагиат, рецензирование, работу корректора и форматирование, выполняются группой единомышленников на общественных началах, а потому мы рассчитываем на такое же отношение к своим правам и обязанностям со стороны авторов. Материалы, опубликованные ранее (полностью или в значительной своей части) в других изданиях, не принимаются. Мы очень надеемся, что предполагаемые авторы избавят нас от работы с такими текстами.

Первое, что предлагается автору, желающему опубликовать статью в нашем журнале, – это зарегистрироваться в качестве потенциального автора и самому разместить предлагаемый к публикации текст на сайте журнала в отведенном для этого разделе (научные статьи, мнения, обзоры, рецензии, переводы). Тем самым автор принимает условия журнала и дает добро на публикацию своей статьи в журнале после прохождения всех предусмотренных процедур. Статья, прошедшая проверку и рецензирование, получает отметку о том, что она будет опубликована в журнале.

При отборе статей для публикации в очередном выпуске включение статьи в этот выпуск определяется, прежде всего, соотношением объемом материалов, в принципе годных для публикации, и фиксированным (96 страниц 9-м кеглем) объемом выпуска. Во внимание принимается соответствие тематики, время подачи материала и его готовность к публикации.

Полная готовность научной статьи к публикации означает ее соответствие принятому в журнале стандарту, включая правильное оформление списка литературы и ссылок, полные сведения об авторах, индексы JEL, аннотацию и ключевые слова на русском и английском, редактируемые формулы (набранные Word и в нем же редактируемые), ручную нумерацию разделов, рисунков и таблиц. Если нумерация автоматическая, она может сбиться при вставке статьи в общий блок.

Заголовок не должен быть длинным. Иначе в колонтитуле будет бессмыслица. Не надо набирать заголовки большими буквами. Надо использовать опцию «все прописные». Это важно!

В списке литературы научные статьи упорядочиваются по алфавиту, причем сначала идут русскоязычные публикации, потом англоязычные и пр. Это нужно, чтобы не возникло путаницы при формировании транслитерации кириллических статей. Источники данных, нормативные и методические материалы идут отдельным списком. Ссылки на интернет-ресурсы, газетные публикации и т.д. желательно давать в сносках. Ссылки на научные публикации должны быть даны в формате [Автор, 2023]. При необходимости к году может быть добавлена латинская буква 2023a, 2023b.

Публикация статьи означает получение ей метаданных, включая DOI, номер выпуска, страницы. Выпуск журнала делается в формате pdf, причем в таком виде, что его сразу можно отдать в типографию и сделать твердую (бумажную) копию, если кто-то из авторов хочет ее иметь для себя. Бумажная версия выпуска имеет статус буклета, печатается за счет автора (заказчика) и в количестве, определенном заказчиком.

Статьи, размещенные авторами на сайте журнала, доступны читателям немедленно, еще до того, как прошли рецензирование. Они не считаются опубликованными до прохождения рецензирования и технических процедур. Но самим фактом размещения и предварительной регистрации человек решает это опубликовать, отпадает необходимость в письменном договоре. Если автор присылает статью в журнал и просит ее разместить, он нарушает стандартную процедуру и может создать нам сложности в будущем.

Старайтесь следовать правилам и не создавать нам проблем!