

Самообучающаяся ЭС.

Часть 1. Постановка задачи.

Варлам Кешелава, Роман Душкин, Т

имур Кешелава

Оглавление

1	Стартовая позиция.....	1
2	Исходные позиции: нейронные сети и семантические сети.....	2
2.1	Искусственные нейронные сети.....	2
2.2	Семантические сети	3
2.3	Одно отличие, ставшее «судьбоносным».....	4
3	Наш ответ Г. Маркусу: гибридные ИИ-системы.....	5
3.1	Гибридные системы ИИ	5
3.2	Дискутируем с Г. Маркусом.....	7
3.2.1	ГО нуждается в данных	7
3.2.2	ГО пока что имеет небольшую глубину и плохо переносит полученные знания на другие данные	8
3.2.3	У ГО пока нет естественного способа работы с иерархической структурой	8
3.2.4	ГО до сих пор с трудом работает с неточно определяемыми понятиями	9
3.2.5	ГО до сих пор недостаточно прозрачно	10
3.2.6	ГО пока плохо интегрируется с уже существующими знаниями	10
3.2.7	ГО пока что неспособно автоматически отличать причинно-следственную связь от корреляции	11
3.2.8	ГО пока что хорошо работает в роли аппроксимации, но его ответам нельзя полностью доверять.....	11
3.2.9	ГО тяжело использовать в прикладных целях	12
4	Ключевая задача.....	13
4.1	При всем богатстве выбора	13
4.2	... выбор очевиден	13
5	Заключение	14
	Список литературы.....	15

Авторы не понаслышке знают о неудаче предыдущих попыток решить подобную задачу, поэтому предлагают революционно новый подход: «Давайте сперва подумаем!»

1 Стартовая позиция

Компьютерная техника с самого момента своего появления неизменно расширяла своё присутствие как в производстве, так и в быту. В настоящее время невозможно представить себе офис без компьютеров, которыми оборудовано практически каждое рабочее место, магазин без кассовых аппаратов, объединённых в сеть, банк без банкоматов, знающих состояние счёта клиента, учебный процесс в школе, равно как в высшем учебном заведении, без разнообразных вычислительных устройств и т. д. Более того, современные вычислительные устройства, уменьшившись в размерах, превратились в «гаджеты»: планшеты, смартфоны, плееры и целый ряд других устройств, постоянно носимых с собой большинством людей. Дальше — больше. Уже разработаны и успешно внедряются на практике, в том числе и среди конечных пользователей, разнообразные носимые устройства и импланты — пока это только «умные» часы и разнообразные протезы, которые год от года становятся всё лучше и интеллектуальнее. В ближайшем будущем интерфейс компьютерных программ, улучшившись, перестанет разделять человека и виртуальный мир, а далее возможны нейроинтерфейсы и нейроимплантаты, которые обеспечат непосредственное взаимодействие между гаджетами и нервной системой человека. Все это ведет к сращиванию реального и виртуального миров. Со стороны традиционного реального мира этот процесс представляется его преобразованием под влиянием цифрового мира. Этот процесс «цифровизации» не остановить.

Среди наиболее значительных прорывов, которые можно с уверенностью предсказать, ближайшим видится внедрение технологий обработки компьютером текстовой информации. Избегая дискуссий по поводу искусственного интеллекта и квантовых компьютеров, воспользуемся терминологией компании IBM.

«Когнитивные вычисления» — общее название группы технологий, способных обрабатывать информацию, находящуюся в неструктурированном, чаще всего текстовом, виде (неструктурированные данные). Они не следуют заданному алгоритму, а способны учитывать множество сторонних факторов и самообучаться, используя результаты прошлых вычислений и внешние источники информации (например, Интернет). В настоящее время наиболее продвинутой когнитивной системой является IBM Watson. Близки к ним и наиболее развиты экспертные системы, такие как СYC (www.cyc.com) и Wolfram Alpha (wolframalpha.com).

Возможности современных методов когнитивных вычислений уже достигли вполне дееспособного уровня. Используя этот уже существующий уровень, можно резко снизить трудозатраты на рутинную офисную работу, то есть обработку стандартных документов, включая справки, заявки, заявления, отчёты, платёжные документы, декларации, договоры и т. д. Таким образом, можно предельно автоматизировать основную часть документооборота. Также можно разработать интерактивный помощник-проводник по корпоративной информационной системе, автоматизировать большинство рутинных операций, а также интеллектуализировать поиск и агрегацию информации, максимально увеличив эффективность аналитической деятельности.

Распространение этой технологии будет настолько всепроникающим, а её влияние на социально-экономические процессы будет настолько огромным, что их трудно переоценить.

2 Выбираем: нейронные сети и семантические сети

В качестве технологической основы построения экспертных систем вообще и когнитивных вычислительных систем, в частности, весьма показательную борьбу ведут нейросетевой и семантический подходы. Оба подхода претендуют на выполнение основных функций: обработки, накопления и использования знаний.

Кроме сравниваемых далее нейронных и семантических сетей существует еще несколько потенциально конкурирующих решений. Все они родственны семантическому подходу т.к. предполагают явное представление знаний. Кроме того, в них можно (иногда весьма условно) выделить хранилище информации и (или) знаний и машину вывода, способную «оперировать» этими знаниями. Существует несколько альтернатив: уже упомянутые семантические сети, фреймы, специальные языки типа LISP, Prolog или KL-ONE и др. Однако в последние 30 лет семантические и нейронные сети заняли доминирующее положение.

В свою очередь из этих двух технологий гораздо активнее развивались нейронные сети. Они даже «захватили» ряд практических приложений, которые казалось более подходящими для семантических сетей и, более того, были первоначально реализованы на семантических сетях и других технологиях нисходящего (символьного) подхода. Такими примерами являются машинные переводчики, системы обработки естественного языка, интеллектуальный анализ данных.

2.1 Искусственные нейронные сети

Основным путём усовершенствования искусственных нейронных сетей в настоящее время стала технология глубокого (или глубинного) обучения (ГО). Анализируя актуальные проблемы развития технологии ГО и искусственного интеллекта (ИИ) в целом, Г. Маркус [Marcus, 2017; есть перевод на русский язык¹] приходит к следующим утверждениям:

- Для работы ГО-систем необходимо много качественных данных, которые должны быть предварительно очищены и размечены специалистом. Чем больше слоёв нейронной сети в ГО-системе, тем больше данных требуется;
- ГО-системы работают только с теми типами данных, на которых происходило обучение, и они всё ещё не могут обобщать и переносить найденные закономерности на данные других типов, даже очень близких;
- ГО-системам очень сложно работать с иерархическими структурами, поэтому, к примеру, обработка языка для них — это очень непростая задача, поскольку естественный язык — это многоуровневая иерархическая структура;
- Как следствие предыдущего ГО-системы с большим трудом воспринимают неточные и нечёткие данные, они часто не различают разницы там, где для человеческого интеллекта разница очевидна и значима;
- ГО-системы унаследовали от искусственных нейронных сетей и усугубили проблему высокой сложности (до практической невозможности) объяснения полученных результатов и произведённого вывода;

¹ С переводом можно ознакомиться по адресу <https://geektimes.ru/post/297309/>



- ГО-системы не принимают во внимание уже имеющийся корпус знаний, а заново обучаются на входных данных, иначе их интерпретируя;
- Важная задача выявления причинно-следственных связей и отделения их от простых корреляций ГО-системам пока даётся с трудом;
- ГО-систему легко обмануть, особенно если она находится на «границе переобученности». Эта уязвимость открывает широкий простор для разнообразных атак, последствия которых ещё даже не до конца осознаны. К решению этой проблемы по большому счету не приступали;
- Крупномасштабных прикладных применений ГО-системам так до сих пор и не найдено, и этот пункт является, фактически, следствием всех предыдущих.

В целом, автор без оптимизма описывает нынешнее состояние и настраивает на переосмысление полученных результатов, даже если они и промежуточные.

Далее в оригинальной статье автор делает неутешительные прогнозы о том, что поднятая шумиха по поводу ГО может привести к новой, уже третьей по счёту «зиме» искусственного интеллекта. Однако он также даёт и своё видение того, как всё это можно было бы преодолеть. Среди его советов такие: применение методики спонтанного обучения, попытки решения более сложных задач, использование современных знаний в области психологии и, наконец, применение гибридных моделей с технологиями синтаксической манипуляции символами.

В качестве промежуточного вывода, учитывая долгосрочные последствия, признаем, что чисто нейросетевой подход в широком смысле является, по всем видимости, тупиковым. Во-первых, нейронная сеть, как правило, не позволяет понять (и проконтролировать) как она функционирует, то есть ничего не добавляет в общую «копилку знаний». Во-вторых, как бы хорошо не справлялась нейронная сеть с поставленной задачей, всегда есть опасения, что с такой же уверенностью она выдаст ошибочный ответ в результате маленькой, незамеченной ранее особенности². При этом вопрос снижения частоты ошибок второго рода при использовании нейросетевого подхода в общем случае решить практически невозможно из-за отсутствия формализма объяснения результатов работы нейросетевых моделей.

2.2 Семантические сети

Семантические сети отличаются возможностью полностью представить всю цепочку рассуждений, приводящих её к тому или иному выводу (развитая функция объяснения). Это не только позволяет контролировать её, но также обеспечивает возможность настраивать и управлять её свойствами. В принципе (если пренебречь иногда непреодолимыми техническими сложностями) вся информация (знания) всех семантических сетей могут быть объединены. При создании соответствующего стандарта появляется возможность суммировать результаты, получаемые независимыми группами разработчиков.

Большинство [Jackson, 1999] ранних удачных экспертных систем (ЭС), так же как в наше время большинство интеллектуализированных систем, основанных на знаниях, состоят из двух существенно различающихся частей:

- Тезауруса, состоящего из фактографической базы данных, и базы знаний, содержащей продукционные правила;

² Один пиксель превратит лошадь в корабль для нейронной сети: <https://indicator.ru/news/2017/10/31/odin-piksel-nejroset/>



- «Универсальной» машины логического вывода (УМВ), обеспечивавшей процедурную функцию обработки баз данных и баз знаний.

Обе составляющие являются обязательными для любой интеллектуализированной системы, основанной на явно сформулированных знаниях. Используя алгоритмы, реализующие законы формальной логики, УМВ строит возможные цепочки взаимосвязанных утверждений, манипулируя информацией, содержащейся в базах фактов и знаний. Для предотвращения комбинаторного взрыва и заикливания в ловушке логических петель вводится специальный механизм отбора «наиболее перспективных» цепочек. При росте величины баз последняя задача становится крайне нетривиальной, так как критерии и механизмы реализации такого отбора могут быть довольно разнообразными. Финальной функцией является выбор той цепи «рассуждений», которая кратчайшим путём ведёт к искомому ответу.

В 70 — 80-х годах XX века казалось, что структура представления знаний в разных проблемных областях не должна сильно отличаться, а УМВ, опирающаяся на законы формальной логики, вообще говоря, универсальна. Получалось, что можно взять за основу удачную ЭС, и, заменив фактические данные проблемной области, создавать новые ЭС. Самая известная попытка — это создание EUMYCIN на основе ЭС MYCIN.

ЭС MYCIN предназначена для первичной диагностики бактериальной инфекции у пациента, приходящего к врачу. Те немногие врачи, которые в те времена имели доступ к компьютерной технике и владели ею в достаточной степени, положительно отзывались об этой ЭС.

Система EUMYCIN представляла собой оболочку ЭС: УМВ и размеченная структура баз данных и правил с модулем, предоставляющим возможность эксперту проблемной области заполнять базу знаний.

Однако выяснилось, что:

- создание базы продукционных правил для новой проблемной области оказалось весьма нетривиальным и дорогостоящим делом (здесь родилась «инженерия знаний»);
- трудно описывать проблемные области в существующих понятиях низкого уровня абстракции.

То есть те конкретные решения, которые были положены в основу реализации ЭС, основанных на знаниях, предопределили формирование двух больших проблем. Первая: базы оказались не универсальными, что приводит к необходимости создавать и предметные БД, и предметные машины вывода (МВ) каждый раз практически с нуля. Вторая: базы, как правило, взаимно несовместимы, что в современной литературе описывается как невозможность «повторного использования компонентов» [Болотова, 2012].

В это связи может встать вопрос о возможности представления плохо формализуемых правил из областей знания, не оперирующих строгими зависимостями и понятиями. Стоит отметить, что некоторые исследователи подчеркивают потенциальную возможность формализации даже описательных наук, опираясь на достижения в области математических подходов [Шевченко, 2019].

2.3 Одно отличие, ставшее «судьбоносным»

Обратим внимание на явление, которое можно назвать «барьером вхождения», отличающим семантический подход от нейросетевого.

- Для работы с искусственными нейронными сетями необходим небольшой опыт программирования, наличие размеченной обучающей выборки (для некоторых задач такая выборка даже не требуется) и терпение. Используя свободно распространяемое программное обеспечение для реализации самой нейронной сети, сделать и продемонстрировать руководителю или заказчику работающую систему (или «минимально жизнеспособный продукт») оказывается делом несложным. После этого можно долго её совершенствовать.
- Семантический подход почти во всём противоположен: для его реализации необходимы эксперты проблемной области, и с ними придётся налаживать отношения. После этого вместе с ними придётся проделать значительную предварительную работу по составлению тезауруса. Это включает создание терминологического словаря, словаря синонимов, антонимов и др., создание онтологических шкал атрибутов, формализацию базовых принципов и законов проблемной области, наполнению (вручную, за счёт труда привлечённых экспертов) семантической сети конкретной информацией, согласование полученной базы знаний и т. д. Получить в этом случае прототип, который можно продемонстрировать руководителю или заказчику, стоит большого труда, занимает много времени, требует привлечения целого коллектива разработчиков, исходно не обладающих общей коммуникативной базой и, как следствие, весьма затратно.

Именно этим сиюминутным прагматическим преимуществом (по мнению авторов) во многом объясняется широкое распространение нейронных сетей.

3 Наш ответ Г. Маркусу: гибридные ИИ-системы

Последняя возможность, предложенная Г. Маркусом, является одной из наиболее предпочтительных, так как методы нисходящей парадигмы в области ИИ были разработаны много десятилетий назад [Minsky, 1967] и незаслуженно подзабыты.

3.1 Гибридные системы ИИ

Гибридная ИИ-система включает в себя лучшие стороны как нисходящего, так и восходящего подходов, но при этом старается отвергнуть стороны слабые [Колесников, 2001; Medsker, 1995]. В итоге получается продвинутое интеллектуальное устройство, которое в какой-то мере отображается на устройство интеллекта естественного. Впрочем, если рассмотреть верхнеуровневую архитектуру такой системы, то будет ясно, что это довольно обычная кибернетическая система [Эшби, 1959], взаимодействующая со средой через фильтры, поэтому неудивительно, что она напоминает биологические системы.

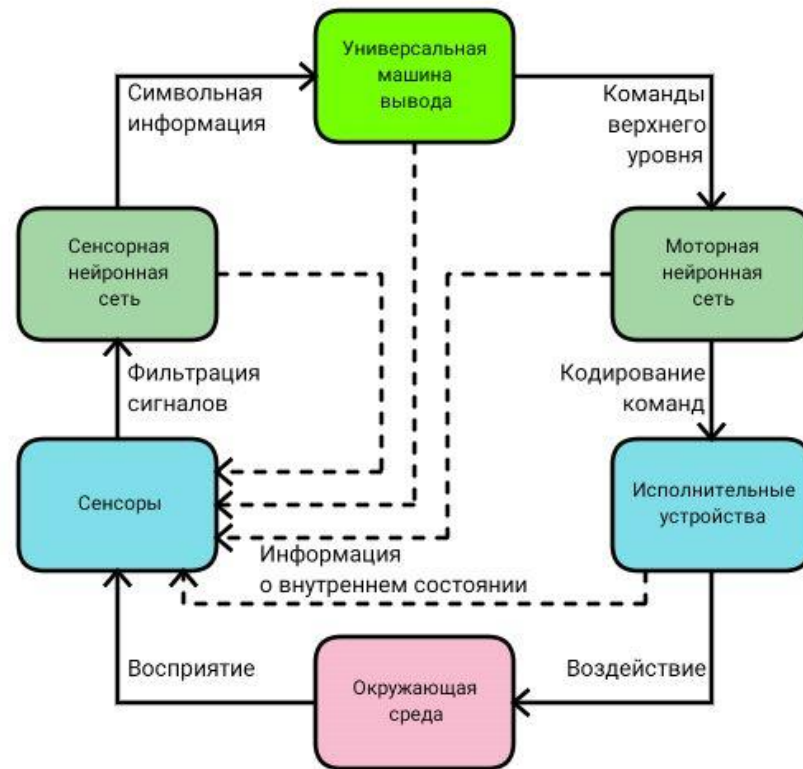


Рис. 1. Верхнеуровневая архитектура гибридной ИИ-системы

Итак, как видно из представленной диаграммы, гибридная ИИ-система представляет собой ни что иное, как универсальную кибернетическую машину, которая имеет три основных элемента: аффлекторы, подсистему управления и эффекторы. При помощи аффлекторов кибернетическая машина воспринимает сигналы окружающей среды, которые обрабатываются в подсистеме управления, сигналы из которой далее поступают в эффекторы, которые на окружающую среду воздействуют. Это общая схема любого автономного агента, поэтому гибридная ИИ-система одновременно является и интеллектуальным агентом, реализуя агентный подход [Weiss, 1999; Wooldridge, 2002; Зайцев, 2008].

Гибридная ИИ-система отличается тем, что её аффлекторы (сенсоры, датчики) и эффекторы (исполнительные устройства) связаны с подсистемой управления и принятия решений через нейронные сети. Тем самым используются сильные стороны восходящего, или «грязного» подхода. Аффлекторная нейронная сеть принимает очищенные сенсорами сигналы внешней среды и преобразует их в символы, которые подаются на вход универсальной машине вывода. Последняя осуществляет вывод на основе символьных знаний из своей базы знаний и выводит результат, который тоже представляется в виде символов. Тем самым реализуются сильные стороны нисходящего, или «чистого» подход. Символьный результат подаётся на вход моторной нейронной сети, которая преобразует высокоуровневые символы в конкретные сигналы управления исполнительными устройствами.

Кроме всего прочего внутри гибридной ИИ-системы должны быть реализованы контрольные связи от всех её элементов к её сенсорам. Тем самым реализуются адаптационные механизмы, основанные на гомеостазе (саморегулирующимся постоянстве) внутреннего состояния системы [Martin, 2009]. Сенсоры фиксируют изменение внутреннего состояния каждой подсистемы, их

элементов и комплексов, и в случае выхода контролируемых значений за пределы установленных гомеостатических интервалов подсистемой управления принимается решение, целью которого будет возврат изменённых показателей в установочный интервал [Бир, 1993].

Именно система с такой архитектурой при переходе через определённый порог сложности может считаться «разумной». При этом «разумность» в этом смысле определяется как адекватное реагирование не только на стимулы внешней среды, но и на внутренние состояния, что также включает в себя постоянный мониторинг состояния собственной подсистемы управления, что называется «саморефлексией», что можно интерпретировать как приход к «осознанию».

Именно на слиянии двух подходов — нисходящего и восходящего — с построением ИИ-систем гибридного типа имеет смысл развивать технологии и преодолевать возникшие сложности в области ГО.

3.2 Дискутируем с Г. Маркусом

Разберемся подробнее в том, как обозначенные проблемы могли бы быть преодолены при помощи гибридных ИИ-систем и методов нисходящей парадигмы ИИ.

3.2.1 ГО нуждается в данных

Чем больше данных, особенно качественно размеченных, — тем лучше результат работы глубинных нейронных сетей [Гудфеллоу, 2017]. Более того, они могут обобщать информацию и находить скрытые зависимости, строя что-то типа продукционных правил «если ..., то ...» [Miller, 1957]. Фактически, любая нейросеть при обучении делает это, так как в действительности строит правила преобразования входных данных в выходные, что на более высоком уровне абстракции можно представить в виде продукций «Если вход принимает такие-то значения, то выход равен тому-то» [Каллан, 2001].

Однако проблема в том, что даже если нейронная сеть делает подобные обобщения, то их представление скрыто в её глубинах в виде неявной информации, отображаемой только в весовые коэффициенты на связях между нейронами [Круглов, 2001]. Интерпретация этих коэффициентов очень затруднительна, если вообще возможна, особенно с учётом того, что одна и та же межнейронная связь может делать вклад в несколько различных правил [Николенко, 2018].

Если рассмотреть пример, приведённый в оригинальной статье, где введено слово «шместра», которое определяется как «сестра, возраст которой от 10 до 21 года», то нейронная сеть вполне может научиться определять, кто шместра, кто просто сестра, а кто вообще ни та, ни другая. Вопрос лишь в обучающей выборке. И после того, как нейросеть обучилась, она отлично разделит шместёр от сестёр, но никогда не сможет объяснить, кто такая шместра и почему она распознаёт шместру в том или ином случае.

Однако если построить гибридную ИИ-систему, то её архитектура может быть устроена следующим образом. Сенсорная нейронная сеть определяет базовые параметры входа. Например, то, что на изображении женщина и ей по оценкам 18 лет. Это то, что нейронные сети с ГО сегодня умеют делать очень хорошо. А вот дальше эти распознанные базовые характеристики подаются на вход универсальной машине вывода, в базе знаний которой есть правило «Если женщина является сестрой и ей от 10 до 21 год, то это шместра», после чего запускается процесс вывода — система проверит, является ли распознанная женщина сестрой, и если да, то будет

абсолютно ясно, что это шместра. Результат такого вывода может быть объяснён очень просто, спускаясь до базовых характеристик, распознанных нейронной сетью.

И обучение такой гибридной ИИ-системы может осуществляться как в процессе разработки, так и в процессе взаимодействия с ней [Флах, 2015]. Например, если мы общаемся с ней при помощи диалогового интерфейса (чат-бота), то при использовании в тексте слова «шместра» ИИ-система может поинтересоваться, что это такое. И тогда можно дать ей определение.

Так что первая проблема систем ГО понятным образом решается при помощи гибридного подхода.

3.2.2 ГО пока что имеет небольшую глубину и плохо переносит полученные знания на другие данные

Системы ГО находят интересные закономерности, но не могут их обобщить, и отсюда происходят различные смешные казусы, описанные в оригинальной статье [Huang, 2017; Kansky 2017]. Обобщение — это операция более высокого уровня абстракции, чем простое нахождение закономерностей, даже скрытых. И если с последним искусственные нейронные сети легко справляются и уже в некоторых областях превосходят в этом людей, то с обобщением у них серьёзные проблемы.

Фактически, обобщение — это символьная операция. Обобщая, человек переносит найденные закономерности с отдельных экземпляров на типы и классы объектов внешнего мира или абстрактных сущностей [Кондаков, 1975]. И делается это именно на символьном уровне, когда правила вывода модифицируются при помощи метаправил. Например, одним из таких метаправил может быть следующее: «Если несколько объектов одного класса имеют одно и то же свойство, то предположить, что все объекты этого класса имеют это свойство». Классы могут объединяться в классы более высокого ранга, и для них тоже можно делать обобщающие выводы, и т. д. Всё это можно реализовать при помощи самообучающейся универсальной машины вывода, в которой есть метаправила, в том числе и для изменения самих метаправил.

Но для этого необходимо иметь иерархию понятий. Но у систем ГО реализация иерархий пока очень слабая, поскольку у нейросетевых моделей нет естественного способа представления иерархических понятий [Душкин, 2011].

3.2.3 У ГО пока нет естественного способа работы с иерархической структурой

У систем ГО пока нет естественного способа представления и работы с иерархическими структурами, так как нейронные сети пока с трудом справляются с иерархиями понятий. Попытки представить иерархические структуры в нейросетевых моделях приводят либо к переобучению нейросетей, либо к тому, что они не могут полноценно дифференцировать объекты распознавания при спуске по иерархии вниз. Кое-каких успехов в этом вопросе достигли нейросети именно глубинного обучения, так как в них неявно и самостоятельно образуются иерархии образов, но ведь весь вопрос в том, чтобы вложить в нейросеть некоторую заданную иерархию в явном виде. Пока это вызывает определённые затруднения [Marcus, 2001].

Хотя, к примеру, существуют нейросетевые модели, которые преобразуют последовательности слов в некие «векторы в трёхсотмерном пространстве смыслов, суть которых человеческому сознанию не ясна», которые группируются в некие кластеры. Но это опять же не иерархии, а всё

такие же векторные преобразования, которыми, по сути, нейросетевые модели и являются [Mikolov, 2013; Socher, 2012].

С другой стороны, нисходящая парадигма имеет все средства для непосредственного представления иерархических понятий. Это можно сделать даже при помощи продукций, не говоря уже о семантических сетях и онтологиях, которые именно для этого и предназначены [Добров, 2009]. Но здесь речь идёт опять о явном представлении знаний и дедуктивном обучении с учителем, хотя вполне возможно построение искусственных интеллектуальных систем, которые обучаются не в специальном режиме обучения, а в процессе взаимодействия с окружающей средой. Главное, чтобы из среды приходило правильное подкрепление.

Так что при использовании гибридного подхода опять можно взять лучшее от обеих парадигм. На нижнем уровне распознающие нейросети будут определять конкретные объекты и понятия, с которыми приходится взаимодействовать системе, а на уровне выше эти распознанные понятия могут укладываться в семантические сети для определения иерархических взаимоотношений и получения обобщённых выводов на основе классов высших порядков, в которые входит распознанный объект [Sowa, 1991]. Например, визуальная нейросеть распознаёт лицо мужчины на фотографии, а на верхнем уровне к конкретным характеристикам распознанного лица добавляются не только персональные данные и физические характеристики, но и такие понятия, как «взрослый», «самец», «человек», «примат», «млекопитающее» и все остальные классы, в которые входят все нижележащие объекты иерархии. А далее к ним могут быть применены все правила и метаправила вывода для получения как заключений по конкретной ситуации, так и новых знаний.

3.2.4 ГО до сих пор с трудом работает с неточно определяемыми понятиями

В процессе своей жизнедеятельности люди постоянно сталкиваются с ситуациями, где данные неполны, информация неточна, знания противоречивы, нечётки и неопределённые. Эти так называемые НЕ-факторы знания пронизывают всю информационную ткань нашей реальности уже в силу самой её природы, поскольку точность данных невозможно повысить до бесконечности в силу фундаментальных ограничений [Душкин, 2011]. Человеческий разум научился обрабатывать такие ситуации, но системы глубинного обучения — нет, поскольку требуют чётко размеченных данных и максимально полный набор значений входных параметров для обучения. Только в последнее время начали появляться комбинированные механизмы для обработки при помощи нейронных сетей знаний с нечёткостью или неопределённостью [Bart, 1992; Abraham, 2005].

Вместе с тем символичный подход в искусственном интеллекте прекрасно справляется с НЕ-факторами. Для этого предназначены разнообразные формализмы от метода Демпстера — Шафера [Dempster, 1968; Shafer, 1976], нечёткой логики Л. Заде и заканчивая мягкими и лингвистическими вычислениями [Заде, 1976]. Эти формализмы позволяют осуществить вывод даже в случае неопределённости высокой степени, получив вполне приемлемые результаты.

Тем не менее, знания обладают ещё и НЕ-факторами другого рода, которые относятся не к конкретной информации, но к знаниям в целом. Это уже упомянутая неполнота, это противоречивость и некорректность, это неадекватность и некоторые другие свойства знания [Душкин, 1999], которые проявляются у людей в разной степени, но характеризуют каждую личность по степени её экспертизы в той или иной области (можно быть экспертом в одной области, обладая всей доступной полнотой знания в ней, и профаном в другой, не имея в ней

знаний вообще). Гибридные ИИ-системы могут объединить мощь глубинного обучения для выявления новых знаний и формальные методы обработки НЕ-факторов для работы в условиях неопределённости и неполноты информации.

3.2.5 ГО до сих пор недостаточно прозрачно

О «невозможности» интерпретировать результаты обучения нейронных сетей уже рассказано очень много. Тут вопрос не в непознаваемости, а в слишком высокой вычислительной требовательности задачи понимания результатов настройки весовых коэффициентов связей между нейронами, полученных в результате обучения [Хайкин, 2006]. Если сеть состоит из сотен тысяч нейронов и тысяч слоёв, то количество весовых коэффициентов связей колоссально, и «дизассемблирование» результатов обучения просто невыполнимо с вычислительной точки зрения.

Эта проблема уже привела к появлению новой парадигмы — ХАИ, eXplainable Artificial Intelligence (объяснимый искусственный интеллект) [Knight, 2017]. ХАИ должен объяснять свои решения и результаты, а это вполне возможно для гибридных ИИ-систем, поскольку в рамках нисходящего подхода объяснение — это одна из ключевых особенностей. Символьные системы всегда могут объяснить результаты вывода, так как известны как правила вывода, так и входные данные [Regina, 1998]. Даже в случае обработки НЕ-факторов знаний вполне можно объяснить, как получены те или иные результаты или приняты те или иные решения.

В случае создания и использования гибридных ИИ-систем, сети глубинного обучения готовят для принятия решений универсальным решателем базовую информацию. Фактически, именно на этом уровне и будет заканчиваться «необъяснимое» в работе таких систем. Но человеческое сознание работает примерно так же. Например, если человек видит кошку, то он обычно не может объяснить, как именно он распознал в животном именно кошку (все рационализирующие объяснения на тему «это маленькое животное с острыми ушками и полосатым хвостом» не годятся, так как это именно рационализация постфактум). Но далее символ «кошка» и все его над-символы могут участвовать в дальнейшем выводе, и результаты такого вывода могут быть вполне объяснены.

3.2.6 ГО пока плохо интегрируется с уже существующими знаниями

Эта проблема является следствием всех перечисленных ранее. Чтобы интегрировать искусственные нейросети с уже накопленными и даже формализованными знаниями, необходимо проделать большую работу по приведению таких знаний в тот вид, который могут воспринять системы ГО. Однако многие знания уже приведены к виду, в котором их могут воспринять символьные системы, а, значит, и гибридные. И работа по формализации знаний ведётся именно в этом направлении — знания представляются в рамках символьных формализмов, а не размечаются для потребления нейронными сетями. Размечаются данные в каждом конкретном случае, а общие знания формализуются.

Это значит, что имеет смысл сосредоточиться именно на совмещении двух подходов, а это путь к гибридизации нисходящей и восходящей парадигмы. Гибридная ИИ-система по своей природе будет подготовлена для интеграции со всем корпусом имеющихся знаний.

3.2.7 ГО пока что неспособно автоматически отличать причинно-следственную связь от корреляции

Фактически даже люди со своим естественным интеллектом часто не могут отличить причинно-следственную связь от корреляции. Естественный отбор способствовал тому, что неокортекс человека склонен находить причинно-следственные связи там, где их может и не быть [Haselton, 2000]. С точки зрения эволюции это вполне разумно, так как лучше избежать мнимой угрозы и выжить, чем долго размышлять на тему, является ли одно событие прямым следствием другого, в результате чего погибнуть в пасти тигра или крокодила. Так что излишняя склонность видеть причинно-следственные связи у людей вполне может быть инструментом снижения ошибок второго рода в распознавании образов [Кэрролл, 2005].

Вместе с тем формальная логика даёт все необходимые инструменты для определения причинно-следственных связей в результатах наблюдений. Чаще всего приходится использовать аппарат нечёткой логики и теории доверия, поскольку наблюдения или эксперименты обычно никогда не бывают «чистыми». Так что приходится закладывать некоторую нечёткость на случай, что в результатах есть примеси невыявленных факторов. Поэтому одновременное применение методов математической статистики (нахождение корреляций) и формальной логики (нахождение причинно-следственных связей) работает вполне достоверно.

Технологии ГО с лёгкостью имплементируют статистические методы, но формальная логика — это вотчина символического подхода. Поэтому для разрешения этой проблемы нейронных сетей опять требуется гибридизация парадигм.

3.2.8 ГО пока что хорошо работает в роли аппроксимации, но его ответам нельзя полностью доверять

Эта проблема является следствием сложности интерпретации результатов, получаемых на выходе нейронных сетей ГО, и предыдущей проблемы. Да, ГО неплохо справляется со статистическими моделями, как это уже было показано. Аппроксимация — это упрощение, замена сложного вычислительного процесса более простым [Лоран, 1975]. Если некоторый вычислительный процесс можно хотя бы приблизительно представить в виде произведения матриц, то методами ГО вполне можно аппроксимировать такой процесс.

Вполне возможно, что это сродни тем процессам, которые происходят внутри естественного интеллекта. Нейронные сети в нервной системе человека (и даже многих животных) вполне успешно аппроксимируют довольно непростые вычислительные задачи. Как поймать рукой брошенный мяч? Быстро решить систему дифференциальных уравнений второго порядка? Есть большие сомнения в том, что какой-то биохимический процесс в нейронах решает подобные уравнения. Но, напротив, нейронные сети вполне успешно аппроксимируют решение таких задач после многочисленных актов обучения. Однако при этом любому человеку будет крайне сложно объяснить, как он ловит брошенный мяч.

И опять на помощь приходит гибридная парадигма. Не важно, как настроились весовые коэффициенты в нейронной сети. Если сеть обучилась успешно аппроксимировать какой-либо процесс, то на более высоком уровне абстракции такие результаты можно объяснить при помощи применения тех или иных правил (как минимум, объяснить можно так же, как эти результаты объяснил бы человек). И тогда далее можно было бы использовать аппарат какой-либо теории

доверия (например, той же теории Демпстера — Шафера) для получения степени доверия к полученному выводу и его объяснению.

3.2.9 ГО тяжело использовать в прикладных целях

Ну и, наконец, в оригинальной статье Маркуса описана проблема в некотором роде упаднического характера. Но, на самом деле, проблемы тут нет. Системы ГО вполне можно применять в различных прикладных задачах, просто сама отрасль ещё достаточно молода, чтобы получить прикладное развитие. Тем не менее, процессы уже идут. Системы машинного зрения уже давно становятся на службу таких отраслей, как обеспечение безопасности и организация дорожного движения. Распознавание инцидентов на автомобильных дорогах и в зонах контроля видеонаблюдения — самые что ни на есть прикладные задачи. Рекомендации по результатам покупок или голосований в социальных сетях — тоже прикладные задачи, успешно решаемые методами глубинного ГО.

Конечно, методы ИИ вообще ещё не так широко применяются, как того хотелось бы. В целом, ИИ войдёт во все сферы жизни человека и общества и изменит их, часто до неузнаваемости. В частности, точно подвергнутся самому серьёзному влиянию следующие аспекты [Душкин, 2017]:

- Государственное и муниципальное управление;
- Обеспечение личной и общественной безопасности;
- Транспорт и логистика;
- Промышленность;
- Образование;
- Наука;
- Здравоохранение;
- Культура.

Можно предположить, что во всех этих сферах жизни появятся и будут развиваться мощные системы поддержки принятия решений (СППР), которые, по своей сути, представляют яркие примеры реализации нисходящей парадигмы [Терелянский, 2009]. Но эти СППР далее будут впитывать в себя методы ГО и другие технологии восходящей парадигмы, и, в конечном итоге, станут гибридными СПР — системами принятия решений, из которых пресловутый человеческий фактор зачастую исключён в принципе.

Можно показать, что использование гибридных систем даёт ключ к решению всех девяти проблем, обозначенных Г. Маркусом.

Всё перечисленное подводит к главной идее — для разработки систем искусственного интеллекта необходимо использовать оба подхода для получения наиболее эффективных результатов. При этом каждая из технологий будет применяться для реализации той части задач, к которым она максимально приспособлена. При таком гибридном подходе каждая технология предстанет со своими сильными положительными свойствами, а проявления слабых свойств можно будет избежать. Такой подход позволит создать ИИ-системы, которые будут обладать мощностью ГО для распознавания образов и других подходящих для машинного обучения задач, но при этом будут хранить и обрабатывать полученные данные в семантическом виде и, соответственно, иметь возможность объяснения полученных результатов и принятых решений. В

свою очередь, это позволит снять многочисленные вопросы этического характера, возникающие при использовании ИИ-систем.

4 Ключевая задача

4.1 При всем богатстве выбора ...

Суть нашего подхода состоит в том, что решение проблемы представления знаний и создание соответствующей МВ на начальном этапе могут рассматриваться как самостоятельные задачи. При этом приоритетной является решение проблемы представления знаний.

Наше отношение к вопросу очередности создания обсуждаемых подсистем состоит в том, что задача формирования структуры данных должна быть решена до начала совершенствования МВ. С целью обоснования сформулированной позиции обратимся к образу той целевой МВ, которая в конечном итоге понадобится для задачи реализации УМВ.

В описываемом случае УМВ должна представлять собой гораздо более сложный программный комплекс, чем представлялось ранее. Активность этого комплекса должна модифицироваться в зависимости от целого ряда контекстных переменных (конкретного предметного мира, масштабных параметров времени, размеров и пр.). Такие модификации должны приводить к настройке алгоритмов работы УМВ. Например, в случае разных масштабов рассматриваемых миров (квантовый мир — мезомир — мегамир) должны меняться и контекст (активная часть фактографической БД), и действующие в этом мире закономерности (активная часть БЗ).

Учитывая специфику необходимой в нашем случае УМВ, можно смело утверждать, что её создание является отдельной задачей, требующей специального подхода и знаний. Обратим также внимание на то, что во всех случаях (разработка, тестирование, модернизация, эксплуатация) УМВ должна использовать какие-то уже существующие данные: модельные или реальные. Это и даёт основания утверждать, что решение проблемы представления знаний — ключевая и приоритетная задача. Адекватность этого решения будет существенно влиять на свойства любой надстраиваемой на этой основе ИИ-системы.

4.2 ... выбор очевиден

Обратим внимание на то, что

- адекватность разработанного представления знаний может быть проверена только в условиях её эксплуатации;
- для эксплуатации разработанного представления знаний надо создать некую интеллектуализированную систему;
- семантический «барьер вхождения» представляет собой очень серьёзное препятствие на пути создания работающей версии любого варианта ИИ-системы (без этого любая ИИ-система существенно ущербна);
- признаем, что наполнить «вручную» любой вариант развитой ЭС или ИИ-системы необходимыми им знаниями, практически невозможно.

Решить весь комплекс указанных задач представляется возможным за счёт целенаправленного создания специализированной ЭС, предназначенной для обработки текстовой информации с целью пополнения своей базы знаний и замещения в ней устаревших знаний. В процессе работы такая ЭС должна постоянно «умнеть» и иметь возможность откладывать «непонятые» тексты на

повторную обработку через некоторое время. Это должно существенно сократить объём экспертных и человеческих ресурсов в целом, требуемых в процессе первичного наполнения, а в дальнейшем привести к возможности «самообучения».

Специалисты практически едины во мнении, что семантические сети потенциально способны к самообучению³, однако для современных семантических сетей реализация этой функции является проблематичной и требует постоянной работы предметных специалистов. Более того, по мере роста базы знаний потребность в присутствии предметных специалистов не уменьшается, а требования к их квалификации только повышаются.

Уточним, что «самообучением» семантической сети и ЭС, основанной на ней, в этом документе мы называем процесс обработки текстовой информации, который:

- может частично производиться в автономном режиме;
- должен требовать вмешательства эксперта и инженера по знаниям только в специально предусмотренных случаях;
- должен обладать всеми необходимыми свойствами, обеспечивающими рост автономной части работы и уменьшение потребности во вмешательстве человека.

При этом, скорее всего, достичь полной автономности самообучения можно только в далёкой перспективе. В обозримом будущем обучающаяся семантическая сеть всегда будет обращаться к эксперту за помощью (функции «переспросить» и «уточнить» в ситуациях «не_понял» и «обнаружено_противоречие»). Важно, что по ходу обучения системы среднее число таких обращений на единицу обрабатываемого текста должно неуклонно падать.

Фактически конечным результатом разработки самообучающейся семантической сети является создание специфической ЭС (когнитивного вычислителя), способной обрабатывать тексты, пополняя новой информацией свою базу знаний.

Подчеркнем, что в данном тексте самообучением ЭС мы называем осуществляемый ею самой процесс обработки текстов с целью пополнения своей базы знаний.

5 Заключение

Рассчитывая создать полноценную информационную систему, обладающую всеми желаемыми «магическими» свойствами, необходимо заложить в её фундамент решения, не ограничивающие её полноты и гибкости. Недопустимо экономить на фундаменте, так как в конечном итоге его свойства определяют максимум количества этажей, которые на нём можно надстроить.

³ В литературе термины «обучение» и «самообучение» часто используется слишком вольно. Так автоматический подбор параметров модели может быть назван её «обучением», что представляется нам некорректным. В этом тексте указанные термины используются в их исходном общезыковом значении.

Список литературы

1. *Abraham A.* Adaptation of Fuzzy Inference System Using Neural Learning, *Fuzzy System Engineering: Theory and Practice*, Nadia Nedjah et al. (Eds.), *Studies in Fuzziness and Soft Computing*, Springer Verlag Germany, Chapter 3, pp. 53–83, 2005. ISBN 3-540-25322-X.
2. *Bart K.* *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence.* Englewood Cliffs, NJ: Prentice Hall, 1992. ISBN 0-13-611435-0.
3. *Dempster A. P.* A generalization of Bayesian inference, *Journal of the Royal Statistical Society, Series B*, Vol. 30, pp. 205–247, 1968.
4. *Haselton M.* Error Management Theory. *Journal of Personality and Social Psychology.* January 2000.
5. *Huang S., Papernot N., Goodfellow I., Duan Y., Abbeel P.* Adversarial Attacks on Neural Network Policies, 2017. arXiv, cs.LG.
6. *Kansky K., Silver T., Mély D. A., Eldawy M., Lázaro-Gredilla M., Lou X.* et al. Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics, 2017. arXiv, cs.AI.
7. *Knight W.* DARPA is funding projects that will try to open up AI's black boxes. MIT Technology Review. Retrieved 2017-11-02.
8. *Marcus G.* *The Algebraic Mind: Integrating Connectionism and cognitive science.* Cambridge, Mass.: MIT Press, 2001.
9. *Marcus G.* *Deep Learning: A Critical Appraisal* — New York University, 2017. — <https://arxiv.org/abs/1801.00631>.
10. *Martin H., Jose A.; Javier de Lope; Darío Maravall.* Adaptation, Anticipation and Rationality in Natural and Artificial Systems: Computational Paradigms Mimicking Nature. *Natural Computing.* Springer, 2009. 8 (4): 757–775. doi:10.1007/s11047-008-9096-6.
11. *Medsker L. R.* *Hybrid Intelligent Systems.* — Boston: Kluwer Academic Publishers, 1995. — 298 p.
12. *Mikolov T., Chen K., Corrado G., Dean J.* Efficient Estimation of Word Representations in Vector Space. arXiv, 2013.
13. *Miller G. A., Chomsky N.* Pattern conception. Paper for Conference on pattern detection, University of Michigan. 1957.
14. *Minsky M.* *Computation: Finite and Infinite Machines.* Englewood Cliffs, N.J.: Prentice-Hall. ISBN 0-13-165449-7.
15. *Regina B., McCullough D., Rambow O., DeCristofaro J., Korelsky T., Lavoie B.* A New Approach to Expert System Explanations. USAF Rome Laboratory Report, 1998.
16. *Shafer G.* *A Mathematical Theory of Evidence,* Princeton University Press, 1976.
17. *Socher R., Huval B., Manning C. D., Ng A. Y.* Semantic compositionality through recursive matrix-vector spaces. *Proceedings from Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning,* 2012.
18. *Sowa J. F., Borgida A.* *Principles of Semantic Networks: Explorations in the Representation of Knowledge,* 1991.
19. *Weiss G.,* ed. by, *Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence,* MIT Press, 1999, ISBN 0-262-23203-0.
20. *Wooldridge M.* *An Introduction to MultiAgent Systems,* John Wiley & Sons Ltd, 2002, paperback, 366 p., ISBN 0-471-49691-X.
21. *Бир С.* *Мозг фирмы.* — М.: Радио и связь, 1993. — 416 с. ISBN 5-256-00426-3.
22. *Голованов Г.* Gartner: «Хайп вокруг ИИ идёт во вред технологии и потребителям» — Хайтек, 19 июля 2017. — <https://hightech.fm/2017/07/19/ai-hype>.

23. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение = Deep Learning. — М.: ДМК Пресс, 2017. — 652 с. — ISBN 978-5-97060-554-7.
24. Добров Б. В., Иванов В. В., Лукашевич Н. В., Соловьёв В. Д. Онтологии и тезаурусы: модели, инструменты, приложения. — М.: Бином. Лаборатория знаний, 2009. — 173 с. — ISBN 978-5-9963-0007-5.
25. Душкин Р. В. Методы получения, представления и обработки знаний с НЕ-факторами. — 2011. — 115 стр., ил.
26. Душкин Р. В., Рыбина Г. В. Об одном подходе к автоматизированному извлечению, представлению и обработке знаний с НЕ-факторами. В кн.: Известия РАН. Теория и системы управления, № 5. — М.: 1999. — стр. 34-44.
27. Душкин Р. В., Жарков А. Д., Иванов Д. А. От безопасного к умному городу — Нижний Новгород: ИТ Форум 2020. «Развитие цифрового государства. Создание систем весогабаритного контроля, интеллектуальных транспортных систем». — 12-14 апреля 2017 года.
28. Заде Л. Понятие лингвистической переменной и его применение к принятию приближённых решений. — М.: Мир, 1976. — 166 с.
29. Бродский Ю.И. 2013. Модельный синтез и модельно ориентированное программирование. ВЦ им.Дородницына
30. Зайцев И. М., Федяев О. И. Агентно-ориентированный подход к моделированию интеллектуальных распределённых систем. Сб. — Донецк: ДонНТУ, 2008. — С. 337-338.
31. Каллан Р. Основные концепции нейронных сетей = The Essence of Neural Networks First Edition. — М.: Вильямс, 2001. — 288 с. — ISBN 5-8459-0210-X.
32. Колесников А. В. Гибридные интеллектуальные системы: Теория и технология разработки. — СПб: Изд-во СПбГТУ, 2001. — 711 с.
33. Кондаков Н. И. Обобщение // Логический словарь-справочник, Изд. 2, М.: Наука, 1975, с. 395.
34. Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика. — М.: Горячая линия — Телеком, 2001. — 382 с. — ISBN 5-93517-031-0.
35. Кэрролл Р. Т. Энциклопедия заблуждений: собрание невероятных фактов, удивительных открытий и опасных поверий. — М.: Издательский дом «Вильямс», 2005. — 672 с. — ISBN 5-8459-0830-2, ISBN 0-471-27242-6.
36. Лоран П. Ж. Аппроксимация и оптимизация. — М.: Мир, 1975. — С. 496.
37. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение. — СПб.: Питер, 2018. — 480 с. — ISBN 978-5-496-02536-2.
38. Терелянский П. В. Системы поддержки принятия решений. Опыт проектирования: монография / П. В. Терелянский; ВолгГТУ. — Волгоград, 2009. — 127 с.
39. Флах П. Машинное обучение. — М.: ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7.
40. Хайкин С. Нейронные сети: полный курс = Neural Networks: A Comprehensive Foundation. 2-е изд. — М.: Вильямс, 2006. — 1104 с. — ISBN 0-13-273350-1.
41. Эшби У. Р. Введение в кибернетику. — М.: Изд. иностр. лит., 1959. — 432 с.
42. Шевченко. О взаимосвязи формализации описательных наук, когнитивного анализа, «искусственного интеллекта», теории игр и теории КЛС. — Цифровая экономика, 2019