3.2. OPENTALKS.AI: КОНФЕРЕНЦИЯ 20-21 ФЕВРАЛЯ 2020 ГОДА

Милкова М.А. – научный сотрудник ЦЭМИ РАН

Краткий обзор конференции по искусственному интеллекту OpenTalks.AI, прошедшей в Москве 20-21 февраля 2020 года. Конференция была посвящена последним достижениям в области компьютерного зрения, анализа естественного языка, предиктивной аналитики, обучения с подкреплением и общего искусственного интеллекта, а также включала различные дискуссии по вопросам безопасности искусственного интеллекта и перспективам развития общества в целом.

Конференция OpenTalks.AI - независимая площадка для всех участников отрасли ИИ, объединившая предпринимателей, разработчиков, заказчиков, ученых, стартапы и инвесторов. Организатором мероприятия является АйПи Лаборатория, руководитель конференции - Игорь Пивоваров.

Разумным стилем мероприятия было стремление к взаимопониманию между различными категориями участников, синтез докладов как теоретического, так и прикладного характера, дружеская атмосфера и четкие рамки.

Первый день конференции был посвящен компьютерному зрению (Computer Vision, CV) и анализу естественного языка (Natural Language Processing, NLP). Второй день освещал вопросы предиктивной аналитики (Predictive Analytics, PA), обучения с подкреплением (Reinforcement Learning, RL) и общего искусственного интеллекта (Artificial General Intelligence, AGI).

Первый день: Computer Vision, Natural Language Processing

Первый день – высочайшая плотность как докладов, так и их содержания, вдохновляющие многих результаты – все это характеризуют стремительное развитие событий в области CV и NLP.

Доклады Виктора Лемпицкого (Samsung, Skoltech), Антона Конушина (Samsung, ВШЭ) главным образом обсуждали технологии CV, основанные на генеративно-состязательных нейросетях (Generative adversarial network, GAN). Первые GAN-ы были предложены еще в 2014 году (Goodfellow et al., 2014) и с тех пор получили стремительное развитие. Основная идея GAN-ов заключается в том, что первая нейросеть генерирует случайные числа из заданного распределения и генерирует из них объекты, которые идут на вход второй сети. Вторая сеть – дискриминатор – получает на вход объекты из выборки и объекты, созданные генератором, и учится предсказывать вероятность того, что сгенерированный объект – реальный. При этом генератор тренируется создавать объекты, которые дискриминатор не отличит от реальных. Таким образом, GAN обучают создавать структуры, похожие на сущности из нашего мира: изображения, музыку, речь, текст. Последние достижениями в данной области включают возможности генерации лиц (StyleGAN) и основаны на операции Adaptive Instance Normalization (AdaIN) (Huang, Belongie, 2017). Предобученные модели находятся в открытом доступе и свободны для использования.

В докладах Александра Крайнова (Яндекс), Александра Ханина (VisionLabs), Семена Буденкова (IntelliVision), Виктора Субботина (Beta) были приведены более реальные примеры, иллюстрирующие достижения в области компьютерного зрения и касающиеся не генерации, а анализа изображений и видео:

- Контроль соблюдений правил безопасности на производстве: выгоднее устанавливать камеры, позволяющие детектировать возможные нарушения (работник находится без каски; отсутствие перчаток и т.п.), чем нести убытки после произошедшего несчастного случая.
- Осуществление контроля доступа по лицу (на проходных системах, на рабочем месте).
- Системы контроля внимания для анализа состояния водителя за рулем.
- Анализ перемещения посетителей в магазинах для прогнозирования и предупреждения очередей на кассах, анализ поведения у полок с товарами.
- Управление очистными сооружениями на основе видео с камер.

Расширение возможностей по анализу изображений, согласно прогнозу, позволит в ближайшем будущем повсеместно осуществлять переводы по баркоду, который будет привязан к человеку. В дальнейшем будет осуществлен переход к единой системе аутентификации, соединяющей фотографию лица, паспортные данные, информацию о банковской карте. Примером может быть упрощение прохождения регистрации и контроля в аэропортах, обслуживание в банках и банкоматах без документов и карт. В качестве примера, считающегося эталонным, рассматривается Amazon Go – оффлайн-магазины, использующие технологию Selfie ToPay, позволяющие расплачиваться за товары посредством изображения лица. Предельным случаем является активное внедрение электронных адаптивных ценников, показывающих разные цены разным покупателям.

Компьютерное зрение не ограничивается только анализом непосредственно изображения, а может быть расширено другими модальностями. Доклад Ольги Перепелкиной (Neurodata Lab LLC) представил подход, согласно которому распознавание эмоций может происходить в том числе с помощью анализа частоты сердцебиения - изменения цвета пикселей в отдельных областях лица. Более того, технологии

позволяют достаточно успешно учитывать и дыхание – воспринимать микродвижения груди (в лабораторных условиях с помощью обычной веб камеры удается достичь высокой точности – ошибка составляет менее 1 вздоха в минуту). Применение такого рода технологий планируется использовать для таргетирования оффлайн рекламы, грубого фильтра кандидатов при приеме на работу (полиграф). Несмотря на то, что в Трудовом кодексе нет упоминания о такой форме взаимодействия работника и работодателя, как полиграф, это не мешает некоторым из компаний использовать такого рода средства для первичного отбора кандидатов.

Интересным примером применения компьютерного зрения является поиск плагиата в логотипах (Роман Доронин, Eora). В отличие от обычных изображений, у логотипа есть сущность, что не позволяет применять для работы с ним стандартные подходы анализа изображений. Например, перевернутое изображение пумы с логотипа Puma, значок Adidas с двумя, а не тремя листками или же растянутый логотип Макдональдс. Для человека сходство искаженных вариантов с оригинальными очевидно, но не для компьютера. Однако и такие вопросы в настоящее время удается решать с помощью аугментации (добавления, изменения признаков). Данная технология позволяет значительно сократить время работы патентного поверенного, а также избежать потенциальных претензий за нарушение авторских прав.

«Большие тексты – зеркало в человеческий мозг» (Татьяна Шаврина, Сбербанк, НИУ ВШЭ)

Сессия по NLP включала доклады Валентина Малых (Huawei), Аркадия Сандлера (МТС), Татьяны Шавриной (Сбербанк, НИУ ВШЭ). Основные тренды в данной области безоговорочно включают BERT (Bidirectional Encoder Representations from Transformers) – технологию предварительного обучения, основанную на нейронных сетях и разработанную Google в 2018 году (Devlin et al., 2019). Основная идея BERT состоит в том, чтобы подавать на вход нейросети (на архитектуре Transformers - Vaswani, 2017) фразы, в которых 15% слов заменены на маску. Нейронная сеть учится предсказывать эти закрытые маской слова. Отметим, что BERT была создана на основе разработки компании OpenAI – нейронной сети GPT. Предобученные модели BERT находятся в открытом доступе. Использование BERT позволяет решить две задачи: предсказывания слова по контексту и определения, является ли одно предложение продолжением другого.

Впоследствии BERT была оптимизирована в версии A Lite BERT (ALBERT) - с помощью снижения количества параметров нейросети (параметризация векторных представлений и обмен весов между слоями нейросети). Предварительно обученные языковые модели, как правило, требуют больших вычислительных затрат и занимают много памяти, поэтому их сложно эффективно применять на устройствах с ограниченными ресурсами. Чтобы ускорить вывод и уменьшить размер модели при сохранении точности, в конце 2019 года была предложена еще одна модель – TinyBert, использующая в своей основе новый метод Transformer distillation (Jiao et al., 2019).

Альтернативным подходом к векторному представлению слов является ELMo (Embeddings from Language Models) от компании AllenNLP. ELMo применяет языковые модели, чтобы получить векторное представление для каждого слова, и использует двунаправленную долгую краткосрочную память (Long short-term memory).

Все модели относятся к машинному обучению без учителя, то есть основаны на тренировках на большом массиве данных, как правило, это Wikipedia, Reddit и др. Русскоязычные версии обучаются на аналогичных русских ресурсах. Общий подход к NLP в настоящее время заключается в том, чтобы адаптировать заранее обученную модель (BERT, ELMo или т.п.) для решения конкретной задачи.

Продвижения в области NLP показывают свою эффективность, например, при прохождении экзамена ЕГЭ по русскому языку (включая не только выполнение задания, но и распознавание и «понимание» условия задания). Система способна набирать около 70 баллов.

Что касается рекомендательных систем, то помимо стандартных, используемых на веб страницах, рекомендации создаются для выдачи автоматизированных подсказок сотрудникам call-центров (Михаил Горбатовский, IBM). Также рекомендации актуальны для составления подборок документов, например, научных статей (Константин Воронцов, МФТИ, ФИЦ ИУ РАН). Анализ потока новостей также использует методы автоматического текстового анализа (Алексей Бурнаков, ИТАР-ТАСС).

Однако актуальна не только автоматическая обработка текстов. Так, профессией будущего в Китае считается разметчик – человек, вручную проставляющий признаки и типовые характеристики слов, текстов, картинок и т.п. По-прежнему остаются области, для которых использование преднастроенных нейросетей неприменимо – например, в области машинного перевода в узких областях (Елизавета Иванова, ЭГО Транслейтинг). В России популярность набирает Яндекс.Толока – краудсорсинговый проект, позволяющий быстро разметить большое число данных, которые будут впоследствии использоваться в машинном обучении. Использование данного сервиса для обучения голосового ассистента Huawei было изложено в докладе Мурата Апишева (Digital Decisions) и Ирины Пионтковской (Huawei)

Второй день: Predictive Analytics, Reinforcement Learning and Artificial General Intelligence

Если в области CV, NLP множество задач остаются нерешенными, то в области предиктивной аналитики кажется, что все стандартные задачи уже решены и, в случае необходимости, автоматизированы. В противовес все более сложным алгоритмам CV, NLP, спикеры сессии по PA (Алексей Драль, BigData Team; Эмили Драль, Mechanica.ai; Андрей Устюжанин, НИУ ВШЭ, Яндекс) советовали отдавать предпочтения простым алгоритмам, однако основывать их на большей выборке. Кроме того, на первый план

выходит требование объяснимости моделей и их устойчивости. Происходит переход от прогнозирования к сценарному анализу.

Второй день во многом затрагивал вопросы безопасности и доверия к ИИ (Александр Поляков). Так, согласно оценкам, доверия к ИИ со стороны людей оправданно нет. Безопасность – та вещь, которая непосредственно пересекается с доверием. Библиографический анализ статей в области безопасности ИИ отмечает стремительный рост публикаций, начиная с 2017 года. Согласно историческим данным, временной промежуток от начала первой статьи по безопасности чего-либо до первого инцидента в реальной жизни составляет 5-7 лет, таким образом, ближайшее время будет очевидно сопряжено с явными проблемами безопасности ИИ: фейковые изображения, аудио, видео, инъекции в исходные данные, на которых происходило обучение, позволит успешно обходить различные системы распознавания.

Отдельная панельная дискуссия была посвящена этике и регулированию в области ИИ (Елизавета Афанасьева, IPChain; Екатерина Калугина, «Дабл»; Елена Введенская, РНИМУ им. Н.И. Пирогова, ИНИОН РАН; Николай Лукашов, Академия управления МВД; Юрий Цветков, МИД РФ). Сейчас ИИ находится в эпицентре создания новых как правовых, так и этических норм. Так, в связи с возможностью создания ИИ разнообразного «творческого» контента (генерация музыки, изображений, текстов), возрастает актуальность вопросов, связанных с охраной такого рода произведений. Возможно ли признать систему ИИ субъектом права?

Отдельной проблемой является монополизация данных. Сейчас большая часть данных сосредоточена в руках цифровых гигантов. Так, в США компания HiQ, занимающаяся созданием сервиса для HR аналитики и анализирующая открытые профили сети Linkedin, была обвинена в нарушении антихакерского законодательства. Судом было постановлено, что запрет доступа к открытой информации позволит корпорациям, обладающим самыми большими базами данных, контролировать использование общедоступной информации в Интернете, то есть ограничит конкуренцию. В России на данный момент отсутствует ясность по теме доступных данных. К примеру, спор между компанией Stafory, разрабатывающей ПО по автоматизированному подбору персонала (робот Вера), и HeadHunter, блокирующей сторонних пользователей, собирающих информацию, был разрешен ФАС в пользу Stafory. Аналогичный спор между компанией Double Data, воспользовавшейся открытыми данными сети ВКонтакте для создания собственного поискового индекса, и сетью ВКонтакте был начат в 2017 году и до сих пор не разрешен.

Неоднозначен и вопрос в целом о том, что именно можно считать персональными данными. Так, в Европе лицо считается персональными данными, что, к примеру, не позволяет размещать камеры ближе 5 метров до человека.

Остро и неоднозначно вопросы применения ИИ стоят и в медицине. В настоящее время активно разрабатываются алгоритмы для оценки вероятности осложнения заболеваний, помощи в постановке диагноза и назначения лечения, анализа тяжелобольных пациентов в режиме реального времени. Известен случай, когда IBM Watson выявил у пациентки редкую форму лейкемии, изучив 20 миллионов научных статей о раке всего за 10 минут. Однако несмотря на некоторые положительные примеры, вопросов и сомнений в данной области пока больше. Использование ИИ нарушает права пациентов на сохранение конфиденциальности личных данных, обнародует врачебную тайну. Кроме того, ИИ может по-разному влиять на принятие решений врачом. Врач рискует попасть в своего рода когнитивную ловушку. Так, по данным экспериментов, система Botkin способна принимать решения на одном уровне с московским врачом и значительно лучше провинциального врача. Каковы должны быть действия в таком случае провинциального врача? Должен ли он признать компетенции системы и просто следовать ее указаниям или же стремиться наращивать свой уровень?

Другим вопросом является необходимость самоидентифицирования системы ИИ. В США принят однозначный положительный ответ на данный вопрос, в России данной проблемой на текущий момент не занимаются, но к согласию не удалось прийти и в ходе дискуссии на конференции. С одной стороны, для гражданина должно быть только важно иметь обратную связь для гарантированного решения проблемы, с другой – человек обладает правом знать, общается ли он с человеком или с искусственной системой.

Отдельная секция была посвящена обучению с подкреплением (RL) и Общему искусственному интеллекту (AGI). Обзорная сессия по RL включала доклады Сергея Николенко (ПОМИ РАН, Neuromatio), Сергея Свиридова (Цифра), Сергея Колесникова (Тинькофф). RL – способ машинного обучения, в ходе которого обучение происходит через взаимодействие с окружающей средой. Классическими примерами являются системы игры в шахматы. Так, система AlfaGo в 2016 одержала победу над корейским шахматистом Ли Седолем со счетом 4-1. Однако единственное победное очко стало последним в глобальном шахматном соревновании человека с машиной. С того времени усовершенствованные системы (AlfaZero, MuZero) отошли от подхода анализа лучших партий лучших шахматистов и основаны на функции оценки позиций (AlfaZero) и функции динамики окружающей среды (MuZero). Последние версии систем одерживают безоговорочные победы над шахматистами, а также в играх StarCraft, Dota и др.

Другим примером является разработка робота, обучающегося на синтетических данных и способного собирать кубик-рубик.

Использование RL алгоритмов для решения реальных задач без участия человека пока сильно ограничено. Однако некоторые примеры есть. Так, агент, обученный с помощью RL, управлял системой

охлаждения больших дата-центров. Есть примеры успешного обучения сортировочных роботов, а также создания конфигураций труб со сложными гидродинамическими свойствами, которые не удавалось сделать человеку.

Последние исследования в области RL посвящены вопросам того, как заставить виртуального агента придумывать новые стратегии взамен поиска одной наилучшей. Ответ – посредством обучения через конкуренцию.

Итоги подводил Евгений Кузнецов (Orbita Capital). Завораживающая и драматичная картина, разворачивающаяся на наших глазах, не оставляет никого в стороне. Согласно историческим данным, каждая новая технология закрепляется все быстрее и массовое распространение ИИ произойдет за годы. В настоящее время происходит совместное познание мира как людьми, так и роботами. Вырисовывается новый тип познания, где человек и робот работают колоборативно, но размываются границы, где заканчивается человеческое и начинается машинное. Происходит интегрирование своего индивидуального сознания с сознаниями других людей, которые создавали внешние системы и от того, насколько это интегрирование будет успешным, и будет зависеть то, насколько общество будет комфортным в итоге.

Литература

1. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Networks. arXiv:1406.2661v1
2. Huang, X., Belongie, S. (2017). Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. arXiv:1703.06868v2
3. Devlin, J., Chang, M-W, Lee, K.,Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). Attention Is All You Need. arXiv:1706.03762v5
5. Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., Liu, Q. (2019). TinyBERT: Distilling BERT for Natural Language Understanding. arXiv:1909.10351v4

Милкова Мария Александровна –научный сотрудник лаборатории экспериментальной экономики ЦЭМИ РАН (m.a.milkova@gmail.com)

Ключевые слова

искусственный интеллект, компьютерное зрение, анализ естественного языка, предиктивная аналитика, безопасность искусственного интеллекта

Maria Milkova, OpenTalks.AI: Conference 20-21 February 2020

Keywords

artificial intelligence, computer vision, natural language processing, predictive analytics, artificial intelligence security.

DOI: 10.34706/DE-2020-01-08
JEL classification: D83 Search, Learning, and Information

Abstract

A brief overview of the conference on artificial intelligence OpenTalks.AI, held in Moscow on February 20-21, 2020. The conference was devoted to the latest achievements in the field of computer vision, natural language processing, predictive analytics, reinforced learning and general artificial intelligence, and also included various discussions on the security of artificial intelligence and the prospects for the development of society as a whole.