SMART Закупка

14 апреля 2021

Прорыв в области искусственного интеллекта: MMT от Facebook переводит 100 языков

Развитие технологии искусственного интеллекта достигло значительных вех. От редактора ИИ, получившего награду в области цифровой журналистики, до Facebook, представляющего многоязычную модель машинного перевода для 100 языков, на этой неделе Toolbox освещает важные достижения в области ИИ.

Глобальный экономический климат побудил несколько организаций сделать стратегический шаг в сторону автоматизации и искусственного интеллекта. Неудивительно, что наблюдается всплеск инвестиций в ИИ и автоматизацию. Такое широкое внедрение ИИ изменило представление о бизнес-операциях по всему миру. Благодаря множеству преимуществ, исследователи и специалисты по обработке данных постоянно стремятся создать новые эталоны в мире искусственного интеллекта. На этой неделе Toolbox делится значительными достижениями в области искусственного интеллекта.

Редактор AI получил премию в области цифровой журналистики
Искусственный интеллект и машинное обучение принесли огромную пользу различным новостным публикациям. Wall Street Journal использует платный доступ на основе машинного обучения для персонализированного контента на основе подписки, в то время как Reuters разработало инструменты на базе искусственного интеллекта, такие как отслеживание новостей и Lynx Insight для курирования новостей. В мае 2020 года Microsoft заменила более 50 журналистов программным обеспечением AI для Microsoft News и MSN.

В рамках оценки своего бизнеса гигант из Редмонда теперь полагается на ИИ, чтобы курировать новости для своих новостных каналов. Вместе с несколькими технологическими компаниями, экспериментирующими с ИИ в журналистике, в мае 2020 года Google представила бесплатный курс ИИ в сотрудничестве с JournalismAI и VRT News, чтобы помочь журналистам понять силу машинного обучения. 14 октября 2020 года Google запустил Journalist Studio на базе искусственного интеллекта, чтобы журналисты могли работать более эффективно.

В этом году на премии онлайн-журналистики 2020 года канадская компания The Globe and Mail Sophi.io, редактор веб-сайтов AI, получила награду за лучший стартап в области цифровых новостей. Это огромный прорыв в индустрии искусственного интеллекта и показатель того, что искусственный интеллект действительно существует.

Обученный редакционной группой компании, Sophi.io самостоятельно курирует и размещает 99% контента на всех своих цифровых страницах. Он также использует интеллектуальную технологию распознавания лиц, чтобы гарантировать, что цифровая домашняя страница отражает разнообразие страны. На повестке дня компании был редактор веб-сайтов с искусственным интеллектом, чтобы помочь журналистам сосредоточиться на глубоком

анализе новостей и отделе новостей для принятия важных стратегических и тактических решений.

Дэвид Уолмсли, главный редактор The Globe and Mail, сказал : «Отдел новостей будущего - это то место, где журналисты могут сосредоточиться на поиске и рассказе отличных историй - то, что машина не может сделать. Наши специалисты по обработке данных медленно и осторожно автоматизировали веб-страницы, тестируя результаты, прежде чем постепенно внедрять их практически на весь сайт ».

Технические новости: Microsoft объединяется с некоммерческими организациями для разработки инклюзивных моделей искусственного интеллекта

Новая модель многоязычного машинного перевода Facebook переводит 100 разных языков Машинный перевод - прибыльный рынок, который в последнее время приобрел популярность. По прогнозам, к 2025 году рынок машинного перевода достигнет 1042 млн долларов, и этот рост связан с огромным потребительским спросом на переведенный контент и расширением клиентской базы предприятий по всему миру.

На этой неделе, 19 октября 2020 года, Facebook AI предоставил открытый исходный код M2M-100, многоязычной модели машинного перевода (MMT), которая может переводить на 100 языков, не полагаясь на данные на английском языке. На сегодняшний день в моделях MMT используются данные по обучению английскому языку в качестве посредника для перевода на два разных языка.

Компания утверждает, что это первая модель MMT, которая может напрямую переводить 100 × 100 языков в любом направлении, не полагаясь только на англоязычные данные. M2M-100 обучен по 2200 языковым направлениям и достиг 10 баллов по показателям двуязычного оценивания (BLEU) по сравнению с MMT, ориентированным на английский язык.

Анджела Фан, научный сотрудник Facebook, сказала : «Единая модель, поддерживающая все языки, диалекты и методы, поможет нам лучше обслуживать большее количество людей, поддерживать переводы в актуальном состоянии и создавать новые впечатления для миллиардов людей в равной степени. Эта работа приближает нас к этой цели ».

Команда исследователей искусственного интеллекта Facebook создала обширный набор данных с 7,5 миллиардами пар предложений на 100 языках и использовала несколько инструментов интеллектуального анализа данных с открытым исходным кодом, включая ccAligned, ccMatrix и LASER, для поиска миллиардов предложений из Интернета. Затем команда представила стратегию разработки мостов, в которой языки сгруппированы в 14 групп на основе классификации, географии и культурного сходства.

Фан сказал : «Люди, живущие в странах с языками одной семьи, как правило, чаще общаются и выиграют от высококачественных переводов. Например, одна группа будет включать языки, на которых говорят в Индии, такие как бенгали, хинди, маратхи, непальский, тамильский и урду. Мы систематически искали все возможные языковые пары в каждой группе ».

Затем команда собрала обучающие данные для всех возможных комбинаций языков моста и собрала 7,5 миллиардов параллельных предложений данных, соответствующих 2200 направлениям. Фан далее пояснил: «Так как добытые данные можно использовать для обучения двух направлений данной языковой пары (например, en-> fr и fr-> en), наша стратегия добычи помогает нам эффективно выполнять редкую добычу, чтобы наилучшим образом охватить все 100 × 100 (всего 9900) направлений в одной модели ».

Для дальнейшей оптимизации модели MMT исследователи из Facebook использовали стратегию обратного перевода, которая генерировала синтетические данные в дополнение к уже добытым параллельным данным. Фан добавил : «В целом, комбинация нашей стратегии моста и данных с обратным переводом улучшила производительность по 100 направлениям с обратным переводом в среднем на 1,7 BLEU по сравнению с обучением только на добытых данных. Имея более надежный, эффективный и высококачественный обучающий набор, мы получили прочную основу для построения и масштабирования нашей модели «многие ко многим» ».

Facebook выпустил ресурсы модели M2M (набор данных, модель, обучение и настройку для оценки) для исследовательского сообщества, чтобы сделать новые усовершенствования в многоязычных моделях машин. В настоящее время модель M2M-100 находится на стадии исследований, и компания планирует и дальше совершенствовать свою модель MMT и постепенно запускать ее в производство.

Искусственный интеллект является ключевым фактором разрушительных преобразований и может способствовать росту огромной экономики. Искусственный интеллект - это быстро развивающаяся область, которая постоянно учится, вводит новшества и трансформирует предприятия по всему миру. Эта новая волна приведет компании будущего и предложит трансформационные решения как для бизнеса, так и для потребителей.

Ссылка на статью: Прорыв в области искусственного интеллекта: MMT от Facebook переводит 100 языков