



Наталья Соколова

CEO компании Brand Analytics

3.7 Медиа и реклама

Персонально ваш. Технологии NLP для рекламы, для медиа... для жизни

Когда ты живешь в пещере с ближайшей сотней со-пещерников — тебе нет нужды в «переносчиках» коммуникаций: все новости и мечты о будущем тарабанятся около костра во время обгладывания тушки мамонта.

Когда ты живешь в персональной пещере, окруженный миллионами персональных пещер, вообщем-то, неизвестных личностей — и «поговорить-то не с кем!» — вся надежда на единые, исторически сложившиеся каналы потребления информации: папирусы, книги, радио, ТВ, интернет... Да вот беда — если 100 лет назад газет в городе было раз-две, а информации хватало на пару слухов на рынке, то теперь от коммуникаций приходится прятаться за ширму «персональных предпочтений»: кому наушники в ухо, кому смартфон в глаза — лишь бы оградиться от водопада беснующихся букв, иероглифов и образов!

Итого, к началу тысячелетия: медиа разучились интересоваться людьми — потому как не понимают, что людям нужно прямо здесь и сейчас (а не в искривленном воображении редактора или журналиста), а реклама, наоборот, из полезной информизирующей стала навязчивой как старая цыганка — всё-то она про тебя знает, даже лучше тебя самого, а ты нужен лишь чтобы «ручку позолотить».

И вот, к 2020-му году (всего-то через 15–20 лет после начала информационного кризиса), когда информации стало бесконечно много, а количество мозгов для её обработки не увеличилось, в промышленно-государственных масштабах развились системы анализа социальных медиа. С помощью подобного рода систем теперь государство, с трудом, но начало понимать желания, боли и мечты людей, а бизнес — на какие клавиши души лучше всего давить для продвижения очередных материальных благ и оперативно реагировать на проблемы с товарами и услугами.

Еще интереснее поразмышлять, а в каком мире мы окажемся еще через 15–20 лет?

На радость нам оказывается, что системы анализа потоков данных (заметьте, уже речь про «потоки данных», а не генерируемого людьми «соцмедиа» — т.е. и IoT, и датчики на смартфонах, и замеры температуры Гольфстрима, потоков

воздушных масс или транспорта на дорогах) вполне начинают справляться с «задачей будущего»: знать, что будет завтра, а, значит, и что делать, чтобы такое будущее (не)наступило.

А ПОКА БУДУЩЕЕ ЕЩЕ НЕ НАСТУПИЛО

Сегодня у тебя нет второго шанса привлечь внимание пользователя, если ты сразу не попал в точку. Своей рекламой, своей новостью, своим продуктом — своим предложением. И хочешь ты того или нет, ты должен очень хорошо понимать своего адресата, каждого из многих миллионов. Это могут называть изучением клиентского опыта или потребительских трендов — неважно. Суть одна — только слушая пользователя на «родном» языке можно определить его намерения.

Отлично, вы узнали намерения пользователя, но что ему предложить? Вы должны налету анализировать доступные предложения и формировать персональные с учетом клиентских ожиданий. Для этого в интернет-сервисах, а именно они сейчас определяют будущее рекламы и медиа задействован целый стек технологий, относящихся к NLP. Обработка языка и извлечение знаний, классификация и кластеризация, sentiment-анализ и анализ настроений, синтез речи и генерация контента — сегодня все поставлено на карту для завоевания внимания пользователей.

При этом к компьютерной лингвистике, используемой в рекламе и медиа, сегодня предъявляются новые требования. Все алгоритмы должны работать на высоких скоростях в режиме реального времени, «понимать» неграмотный язык социальных медиа и обрабатывать непредсказуемо большой поток данных. Именно поэтому круг разработчиков современных технологий NLP для рекламы, медиа и маркетинга весьма узок.

NLP В РЕКЛАМЕ

В 2018 году произошло знаменательное событие — по данным АКАР^[83] объем интернет-рекламы превысил объем телевизионной, которая продолжает снижаться и в 2019 году. 203 млрд рублей по результатам 2018 года потратили рекламодатели на рекламу в интернете и только 187 млрд рублей на телевизионную рекламу.

При этом львиная доля интернет-рекламы — это контекстная реклама. По данным исследования

«Экономика рунета 2018»^[84] объем контекстной рекламы вырос в 2018 году на 18% и составил 165 млрд рублей. К контексту в данном исследовании отнесена как традиционная контекстная реклама, так и таргетированная реклама в соцсетях. В обоих случаях текстовый анализ лежит в основе упомянутых рекламных систем.

Основными игроками рынка интернет-рекламы являются поисковики **Яндекс** (Директ, РСЯ) и **Google** (Adwords, Double Click) с сервисами контекстной рекламы и партнерскими сетями для тематической. И соцсети **ВК** (MyTarget), **ОК** (MyTarget), **Instagram**, **Facebook** и **Twitter**, представляющие возможности таргетированной рекламы. С полной рекламной экосистемой, включающей рекламные сети, прогностик-платформы, сети нативной и видео рекламы можно ознакомиться на Technology Map от Adindex^[85].

Вывод очевиден — рекламный канал в интернете стал главным рекламным каналом для брендов, а сами рекламные технологии и форматы с успехом применяют NLP для построения нужных рекламодателю аудиторных сегментов, создания рекламных сообщений, выбора точек контакта и последующей аналитики.

Например, если раньше целевая аудитория «задавалась» соцдемом или вы напрямую указывали, например, интересы нужных вам пользователей или профессии, то сегодня аудитории все чаще строятся по принципу look-alike с помощью машинного обучения.

Взяв в качестве эталона, например, часть базы совершивших у вас покупку пользователей, вы можете найти новых потенциальных покупателей, похожих на эталон. Наиболее удачная реализация машинного look-alike у Facebook, который анализирует для поиска «похожих», в том числе, контент, которым интересуется пользователь.

Согласно исследованию Brand Analytics «Социальные сети в России»^[86] ежемесячно в соцсетях высказываются более

46 млн пользователей, публикуя более 1,8 млрд сообщений. Такое тотальное проникновение в нашу жизнь социальных медиа позволяет анализировать мнения миллионов пользователей

46 млн
пользователей

и предоставлять рекламодателям новые возможности для повышения эффективности рекламы.

Системы анализа соцмедиа позволили брендам:

- сегментировать и выявлять новые группы целевой аудитории
- определять интересы пользователей и находить инсайты для рекламных креативов
- медиапланировать размещения на площадках с нужной аудиторией
- выявлять блогеров и инфлюенсеров для продвижения по нужной тематике
- оценивать восприятие креатива и эффективность рекламы в целом

Наиболее распространенные системы мониторинга и анализа соцмедиа представлены в Technology Index от Adindex^[87] (категория Social Media Monitor). Лидером рейтинга является система Brand Analytics, в Топ-5 вошли также Медиалогия и YouScan. Однако для реализации вышеобозначенных возможностей необходимо использовать не мониторинговую, а именно аналитическую систему.

NLP В МЕДИА

Если говорить о классических СМИ, то NLP сейчас дает шанс наиболее прогрессивным издателям создавать действительно востребованный контент и не тратить ресурсы на то, что в итоге никто не читает. Это шанс выжить, стать лидером и создать медиа нового поколения.

При этом конкурировать за читателя СМИ придется с быстрорастущими социальными медиа и новыми платформами по аккумулярованию и распространению авторского и издательского контента. Наблюдению за трендами медиапотребления посвящен регулярный рейтинг от Brand Analytics русскоязычных Топ-платформ и Топ-100 медиаресурсов, наиболее цитируемых в социальных медиа^[88].

Что же NLP дает современному издателю? Поскольку главным и единственным адресатом для СМИ является читатель, то измеряя его реакцию можно с уверенностью определять интересные читателю инфоповоды (рейтинг самых обсуждаемых постов^[89]), темы, форматы, персонажей (рейтинг внимания^[90] к медиаперсонам), авторов, ресурсы и т.д. и т.д. Или даже переа-

пустить СМИ с прицелом на новую аудиторию и востребованную тематику. Чтобы не 5% как сейчас, а 95% создаваемого контента пользовались спросом и медиа вновь стали бизнесом.

Новостные агрегаторы за последние два года потеряли часть внимания пользователей в пользу медийной экосистемы Telegram и персонализированного ньюсфида Яндекс.Дзен. Однако агрегаторы планируют бороться и делают ставку на персонализированный контент, подбираемый в результате наблюдения за интересными читателю материалами и с использованием различных вариантов коллаборативной фильтрации.

У Рамблера — это совсем свежий проект умной новостной ленты на основе ваших интересов Рамблер/Бит^[91], у Mail.ru — рекомендательный сервис «Пульс»^[92]. Туда же движется и Google со своей лентой рекомендаций Discover^[93]. В отличие от других «Пульс» хоть и работает сейчас только с материалами СМИ и блогерами, в перспективе планирует стать полноценной авторской платформой, как и Яндекс.Дзен. Буквально «на днях» Mail.ru приобрел сервис рекомендаций статей и нативной рекламы relap.io^[94], технологии которого могут быть использованы в развитии «Пульса».

Например, если раньше издатели под публикацией на сайте размещали блок «Почитать еще», в котором предлагались либо материалы той же рубрики, либо материалы с теми же ключевыми словами, то сейчас рекомендательные системы показывают и тексты и рекламу, «похожие» на те, которыми пользователь уже интересовался. «Похожесть» может подразумевать множество атрибутов. И тематику, и длину текста, и стиль изложения, и долю визуального контента и т.д. и т.д.

А что у «больших», что у социальных сетей? Несмотря на неудачи с алгоритмами умных лент, соцсети не планируют делиться с новичками своими пользователями и предпринимают очередные попытки лучше представить новостный и авторский контент непосредственно внутри соцсети. И если от искусственного интеллекта поддержки авторов «Прометей»^[95] (ВКонтакте) давно не было новостей, то Facebook в последнее время разразился инициативами^[96], направленными на привлечение издателей в социальную сеть и создания нового новостного раздела опять же с учетом интересов пользователей.

Вывод тут только один. Все надежды удовлетворить ожидания читателей связаны с персонализацией контента для пользователей и узкоцелевых групп с учетом их интересов. Многократные эксперименты подтверждают гипотезу, что ИИ способен выявлять и интересы, и соответствующий им контент. Уже в ближайшее ИИ, возможно, вернет нам веру в интересные медиа.

ГОТОВЫЕ К ИСПОЛЬЗОВАНИЮ

Компьютерная лингвистика сейчас это не только лингвистика, это все большее применение машинного обучения для автоматической обработки в реальном времени возрастающего потока неструктурированных данных с целью отсеять мусор и «намыть золото». Это противодействие растущему влиянию ботов, выявление коммерческого и иного спама, агрессивных настроений, исключение нерелевантной бизнес-задачам информации. И, наоборот, обнаружение зарожда-

ющихся и умирающих трендов, выявление «живых» людей, их мнений и интересов, текстовый анализ информации на картинках и в видео и т.д. и т.д. И это перечислены только работающие технологии на примере решений Brand Analytics.

Упомянутые выше алгоритмы и решаемые ими задачи без машинного обучения реализовать крайне сложно или невозможно. Один из самых часто задаваемых вопросов любителями потролить на конференциях — это вопрос об автоматическом выявлении сарказма и иронии. И что? В ближайшее время это задачу обещают решить. Тем самым пройден путь от сентимент-анализа на основе словарей, до обученных автоматических классификаторов, которые готовы работать со столь сложными языковыми конструкциями.

А завтра — больше. Больше полезной информации для принятия решений и рекомендаций персонально вам.