|  |  |
| --- | --- |
| Guest Post: Forget big data… here comes machine learning  Guest Post: Forget big data… here comes machine learning  ***Richard Harris, chief executive and founder of Intent Media explains what the future of human-machine interaction means for online travel booking***  The world generates a staggering 2.5 quintillion bytes of data per day. That is written 2,500,000,000,000,000,000. To give you an understanding of how much data that really is, just one character of the words you are reading right now requires 1 byte of data. 2.5 quintillion characters would fill 705 billion King James Bibles, containing more than 780,000 words each. Stacked one on top of another on the ground in front of you, those books would stretch 239,000 miles into the sky, through space, and all the way to the moon…over 400 times.  Meanwhile, back here on Earth, we would carry on generating 3.6 million Google searches, 103 million spam mails, 15 million text messages and 18 million weather forecast requests…every single minute. Is it any wonder that the phrase “big data” is thrown around as much as it is?  **Data doesn’t solve problems…Machine Learning does**  For all the talk about how important this gargantuan volume of data is, I’ve got a dirty little secret to tell you….data is dumb. By itself, data holds no meaning and delivers no value. In order to derive benefit from the information we produce, first we must sift through that data, determining what is relevant and what is not, recognizing patterns, and putting data to work for us through complex predictions of what it all means. And that, undoubtedly, is too difficult for even the brightest of unaided Earthling minds.  Enter the Chinese strategy game Go. “What in the world does untangling quadrillions of bytes of data have to do with a 2,500 year old game of black and white pebbles?” Go, played with 361 moveable stones, is said to have more possible positions than the Universe has atoms. The best players perform on intuition, calculating moves and predicting outcomes on a level too sublime to breach cognitive awareness. Because of this, Go makes for an especially enticing challenge to Machine Learning (ML) researchers hoping to mimic human capabilities. In 2016, Google DeepMind’s AlphaGo system defeated reigning World Go champion, Lee Sedol using a “deep neural network” trained on 30 million moves by human players. By analyzing those moves, AlphaGo independently learned to improve its own choices, effectively becoming the best Go player in the World. What’s amazing about that is that some kinds of Machine Learning are already surpassing human analytical abilities.  But Machine Learning does not show up only in the advanced research labs of data scientists, working on esoteric intellectual problems. Of the tens of billions of dollars that companies invest in Artificial Intelligence each year, about 60% is now focused on ML, propelling real-world solutions into the marketplace.  **Machine Learning isn’t sorcery, but it does work remarkably well**  One area where there’s a huge opportunity to be explored with Machine Learning is the travel booking industry. In 2016, online travel sales worldwide exceeded $550 billion, and are predicted to climb another $100 billion per year through 2019. But of the hundreds of millions of travelers who visit an online travel agent at any given moment, less than 5% of them actually complete a booking.  **What’s going on here? How can we make sense of this contrary data?**  Machine Learning to the rescue. At Intent Media, we trained our ML systems on billions of online travel booking site page views, hoping to distill useful information about how travelers really buy. What we found was that each visitor sends hundreds of individual “signals”— demographic, behavioral and contextual data points generated during an online shopping session. With Machine Learning, travel websites can use those signals to segment different types of travel shoppers and accurately predict their behavior at any point along the transaction path.  Is this visitor ready to make a purchase now? Would an enticing offer at this moment likely result in a purchase? Should we re-target them with advertising? Or is this one simply destined to be ‘the fish who got away?’ Much like the Go playing example we looked at earlier, well-trained ML systems calculate all the possible outcomes simultaneously, in real time, and execute the engagement tactic with the highest chance of driving value out of each visitor. Fed on a steady diet of data, our Machine Learning system seems to know what users intend to do, even before those users themselves. It is truly sublime.  Machine Learning isn’t sorcery, though, and the work of training ML systems is never done. After scoring users we still must A-B test ML-powered predictions against control groups over and over and over, scoring and segmenting users, placing and validating machine predictions, and incorporating the results yet again into our constantly improving machine intelligence model. Why go through such a grueling, iterative process? Simple…because it pays off. Over eight years, the travel sites we support have cashed in on Machine Learning to the tune of billions of dollars of incremental gains.  **The future of machines is the future of humanity**  Despite its graduation from the theoretical to the practical in recent years, however, Machine Learning is still very much in its infancy, and that’s exciting. In the coming decade, ML will become as ubiquitously and invisibly woven into the fabric of all aspects of our lives as it is in industries like travel booking today.  As for those “trips to the moon and back?” Chump change compared with the hockey stick data growth that’s on the launch deck. The volume is so large that Microsoft has been successfully experimenting with a solution to the problem of preserving all that data, using the densest known storage medium in the universe— one that can hold a full quintillion bytes in a single cubic millimeter— DNA. There’s just too much information, coming forth too rapidly to possibly be processed by humans, unaided. Machine Learning is beginning to help define entirely new ways to do live, and to do business. As Jeff Bezos pointed out recently, “We are now solving problems with Machine Learning and Artificial Intelligence that were in the realm of science fiction for the last several decades.” | Гостевой пост: забудьте про big data… встречайте машинное обучение  Guest Post: Forget big data… here comes machine learning  ***Ричард Харрис, директор и основатель Intent Media рассказывает о том, какое будущее человеко-машинного взаимодействия ждет онлайн-бронирование в сфере туризма***  Каждый день в мире создается целых 2,5 квинтиллиона байтов. Пишется это так: 2 500.000.000.000.000.000. Для того, чтобы лучше понять, сколько это: на один символ слова, которое вы сейчас читаете, приходится 1 байт данных. 2,5 квинтиллиона знаков – это 705 миллиардов копий Библии Короля Якова, в каждом экземпляре которой содержится свыше 780 000 слов. Если положить книги друг на друга, то получится башня высотой в 384 633 км, которая устремится ввысь – в небо, через космос, дойдет до Луны… и так 400 раз.  Ну, а здесь, на Земле, мы продолжим генерировать по 3,6 млн Google-поисков, 103 млн спам-сообщений, 15 млн смс и 18 млн запросов о прогнозе погоды… каждую минуту. Стоит ли удивляться тому, что фразой «big data» теперь бросаются на каждом углу?  **Данные не решают проблем… но машинному обучению это под силу**  Говоря о важности всех этих гигантских объемов данных, не могу не поделиться одним маленьким грязным секретом … данные – бесполезны. Сами со себе данные не содержат какого-либо смысла и не несут никакой ценности. Чтобы извлечь выгоду из создаваемой нами информации, необходимо сначала отсортировать данные, разделить их на подходящие и не подходящие, определиться с шаблонами и приспособить данные для работы, используя череду сложных предсказаний всевозможных значений. Несомненно, такая задача непосильна даже для наисветлейших умов землян без использования вспомогательных ресурсов.  Вспомним китайскую игру Го. «Что может быть общего между необъятными квадрильонами байт данных с 2500-летней игрой с черно-белыми камешками?». Считается, что в Го с ее 361 движимым камнем существует больше положений, чем во всей Вселенной – атомов. Лучшие игроки руководствуются интуицией, просчитывая шаги и предугадывая результат на запредельном уровне, не подвластном когнитивному восприятию. Из-за этого Го крайне привлекательная для исследователей машинного обучения (МО), задавшихся целью сымитировать человеческие возможности. В 2016 году система AlphaGo от Google DeepMind одержала победу у действующего чемпиона мира по Го Ли Седоля, используя «глубокую нейронную сеть», обученную на 30 млн ходах человеческих игроков. Анализируя представленные шаги AlphaGo научилась совершенствовать собственный выбор, став лучшим игроком в Го в мире. Самое удивительное здесь то, что в некоторых аспектах машинное обучение уже превосходит аналитические способности человека.  Но машинное обучение присутствует не только в навороченных научных лабораториях у исследователей данных, решающих редкие интеллектуальные проблемы. Из десятков миллиардов долларов, которые компании ежегодно инвестируют в искусственный интеллект, порядка 60% идет на машинное обучение, продвигающее реальные решения на рынке.  **Машинное обучение – это не колдовство, но работает оно на удивление классно**  Одной из областей с огромным потенциалом для машинного обучения является индустрия бронирования в туризме. В 2016 году объем онлайн-продаж туров по всему миру превысил отметку в $550 млрд. Ожидается, что к 2019 году эта цифра дойдет до отметки в $100 млрд в год. Но из сотен миллионов туристов, прибегающих к услугам онлайн-менеджеров по туризму, менее 5% на самом деле завершают бронирование.  **Что тут происходит? Как нам понять столь противоречивые данные?**  На помощь приходит машинное обучение. В Intent Media мы тренировали системы МО на миллиардах просмотров туристических сайтов с онлайн-бронированием, стараясь извлечь полезную информацию о том, как именно туристы покупают что-либо. Мы обнаружили то, что каждый посетитель отправляет сотни отдельных «сигналов»: демографических, поведенческих и контекстных точек данных, которые генерируются во время сессии онлайн-покупки. С машинным обучением туристические сайты могут использовать данные сигналы для сегментации покупателей по различным типам и создания более точного предсказания их поведения в течение всей операции.  Готов ли этот посетитель совершить покупку именно сейчас? Может ли заманчивое предложение в этот конкретный момент времени довести его до осуществления покупки? Стоит ли перенаправить его рекламным сообщением? Или же нам попалась категория посетителей, про которых говорят, «что рыба сорвалась с крючка»? Также, как и в рассмотренном ранее примере игры Го, хорошо натренированные системы МО просчитывают все возможные варианты одновременно, в режиме реального времени, и реализуют тактику взаимодействия, которая с максимальной вероятностью принесет ценность для каждого посетителя. Взращенная на полноценной диете из отобранных данных, наша система машинного обучения узнает о намерениях пользователя быстрее, чем он сам. Вот это действительно грандиозно.  Но все-таки, машинное обучение – это не колдовство, поэтому работа по научению систем МО не закончится никогда. После количественного определения пользователей, мы все равно должны вновь и вновь возвращаться к А/В тестированию результатов предсказаний моделей МО и контрольных групп: подсчитывать и сегментировать пользователей, размещать и проверять машинные предсказания, добавлять полученные результаты в нашу постоянно совершенствующуюся систему искусственного интеллекта. Зачем нам проходить через столь архисложный и итеративный процесс? Все просто… это окупается. За восемь лет работы поддерживаемые нами туристические сайты монетизировали на машинном обучении миллиарды долларов дополнительной прибыли.  **Будущее машин – это будущее человечества**  Несмотря на, что в последние годы отмечался плавный переход машинного обучения от теории к практике, оно все еще находится в зачаточном состоянии. И это здорово. В течение следующих десяти лет МО станет таким же незаметным и полномасштабным спутником нашей жизни, как это уже произошло в сфере туристического бронирования.  Ну а что насчет тех «полетов до Луны и обратно»? Прирост данных здесь сравним с эффектом хоккейной клюшки в стадии активного взлета. Объем информации настолько велик, что Microsoft успешно экспериментирует с решением проблемы сохранения всех этих данных в самой плотной из ныне существующих сред хранения – той, которая сможет удерживать целый квинтильон данных в одном кубическом миллиметре, – ДНК. Колоссальные объемы информации появляются настолько быстро, что их не сможет обработать ни один человеческий ум без посторонней помощи. Машинное обучение готово помочь в создании новых способов ведения бизнеса и жизни. Как недавно сказал Джефф Безос: «Сейчас с помощью машинного обучения и искусственного интеллекта мы можем решать проблемы, которые не один десяток лет относились к разряду фантастики». |

http://travolution.com/articles/106685/guest-post-forget-big-data-here-comes-machine-learning