

Компьютерная аналитика

Семенов Ю.А. (ИТЭФ-МФТИ)

Аналитика существует давно и имеет целью выявление тенденций и прогнозирование. Часто служит для подтверждения уже выбранного курса или принятого решения. Компьютерная аналитика (когнитивный компьютеринг) имеет целью выявление неизвестных тенденций, закономерностей и корреляций. Современная компьютерная аналитика способна воспринимать задания на естественном языке (например, русском). Эта технология позволяет получить объективный прогноз и выявить трудно предсказуемые последствия тех или иных решений. Исполнительный директор компании IBM Ginni Rometty 2-го марта 2012 года сказала:

"В ближайшие пять лет все компании на рынке разделятся на победителей и побежденных в зависимости от качества их аналитики. Опыт и интуиция не помогут. Людей надо переучивать".

Пять лет уже прошли...

На данном этапе востребована аналитика для бизнеса, которая позволяет прогнозировать динамику спроса, производства и изменение цен. Но в перспективе, в частности по мере развития Интернета вещей, можно ожидать использования технологии для выявления неизвестных связей и зависимостей. Компания IBM считает возможным использование аналитики для выявления неизвестных зависимостей и в сфере науки (путем анализа побочных результатов измерений).

В 2015 году объем накопленных человечеством данных достиг 7 зеттабайт (1 зеттабайт = 10^{21} байт эквивалентен объему 50 библиотек конгресса США). Заметная доля этих данных доступна через Интернет (по большей части в виде текстов). Но до недавнего времени люди не могли извлекать пользу из этой информации.

22-го сентября 2011 г. появилась заметка компании IBM о компьютере **Watson** (см. Talking computer will change the world; promises not to kill you). Разработчики (во главе с David'ом Ferrucci) позиционируют Watson, как компьютерную систему искусственного интеллекта. Watson может обрабатывать до миллиона книг в сек. (500 гигабайт). Этот компьютер содержит в себе 90 серверов Power7 750, каждый из которых имеет по 4 восьмиядерных процессора Power7. Оперативная память этой машины содержит 15 Тбайт.

Данный проект можно считать шагом на пути создания искусственного интеллекта к 2023 году (см. <http://book.itep.ru/4/7/resources.htm#20>) Трудно утверждать это точно, но возможно, что этот проект является этапом создания искусственного интеллекта для безлюдного оружия поля боя. Компания IBM в прошлом году открыла Watson для широкого международного сотрудничества.

Хотя компьютер Watson лишь с некоторой натяжкой может быть отнесен к суперкомпьютерам (143-е место в списке Top500 на начало июня 2013), он занимает особую позицию. Это первый компьютер, который может работать с текстами на естественном языке, и в нем реализован важный шаг в обучении машины пониманию контекста и самого задания. Что важно, машина способна пользоваться данными из Интернет для решения стоящих перед ней задач.

Компания IBM пытается коммерциализовать использование вычислительной машины Watson (победитель телевизионного шоу Jeopardy). Среди возможных применений рассматривается сложная диагностика различных заболеваний, в частности раковых. С точки зрения программирования эта машина не является детерминистской, она управляется запросами, анализирует накопленную информацию и возвращает данные, которые с ее точки зрения являются наиболее релевантными полученному запросу. Машина может также использоваться как советчик для врача .

Компания IBM предлагает свой суперкомпьютер Watson в качестве платформы для разработки облачных приложений, ориентированных на когнитивный компьютинг (см. "IBM to offer Watson supercomputer as cloud development platform", Chris Kanaracus, IDG News Service, November 14, 2013). Объемы информации во всех областях быстро растут.

Различные методы экстраполяции и анализа корреляций использовались для прогнозирования и ранее. Компьютерная аналитика предлагает возможности обработки больших объемов данных, например, доступных в Интернет.

Следует, впрочем, всегда помнить, что качество аналитического прогноза полностью определяется качеством и надежностью исходных данных. Информация из Интернет обычно не обладает высоким качеством, но самообучаемость аналитических программ (deep learning) облегчает решение этой проблемы.

Технология глубокого обучения позволила создать программу, которая обыграла чемпиона мира по го, что считалось невозможным при использовании традиционных методов.

Компания IBM усовершенствовала программу анализа данных Watson Discovery Advisor так, что она стала способна ответить на ваши вопросы, до того, когда вы их задали. Считается, что эта программа будет полезной помимо медицины в финансах и юридической деятельности. Можно также считать, что эта система не запрограммирована, а **обучена**.

Информация о здоровье, которую собирает Apple Watch, может передаваться в облако на базе IBM's Watson. Медицинские данные очень неоднородны, большая часть информации имеет текстовый вид. Разрабатываются программы аналитики для реализации индивидуальных планов лечения.

Вслед за Watson появились машины Siri, Cortana и пр. того же класса и назначения.

Если в 2013 году на пике популярности были научные аналитики, в 2014 центр сместился в область информационной бизнес-аналитики (см. "Big Data Analytics predictions for 2014", Michele Chambers, Greg Todd, David Smith, December 31, 2013). За счет программных средств прогнозирование будет осуществляться в реальном масштабе времени. Научным аналитикам останутся стратегические инициативы. В 2014 году предсказательная аналитика начала работать в рамках Hadoop. В 2014 маркетинговая аналитика на основе big data стала реальностью. Смотри также <http://blog.revolutionanalytics.com/2013/12/14-analytics-predictions-for-2014.html>.

Директор подразделения по цифровой преступности компании Microsoft Брайан Херт (Bryan Hurd) утверждает, что аналитика big data начинает играть все большую роль в борьбе с киберпреступностью. Аналитика позволяет выявлять активность, сопряженную с использованием активационных ключей украденных

программ. Компьютерный анализ сетевого трафика способен предсказать будущие кибер преступления.

Торговая компания Macy объявила, что она предлагает покупателям новый сервис - возможность попросить помощи у системы Watson. Система может работать в облаке с английским и испанским языками. Клиент может задавать вопросы на естественном языке (голосовой канал пока отсутствует). Система использует технологию машинного обучения.

В LANL (Lawrence Livermore National Laboratory) создан суперкомпьютер (Catalyst), предназначенный для переработки больших объемов данных. Этот комплекс планируется использовать для бизнес аналитики и для исследования свойств сверхтяжелых элементов.

Многие задачи, например, бизнес не требуют обработки гигантских объемов данных и для таких целей IBM разработала машины размером с коробку для пиццы.

Исследовательская компания Gartner выделяет 10 технологических тенденций в 2015 году (см. "Cloud, web-scale IT and IoT in Gartner's 2015 top technologies", Archana Venkatraman, Thursday 09 October 2014). Среди них облачные технологии, архитектуры SDN, 3D-печать, Интернет вещей, информационная аналитика и WEB-технологии.

Компании IBM и Juniper Networks решили объединить возможности своего оборудования и программного обеспечения для извлечения аналитических данные из информации проходящей через сетевые каналы. Речь идет о данных, поступающих от персональных компьютеров и мобильных устройств. Технология IBM будет встроена в шлюзы Juniper, в то время как более глубокая аналитика будет реализована на отдельных серверах.

В новую версию Microsoft Office встроено средство (BeyondCore Apps) для реализации аналитики. Это аналитическое средство могут использовать не только эксперты, но практически все пользователи Office. Для работы с этим продуктом достаточно навыков применения Office. Ожидается, что это средство будет и дальше совершенствоваться в течение ближайших 5-10 лет.

Аналитика больших объемов данных находит все более широкое приложение. Ниже приведены доли этой технологии в программных разработках разных отраслей ИТ: (см. "How, where and why big data analytics are delivering value", David Loshin, June 2015).

- 31% в сфере торговли
- 30% в мобильных приложениях
- 30% при интеграции данных
- 26% в автоматизации или управлении бизнес-процессами
- в анализе больших объемов данных

Современные системы аналитики могут предсказать кражу данных путем сканирования сообщений e-mail. Токийская фирма UBIC создала программу, которая анализирует тексты почтовых сообщений и выявляет потенциальные угрозы кражи данных

Исследовательская лаборатория армии США (ARL, министерство обороны США) планирует использовать в своих разработках в 2015-30г суперкомпьютер с вычислительной мощностью 100 петафлопс (см. "U.S. Army plans for a 100 petaflop supercomputer", Andy Patrizio, February 25, 2015). Компьютер будет применен и для информационной аналитики. В будущем это направление разработок будет являться стратегическим.

В России проблемами компьютерной аналитики занимается компания Yandex (по некоторым аспектам и лаб. Касперского), но мне представляется целесообразным создание распределенного компьютерного аналитического центра, где смогут работать десятки или даже сотни программистов.

Компании Cisco и IBM сотрудничают в сфере аналитики для Интернета вещей. Планируется использовать инфраструктуру Cisco для аналитики, доступной на платформе IBM Watson. Эта технология будет, например, применена для мониторинга здоровья шахтеров и рабочих характеристик критического оборудования, позволяя поднять производительность и эффективность работ. Планируется внедрение технологии в Объединенных Арабских Эмиратах, а также 136 других странах. Российская авиакомпания Aeroflot приобрела пакет аналитических программ Hadoop-based BigInsights у компании IBM. Эти программы будут использоваться для выявления тенденций и склонностей клиентов, а также для определения наиболее перспективных направлений развития сети маршрутов. Для обработки используются данные из более чем 300 источников.

Ученые компании IBM создали искусственные нейроны и синапсы, используя фазопеременную память. Впервые исследователи смогли создать случайно возбуждаемые нейроны (randomly spiking neurons). Эта технология будет применена для приложений когнитивного компьютеринга. Эксперты IBM считают, что реальный чип процессора, использующий эту технологию, появится только через несколько лет. Искусственный нейрон имеет размер 90нм, но в перспективе может быть уменьшен до 14 нм. Система уже сегодня демонстрирует высокую эффективность выявления корреляционных зависимостей. Понятно, что компьютерная аналитика, чем дальше, тем больше будет базироваться на технологиях искусственного интеллекта.

Считается, что компьютерная аналитика будет эффективной для выявления уязвимостей прикладных программ и детектирования сетевых атак.

Компьютерная аналитика может стать еще одним источником знаний об окружающем мире наряду с экспериментом, теорией и моделированием.