



Тайная сила данных

Машине достаточно всего 70 «лайков», чтобы знать о человеке больше, чем его друзья. По всей видимости, **компания, занимающаяся большими данными,** именно при помощи этих знаний привела Трампа в Белый дом

Судовольствием ходишь в бассейн, любишь косметику MAC и иногда лакомишься спиральной картошкой фри? Тогда ты жизнелюбивый интеллектual с гомосексуальными наклонностями. Увлекаешься Куртом Кобейном, Adidas Originals и биологией? Тогда ты гетеросексуальная невротичка и у тебя множество друзей.

Если, прочитав это, вы сейчас неодобрительно качаете головой, считая это предубеждениями, то ошибаетесь. Потому как не предубеждения это, а утрированные, следует признать, примеры корреляций — связей между признаками, свойствами и состояниями. Этот термин получает широкое распространение вместе с развитием феномена Big Data. Объем данных по всему миру каждые два года увеличивается почти в два раза, как и производительность компьютеров, поэтому, используя машину в анализе данных, можно устанавливать взаимосвязь практически всего, что нас окружает, и составлять прогнозы. Если, к примеру, компания, которая занимается выпуском платежных карт, постоянно проводит оценку миллиардов, если не триллионов, записей о покупках, совершенных ее клиента-

ми, то машинный анализ может выдать почти сразу поразительные предсказания такого рода: браку этой пары грозит серьезный кризис, вот этот мужчина чувствует себя не в своей тарелке, а та женщина в течение трех месяцев тяжело заболит.

Продажа цифровых следов

Данные пользователей платежных карт всегда были чрезвычайно содержательны, а мощные компьютеры и сложные алгоритмы позволяют сборщикам данных узнавать еще больше. Сегодня же, во времена смартфонов и планшетов, датчиков и камер, цифровые следы остаются повсюду. Их можно купить, и более того — нередко они оказываются в свободном доступе. Их ценность заключается в возможности выявить связи. Если составить правильную комбинацию цифровых следов, то человек, оставивший их, становится раскрытой книгой, которую специалисты по анализу данных могут не только прочитать, но даже переписать некоторые главы, навязав чужие потребности и мнения. Довольно много людей считает, что именно таким образом Дональд Трамп победил на президентских выборах в США.

ФОТО: компания-производитель Carlos Barria/Reuters, Tobias Hase/dpa/Global Look Press, Carlo Allegri/Reuters, Lauren Bamford/michalkosinski.com

Научный сотрудник Стэнфордского университета Михал Косински объединил большие данные с психометрией, которая изучает теорию и методику психологических измерений. С 2010 по 2014 годы он работал заместителем руководителя Психометрического центра Кембриджского университета. Изучением корреляций между цифровыми следами и профилем личности он занимается с 2007 года, когда поделился со своими 150 друзьями на Facebook психологическими опросниками (<https://discovermyprofile.com>), которые были рекомендованы множеством раз, а автор получил информацию по тысячам людей. Людям просто доставляло удовольствие отвечать на вопросы, которые напоминали психологические тесты в женских журналах — неслучайно две трети участников оказались женщинами. Четыре года спустя количество участников опроса достигло шести миллионов, а на сегодняшний день их насчитывается около восьми миллионов. Тестирование опирается на теорию так называемой пятифакторной модели личности, которая вот уже около тридцати лет является стандартом в современной психологии. Согласно этой теории, личность человека включает в себя пять общих черт, так называемую «большую пятерку»: открытость опыту, добросовестность, экстраверсия, доброжелательность и невротизм, — которые проявляются в различных формах. Первые буквы названий этих черт на английском языке составляют акроним OCEAN, поэтому такое выведение модели еще называют «метод океана».

Скажи нам, кто твой «лайк»

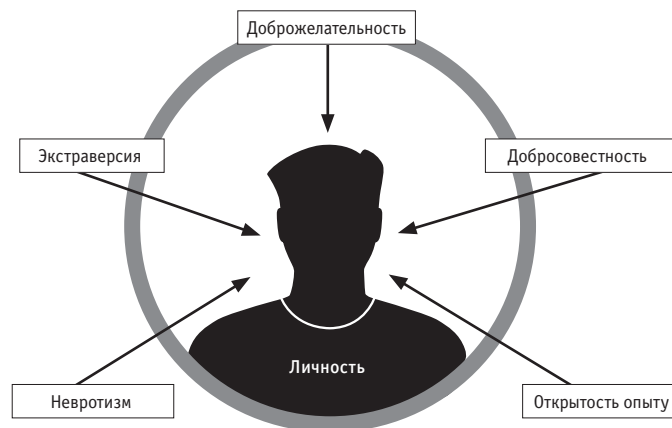
Таким образом психолог-философ Косински в течение короткого промежутка времени получил сотни тысяч психогрaмм. Подготовив выборку из психогрaмм 58 000 испытуемых из США, он начал искать корреляции между результатами тестирования и их профилями в Facebook. Кроме «большой пятерки», для изучения выборки у испытуемых запрашивались или выявлялись такие признаки, как возраст, пол, сексуальная ориентация, политические предпочтения, опыт употребления наркотиков, интеллект и, что самое ценное, действия в Facebook в виде отметок «Нравится», к которым группа исследователей во главе с Косински получила доступ с согласия участников опроса. Связи между «большой пятеркой» и «лайками» оказались чрезвычайно тесными. Располагая сведениями о «лайках» испытуемого, можно было сделать однозначные выводы о чертах его характера и других признаках, которые тем или иным образом фигурировали в опросе, и наоборот: по определенным признакам можно было судить об отметках «Нравится», что было выгодно для целевой рекламной кампании. Десяти «лайков» было достаточно, чтобы узнать личность лучше, чем обычно коллеги по работе. 70 — лучше, чем друзья, 150 — лучше, чем близкие родственники. А если удавалось проанализировать больше 300 «лайков», можно было узнать испытуемого лучше, чем его спутник жизни.

Конечно, чем больше значений охватывает анализ действий в Facebook, тем более точны выводы. Но бывают и отдельные особенно содержательные «лайки», корреляции которых бывает сложно проследить с точки зрения причины — такой типичный феномен больших данных.

Это возвращает нас к началу статьи. Если вы, например, отметили понравившейся публикацию о плавании, то, скорее всего, вы очень довольны жизнью. Если любите спиральный картофель фри, то вероятность того, что вы эрудированны, выше среднего. Это не значит, что можно стать умнее, если нажимать «Нравится» под публикацией на тему картофеля фри в виде спиралек. И даже если их есть. Но если вы не только «лайкаете» публикации о плавании, но еще и регулярно посе-

Большая пятерка

Согласно пятифакторной модели личности человека, также называемой «большая пятерка», личность человека может быть изменена исходя из выявления пяти значений. Характер определяется степенью выраженности факторов.



Михал Косински разработал метод, при помощи которого можно с удивительной точностью составить профиль личности человека на основании отметок «Нравится» в Facebook

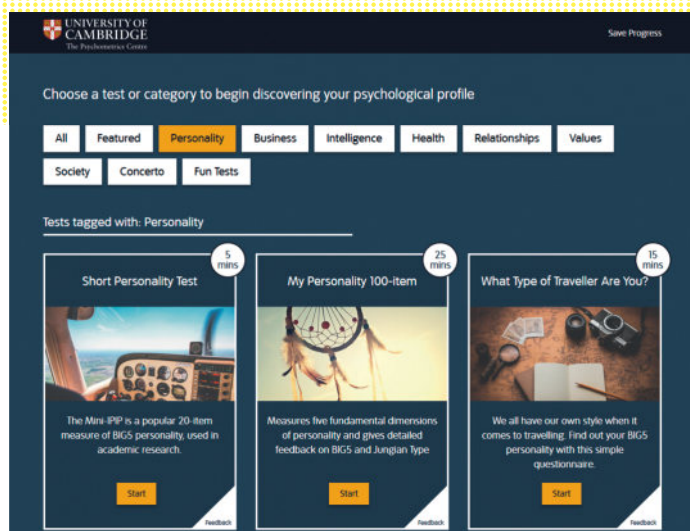


щаете бассейн, то действительно вполне вероятно, что жизнью вы все-таки довольны.

Михал Косински — один из ученых, предупреждающих о том, что данные такого рода могут нести опасность, если попадут не в те руки. В этом отношении его теория об аспектах больших данных на выборах, на которых победил Трамп, кажется правдоподобной: Косински считает, что некая британская компания украла у него «модель океана», которую он разработал для Facebook, и именно с ее помощью Дональд Трамп удалось достучаться до сознания избирателей. Эта зловещая компания, основанная в 2013 году, носит название Cambridge Analytica (CA) — намек на то, что она нанимала сотрудников из Кембриджского университета.

Спусковой механизм политического землетрясения

CHiP встретил руководителя отдела анализа и обработки данных CA Александра Тайлера упреками. По его словам, в его компании об «этом исследователе Косински» впервые услышали в декабре 2016 года, когда швейцарский журнал Das Magazin опубликовал его возмущения. Позже Тайлер добавит: →



Самотестирование

На discovermyprofile.com можно пройти психологическое тестирование на основании OCEAN, в том числе подключив профиль Facebook



Закулисный махинатор

Советник Трампа по стратегическим вопросам Стивен Бэннон, предводитель «альт-правых», является членом совета директоров Cambridge Analytica

«Мы контактировали с некоторыми сотрудниками университета с его факультета, но не напрямую с ним». То, что Косински незнаком СА, несмотря на то, что он работал заместителем руководителя небольшого института, услугами которого компания пользовалась с точки зрения не только кадровой, но и, возможно, содержательной, не для всех звучит убедительно. Но так или иначе, СА часто не проявляет своей позиции, слухи официально не опровергаются, чем подпитывают домыслы. Пока еще неясно, какое влияние оказала кампания СА по Brexit на исход выборов, но в любом случае складывается впечатление, что СА получает удовольствие от того, что именно на нее возлагается ответственность за политическое землетрясение такого класса, как референдум по выходу Великобритании из ЕС и его результаты или победа Трампа.

Сомнения вызывают и слова Тайлера о том, что СА не использует данные Facebook: «Это собственность компании Facebook». Бесспорно, но все же вопрос заключается в том, не использует ли СА данные пользователя — и даже его ни о чем не подозревающего друга на Facebook, когда он открывает свой приватный профиль для прохождения опроса. В 2015 году газета The Guardian тоже попыталась расспросить СА, предъявив результаты исследований, но ответов не добились. Во вся-

ком случае, СА ссылается на наличие различных источников: «Данные, с которыми мы работаем, варьируются в зависимости от страны. В России персональные данные можно собирать только с предварительного заявления о согласии и в определенных целях. А в Америке это происходит по-другому: мы можем лицензировать данные у крупных агрегаторов — Asxiom, Experian или Aristotle». Эти компании собирают персональные данные всех видов, которые могут быть связаны, например, с программами лояльности, с компанией, выпускающей платежные карты, или с дистрибуцией журналов. Еще по оценкам 2013 года компания Asxiom владела подробной информацией о более чем 800 млн человек.

Европейские положения о защите персональных данных могут оказаться полезными в борьбе против действий таких компаний, как СА или Aristotle, которая тоже уже имеет опыт участия в политических кампаниях. Но если во время оформления заказа, прохождения опроса или розыгрыша пользователь нажимает кнопку «Согласен на использование моих персональных данных в маркетинговых целях», то следует ожидать, что им будут манипулировать с помощью рекламных рассылок или даже политической пропаганды.

Психологические трюки и победа на выборах

Без сомнения, СА поспособствовала предвыборной кампании Дональда Трампа психометрическими методами и цифровыми инструментами. Для этого компания использовала «метод океана» и «смоделировала психологический профиль каждого совершеннолетнего американского потребителя», как сказал Тайлер в беседе с CHIP. Что именно это значит? «У нас есть психологические портреты сотен тысяч человек, которые мы соединили с нашим базовым набором данных». Эти базовые данные охватывали все 220 млн совершеннолетних граждан США, которые собирались такими крупными компаниями, занимающимися Data Mining, как агрегаторы, указанные выше, и лицензировались СА для дальнейшего использования. Таким образом были получены данные и информация об образе жизни и потребительском поведении населения. Сюда же вошли и результаты собственных исследований. Затем при помощи машинного обучения были вычислены личные ценности каждого гражданина.

Предвыборная кампания, проводимая СА для Трампа, оказалась невероятно действенной, сконцентрировавшись всего на 17 штатах, где исход выборов было сложно предсказать. И поскольку СА знала не только о политических предпочтениях и о многих других симпатиях граждан, но и об их психологической сути, то могла с удивительной точностью определить, с какими посланиями к каким людям обращаться. Специальное приложение подсказывало волонтерам, проводящим агитацию в пользу Трампа, личностные типы жителей того или иного дома, чтобы они могли направить свой разговор с жителями в нужное русло. Обратную реакцию волонтеры записывали в то же приложение и отправляли прямо в СА в целях оптимизации системы.

Президент США как орудие Big Data

Для распространения посланий Трампа его команда задействовала прежде всего Facebook. Использовались и так называемые «темные публикации», когда некоторые объявления попадали только в ленту определенных групп лиц. Такой микротаргетинг с участием сомнительных новостей отчасти даже сомнительной достоверности определяется совсем не тем, за что выступает кандидат, а тем, что хочет услышать потенциальный избиратель. Послание Трампа было не одно — их было много тысяч, и они были нацелены на совершенно разных лю-

дей. Трамп стал орудием Big Data. Руководитель отдела анализа и обработки данных СА рассматривает политическую рекламу не иначе как кампанию по продвижению продукта. «Вопрос заключается в повышении эффективности — это действительно занимает центральное место. Нужно донести рекламу нужного содержания до нужного человека и таким образом выразить послание, чтобы человек на него отреагировал».

Как только вы попадаете в Интернет, вы тут же оказываетесь под прицелом рекламщиков. Это сильно раздражает, но можно подойти к рекламе с другой стороны, рассматривая ее как сделку, как раз когда речь идет о бесплатных сервисах. Большинство пользователей соглашаются на нее, потому что сценарии Оруэлла и воздействие рекламы на потребителей — все же две разные вещи. Пока она не становится на темную сторону, позволяя использовать данные в политических манипуляциях, стигматизации и прочих видах ущемления интересов, мало что может пойти не так. По этому девизу в цифровой вселенной до сих пор можно было довольно сносно жить. Но победа Трампа дала понять, что вывод этот ложный.

Даже если вклад СА в триумфе светловолосого строительного магната переоценен, все же следует признать, что новое правительство США как минимум косвенным образом имеет доступ к базе данных с психограммами всех граждан, куда относится, помимо всего прочего, и такая конфиденциальная информация, как сексуальная ориентация, а также политические и религиозные предпочтения.

«Альтернативные правые» на заднем плане

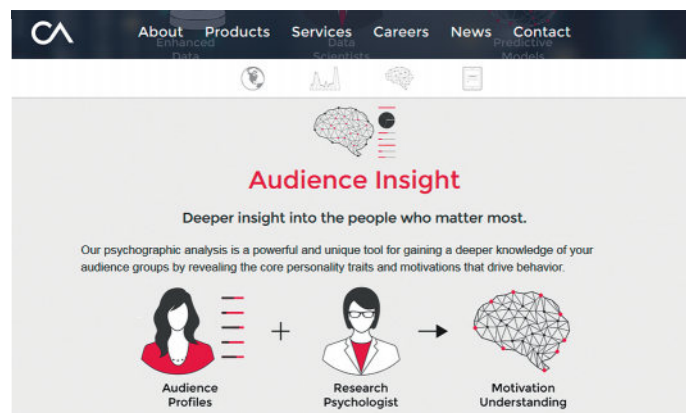
Стивен Бэннон, главный представитель ультраправого движения в США, так называемых «альт-правых», не раз обвинялся в антисемитизме и расизме. Он возглавлял избирательный штаб Трампа, теперь же занимает пост главного советника президента по стратегическим вопросам и даже является членом постоянного состава Совета национальной безопасности США. Но это еще не все. Именно ему приписывается важная функция контроля в совете директоров СА. На наш запрос подтверждения от СА не последовало, опровержения — тоже.

Бэннону приписываются такие идеи, как создание базы данных мусульман и разработка коммуникационной стратегии для президента с использованием больших данных. Кто бы мог воплотить их в жизнь лучше, чем «его» компания СА? Вопрос не совсем риторический: компания Palantir, основным инвестором которой является ЦРУ, также занимается анализом данных в политических целях. Прежде всего она связана со спецслужбами: по слухам, ее предсказательная аналитика помогла США найти, где скрывается сам Усама бен Ладен. Питер Тиль, крупный акционер Palantir немецкого происхождения, сыграл важную роль для переходной команды Трампа и не в последнюю очередь благодаря многомиллионным вложениям в предвыборную кампанию имеет связи с ним.

Тиль является оппонентом Бэннона не только из-за деловых интересов компании Palantir. Этот миллиардер в свое время стал первым внешним инвестором Facebook, получив членство в совете директоров компании. Мог ли глава Facebook Марк Цукерберг через Тиль сдружиться с Трампом — это отдельный вопрос. Но Цукерберг определенно не хотел бы смотреть на то, как СА выполняет правительственные поручения, вызывающие вопросы, сомнительным образом получая данные Facebook или даже, может быть, выкладывая в систему фейковые новости по задумке Стивена Бэннона, играющего двойную роль — представителя СА и советника Трампа. И независимо от того, кто возьмет верх, Трамп остается под влиянием представителей двух компаний,

Big Data и Big Five

Используя большие данные и алгоритмы психологии, компания Cambridge Analytica составила психограммы американских избирателей, благодаря чему Дональд Трамп попал в Белый дом.



Cambridge Analytica характеризует свой сервис как «мощный универсальный инструмент», который выявляет «мотивы поведения»

«Мы смоделировали психологический профиль каждого потребителя США»

Александр Тайлер,

руководитель отдела анализа и обработки данных Cambridge Analytica



которые готовы использовать конфиденциальную информацию населения в политических целях.

Будущее предвыборных кампаний

А что же европейские страны? Возможен ли такой микротаргетинг во время избирательных кампаний по другую сторону Атлантики? В принципе, возможен, считает представитель СА Александр Тайлер: «Нам пришлось бы использовать не те инструменты, что были в США. Но остальное, пожалуй, в целом мало чем отличалось бы».

Впрочем, следующий президент США вполне может обойтись и без помощи таких компаний, как Cambridge Analytica. Крупнейшая социальная сеть Facebook хранит на своих серверах персональные данные более чем миллиарда пользователей, которые сами охотно делятся любой частной информацией. Кроме того, Facebook сотрудничает с Anxiom и другими компаниями, собирающими и обрабатывающими данные. По меньшей мере с точки зрения жителей Америки эти данные принадлежат не пользователям, а социальной сети Facebook, и если у Марка Цукерберга вдруг появится такое желание, он даже сможет обеспечить себе пост президента США на выборах в 2020 году. 