

28 января 2014

Сайты знакомств — развлечение не для всех, но сложно отрицать: многим людям они изменили жизнь к лучшему, дав новый способ заводить друзей и искать партнёров. Эти сайты к тому же — чрезвычайно интересная площадка для исследований, где социологические проблемы нередко удаётся решить при помощи построения математических моделей и анализа данных.

Многие сайты знакомств используют алгоритмы, позволяющие подбирать и рекомендовать потенциальных партнёров на основе общих вкусов, симпатий и антипатий, интересов, хобби и так далее. Сайт с лучшими алгоритмами будет эффективнее (вспомните недавнюю [заметку о том, как большие данные помогают сервису eHarmony создавать счастливые брачные пары](#)) — а значит, привлечёт больше пользователей и заработает больше денег. Рынок онлайн-знакомств оценивается примерно в \$3 млрд в год.

Издание Technology Review [сообщает](#) о последних новостях в этой области. Группа исследователей из Айовского университета, которой руководит Кан Чжао, [разработала](#) специализированный рекомендательный движок. Используя анализ больших массивов данных, он потенциально может вывести эффективность сайтов знакомств на новый уровень. Новая модель не только учитывает вкусы пользователей, но и измеряет их привлекательность для потенциальных партнёров. Система, разработанная Чжао и его командой, предлагает пользователям таких визави, которые с наибольшей вероятностью дадут ответ на сообщение, если тот начнёт диалог.

Широко известно, как работают рекомендательные движки Amazon и Netflix: на основании знаний о том, какие товары выбрали те или иные пользователи, другим пользователям, покупающим схожие вещи, предлагают купить ещё что-то, чего у них нет. По аналогичному принципу движок сайта знакомств анализирует, каких собеседников человек выбирает для переписки, затем находит других пользователей с похожими вкусами и рекомендует потенциальных партнёров, с которыми те уже познакомились, а пользователь, нуждающийся в рекомендации, — ещё нет. Другими словами, получают рекомендации типа «Мальчикам, которым понравилась та же девочка, понравились также эти девочки» и «Девочкам, которым понравился этот мальчик, также понравились эти мальчики».

У этого подхода есть один серьёзный недостаток: он не принимает в расчёт

привлекательность самого пользователя. Если люди, которым он пишет, никогда не отвечают, то никакие рекомендации не помогут. Поэтому Кан Чжао и компания решили добавить ещё один фактор. Их рекомендательный движок анализирует также и получаемые ответы и по ним оценивает привлекательность (или непривлекательность).

Очевидно, что пользователи, получающие больше ответов, более привлекательны. Учитывая этот фактор, сайт может порекомендовать потенциальных партнёров, которые не только соответствуют вкусу, но и, скорее всего, найдут собеседника привлекательным и пойдут на контакт.

Результаты работы алгоритма сравнили с тем, как работают другие методы рекомендаций — например, только по интересам или по другим переменным. Каждый алгоритм применили к уже собранным данным о том, как переписываются пользователи, чтобы узнать, будут ли рекомендации соответствовать реальному положению дел. Для экспериментов использовался анонимизированный набор данных, полученный с сайтов знакомств и содержащий переписку 47 тысяч пользователей на протяжении 196 дней.

Кан Чжао сообщает, что результаты тестов однозначно указывают на превосходство нового метода, хотя количественные показатели не приводятся. «Если пользователь знакомится с кем-то, кого посоветовал наш движок, то шансы на ответ будут выше», — пишет он.

Правда, одна проблема так и осталась неразрешённой. Алгоритм Чжао бесполезен для новичков. Он нуждается в исходных данных, а для того, чтобы их накопить, пользователь должен завязать достаточное количество разговоров (и, возможно, оказаться отвергнутым достаточное число раз). Между тем именно такие люди больше всего нуждаются в качественных рекомендациях.

Как говорит Кан Чжао, эксперимент с сервисом знакомств был интересен и с академической точки зрения. Большие наборы данных, описывающих поведение людей, дают возможность протестировать новые методики и алгоритмы. В теории динамический анализ данных может стать ключом к новым исследованиям в области социологии.

Что до применения нового алгоритма на настоящих сайтах знакомств, то тут Чжао ничего конкретного не сообщает. Впрочем, исследования столь практической направленности не должны долго оставаться невостребованными.

Источник - <http://www.computerra.ru/92805/dating-bigdata/>