

О ПРАКТИЧЕСКОМ ПРИМЕНЕНИИ ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ

Б.А. Низомутдинов, А.С. Тронников, А.Б. Углова

*Университет ИТМО, Российский государственный педагогический университет им. А. И. Герцена.
Санкт-Петербург*

За последние годы, все большая часть социальных интеракций переносится человечеством на цифровые платформы. Благодаря интенсивному росту аудитории социальных сетей [1] и интернациональной политики предоставления доступного интернета [2], изменилась сама основа социальных наук: теперь вместо проведения экспериментальных исследований в лабораториях или выборочных тестирований на тысяче испытуемых, исследователи могут наблюдать и изучать поведение десятков миллионов людей [3]. Итоги подобных исследований могут предложить беспрецедентные по своим масштабам результаты, способные изменить наше понимание социума и его поведения [4].

Проведение подобных масштабных исследований стало возможным благодаря «цифровым отпечаткам» - любой информации, оставляемой пользователем при использовании сети Интернет. Такие «отпечатки» могут включать в себя множество данных: от IP-адресов и истории поисковых запросов, до статистики нажатых кнопок на сайте и количестве друзей в социальных сетях [5].

Используя такую информацию, исследователи научились создавать различные прогностические модели [6], которые, в свою очередь, используются в персонализированных поисковых системах [7], таргетинговой рекламе [8], системах оценки персональных рисков [9] и т.п. Однако, проведение подобных исследований и применение прогностических моделей ведет к противоречивым вопросам в области информационной безопасности и личного пространства.

Практика мирового законодательства, в принципе, как и законодательства Российской Федерации ещё не подготовлена для регулируемого процесса сбора массовых открытых данных пользователей. Вступивший в силу 25 мая 2018 года Общий регламент по защите данных (GDPR) был принят Европейским Союзом как один из основных и первых документов, регламентирующих права пользователя в активно развивающейся среде сбора общедоступных данных.

Благодаря данному постановлению, граждане ЕС (и не только) получили право на полное и всесторонне знание о том, какие данные собираются, предоставляются и обрабатываются как основными Интернет сервисами, так и третьими лицами, выступающими в роле посредника при обработке подобной информации. Кроме того, одним из основных пунктов данного регламента является предоставления пользователям

«Право на забвение» («right to be forgotten»). Данное право гарантирует человеку возможность запросить полное удаление каких-либо пользовательских данных из систем хранения как первоначального сервиса, так и из систем хранения третьих лиц, которым осуществлялась передача подобного рода информации.

Данное постановление также регулирует множество мелких вопросов, связанных с информационной безопасностью, способами хранения персональных данных, использование данных несовершеннолетних, предоставление информации в конфиденциальной политики сервиса и т.п. Однако, процесс сбора общедоступных данных, которые пользователем самостоятельно и сознательно выкладывает в Интернет, остается законным.

В законодательстве РФ есть несколько документов, регламентирующих работу с персональными данными:

- Конвенция о защите физических лиц при автоматизированной обработке персональных данных (Страсбург, 28.01.1981, ратифицирована Федеральным законом от 19.12.2005 года N 160-ФЗ);
- Конституция РФ. Статьи 23 и 24;
- Федеральный закон «О персональных данных» от 27.07.2006 N 152-ФЗ;
- Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 N 149-ФЗ.

Данные нормативные акты регламентируют отношения между оператором персональных данных (лицом, совершающим сбор, выгрузку, обработку, хранение, систематизацию, передачу данных) и человеком, предоставляющим своим персональные данные.

При этом следует отметить, что, согласно статье 22 закона «О персональных данных», исключением является сбор общедоступных данных.

Однако, поскольку данные законодательные документы были составлены ещё до массового применения технологий больших данных при массовом сборе общедоступных данных, то на текущий момент возникают судебные тяжбы между социальными сетями и компаниями, занимающимися выгрузкой пользовательской информации.

Так, в 2017-2018 годах было судебное дело между социальной сетью ВКонтакте и компанией Double Data [10], занимающейся изучением кредитных возможностей людей на основе пользовательских данных (скоррингом). Представители социальной сети утверждают, что бесконтрольный сбор пользовательских данных является нарушением права пользователей ВКонтакте, в то время как представители скорринг-компании утверждали о законной возможности использования общедоступных данных.

По итогу дела суд первой инстанции отклонил иск ВКонтакте, обосновав это законностью деятельности компании Double Data [11]. Тем не менее, при повторном судебном разбирательстве Девятым арбитражным судом, победителем вышел истец, как обладатель исключительного права на владение и использование базы данных, содержащей информацию о своих пользователях [12]. Однако, представители Double Data снова просят повторного рассмотрения дела, так как считают решения суда не обоснованным и ссылаются на то что сбор информации происходит не из баз данных ВКонтакте, а исключительно из «кэшированной» информации, оставляемой самими пользователями. По итогу предыдущие постановления судов были аннулированы, дело передано в Арбитражный суд Москвы.

Приведенный пример демонстрирует несостоятельность текущего законодательства при рассмотрении новых проблем, возникающих при обработке общедоступных пользовательских данных для коммерческой реализации. И поскольку рынок систем, занимающихся разработкой прогностических моделей, активно растет – подобных прецедентов, требующих вмешательства со стороны судебных инстанций, будет становиться все больше.

Кроме правовой и этической стороны вопроса при разработке прогностических моделей пользователей, существует так же технический и методологический.

Для создания прогностических моделей требуется кооперация исследователей из множества областей: психология, социология, лингвистика, прикладная математика, математическая статистика, компьютерное моделирование. Одним из наиболее часто встречающихся способов исследования потенциальных взаимосвязей является регрессионный анализ данных.

При применении регрессионного анализа данных необходимо определить ряд зависимых и независимых величин, корреляция которых будет являться предметом исследования. От выбора этих ряда величин и будет зависеть направленность работ и их результат. Одним из наиболее часто применяемых способов создания прогностической модели является анализирование индивидуальности пользователя при помощи «Большой Пятерки» - пятифакторной модели личности человека. Благодаря данной модели возможно охарактеризовать личность человека согласно определенным диспозициям: экстраверсия, дружелюбие, сознательность, нейротизм, открытость опыту. Преобладание или комбинация определенных факторов позволяют судить о личности человека и его чертах характера.

Для создания прогностических моделей также необходимо определиться с набором пользовательских данных, на основе которых будет проводиться поиск взаимосвязей. Как правило, исследования концентрируются на отдельных аспектах «цифровых отпечатков» пользователя – лайки, сообщения, звукозаписи, статусы и т.д.

Помимо подобных данных, также используются данные о социальной структуре в социальных сетях. Исследования в данной области позволяют проанализировать наличие определенных взаимосвязей между членами сообществ, групп или друзей.

Стоит отметить, что анализ пользовательских данных не ограничивается информацией, осознанно оставленной пользователем: Многие приложения, сервисы и сайты, открыто предупреждают, что собирают данные о потребителях и передают её третьим лицам «в целях улучшения условий».

Помимо «цифровых отпечатков» оставленных в онлайн, обработке могут быть подвержены и «оффлайновые» данные – данные о геолокации со смартфонов, данные о сердцебиении со спортивных браслетов и т.п. Использование подобных данных позволяет создавать картину пользовательской мобильности в социуме: его перемещениях, предпочтениях, реакциях и пр.

Использование подобных данных может производиться без прямого уведомления и желания пользователя, что вызывает ряд этических проблем. Например, использование персональных данных таким образом может подорвать доверие к цифровым технологиям или вовсе заставить пользователей дистанцироваться от применения таковых.

С другой стороны, применение большого количества персональных данных для разработки прогностических моделей позволяет добиться более точечного моделирования человеческого поведения. Появление все новых автоматизированных средств анализа данных позволяет начинать новые исследования и открывать ранее неподтвержденные корреляции, дающие возможность определить тот или иной аспект человеческой природы.

ЛИТЕРАТУРА

1. Sandra M., Oded N. Using Big Data as a window into consumers' psychology // *Current Opinion in Behavioral Sciences*. 2017. № 18. P. 7 – 12.
2. Michal K. Mining Big Data to Extract Patterns and Predict Real-Life Outcomes // *Psychological Methods*. 2016. № 21. P. 493 – 506.
3. Barbakov O.M., Vinogradova M.V., Shatsky A. Social Portrait of Online Mass Media Audience in Russia // *Media Watch*. 2018. № 9. P. 383 – 396.
4. Xenos S., Ryan T. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage // *Computers in Human Behavior*. 2011. Vol. 27, № 5. P. 1658 – 1664.
5. Settanni M., Azucar D., Marengo D. Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis // *Personality and Individual Differences*. 2018. Vol 124. P. 150 – 159.
6. Sophie W.F., Susanne B.E., Patti V. Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp // *SAGE*. 2017. № 20. P. 1 – 19.
7. Kaiqi H., Qiao W., Zhenyang W. Natural color image enhancement and evaluation algorithm based on human visual system // *Computer Vision and Image Understanding*. 2006. Vol 1, № 103. P. 52 – 63.
8. Kim Y., Kim J.H. Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study // *Information Processing & Management*. 2018. Vol. 6, № 54. P. 1101 – 1114.
9. Gosling S., Gaddis S., Vazire S. Personality Impressions Based on Facebook Profiles // *International Conference on Weblogs and Social Media*. 2007.
10. Социальная сеть требует запретить сбор данных для банков // *Коммерсантъ*. URL: <https://www.kommersant.ru/doc/3206044> (дата обращения: 22.04.2019).
11. Право на поиск информации в интернете под угрозой. Решение по делу «ВКонтакте» ограничивает работу поисковиков / *Сколково*. URL: <http://sk.ru/news/b/pressreleases/archive/2018/01/29/pravo-na-poisk-informacii-v-internete-pod-ugrozoj--reshenie-po-delu-vkontakte-ogranichivaet-rabotu-poiskovikov.aspx> (дата обращения: 22.04.2019).
12. Постановление суда по интеллектуальным правам. / Суд по интеллектуальным правам. URL: http://kad.arbitr.ru/PdfDocument/1f33e071-4a16-4bf9-ab17-4df80f6c1556/4c9d2b02-4fbd-4554-82c8-53282523639c/A40-18827-2017_20180724_Reshenija_i_postanovlenija.pdf (дата обращения: 22.04.2019).