Использование результатов социального профилирования в процессе решения актуальных социально-экономических задач

А.С. Бождай, А.Ю. Тимонин

Пензенский государственный университет

bozhday@yandex.ru, c013s017b301f018@mail.ru

Аннотация

Разнородные открытые данные сети Интернет служат основой для различных прикладных исследований, одним из которых является социальное профилирование человека. Для разрешения этой задачи используется комплекс аналитических средств и технологий, позволяющих оперативно обрабатывать большие массивы неоднородных данных. Конечным результатом исследования является социальный профиль или их множество, которые представляют ценность в сфере социально-экономической аналитики. В статье представлен обзор актуального опыта использования персонифицированной информации, связанного с задачей социального профилирования. Даны практические примеры применения персональных социальных профилей в различных социо-экономических задачах. На основе тенденций развития информатизации общества делается вывод о наиболее вероятных в скором будущем сферах применимости социальных профилей, строящихся на основе технологий Big Data.

Ключевые слова: Big Data, анализ данных, неструктурированные данные, открытые источники данных, социальные медиа, социальный профиль человека

1. Введение

Обработка гетерогенных данных сети Интернет задействована в широком круге прикладных задач человеческой деятельности. Одной из них является построение социального профиля. Социальный профиль (СП) – это множество, состоящее из информации, способной тем или иным образом охарактеризовать человека, причем эта информация должна быть сгруппирована для удобства человеческого восприятия [1]. Задача социального профилирования относится к теоретической концепции анализа социальных сетей и использует схожие механизмы. Первоначально для учета всех связей между его элементами разрабатывается математическая модель, затем строится социальный граф [2], в котором узлы представлены социальными объектами, а ребра – социальными связями между ними. Пример такого графа показан на рисунке 1.

В качестве исходной информации социального профиля выступают открытые и персональные данные, получаемые из публичных источников сети Интернет. По своей природе, она является разнородной и разделяется на следующие типы: текстовая, мультимедийная, геопространственная, статистическая. Для их обработки и хранения требуются комплексный аналитический подход, учитывающий особенности каждого типа данных, и специализированные хранилища информации (реляционные и NoSQL). Совместное использование технологий Big Data и нейронных сетей в процессе анализа позволяет повысить точность определения смыслового содержания данных.

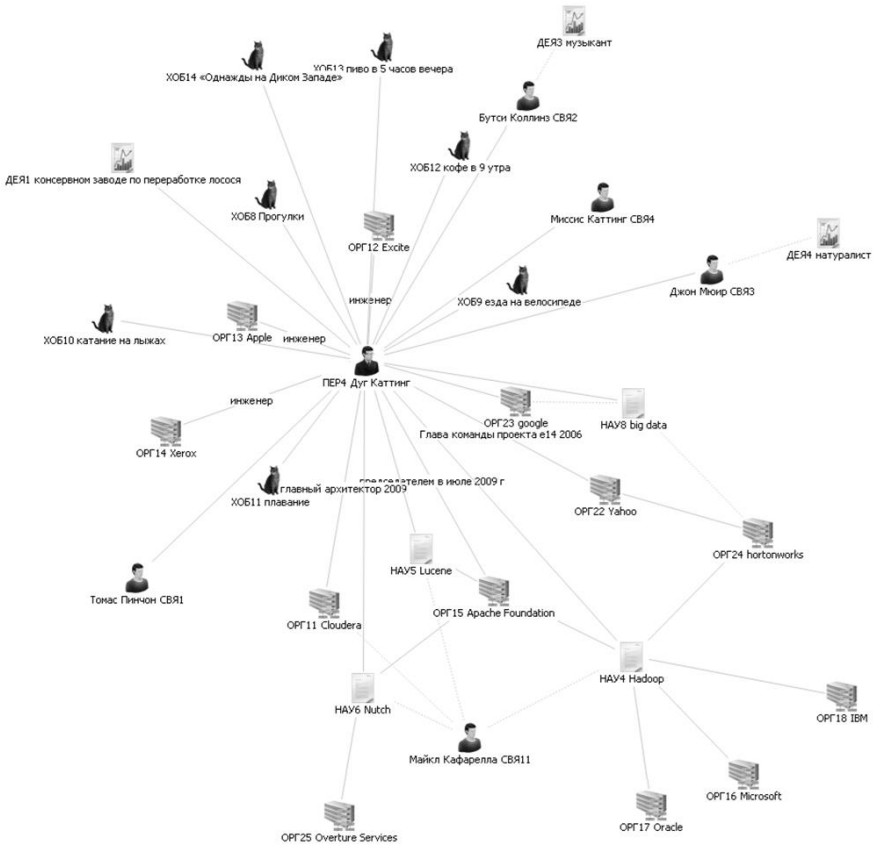


Рис. 1. Пример социального графа

Задача построения социального профиля разделяется на последовательность шагов [3]:

1. Сбор исходной информации из разных источников для большей полноты итогового СП;

2. Фильтрация данных от бессмысленной, противоречивой и конфиденциальной информации;

3. Выделение из множества собранных данных уникальных ключевых сущностей, которые станут центральными узлами сети, и создание на их базе структурированной информационной карты [4];

4. Распределение динамического контента по его типу в соответствующие хранилища информации и дальнейший поиск социальных объектов как в текстовых, так и мультимедиа данных;

5. Выделение связей между узлами (как однонаправленных, так и двусторонних);

6. Расширение сети до состояния, приемлемого для последующего анализа, за счет еще не включенных связанных объектов;

7. Поиск существующих множественных связей;

8. Поиск аналогий в нерассмотренных динамических данных и включение их в описание созданных объектов сети для исключения противоречивости данных;

9. Постройка социального графа на основе полученных сведений (рис. 2).

После завершения создания отдельного социального профиля или их групп, полученные результаты отправляются для рассмотрения специалистам, занимающимся прикладными задачами, у которых в качестве входных данных выступают СП.



Рис. 2. Структурная схема решения задачи социального профилирования

Данная работа посвящена рассмотрению как уже существующих подобных видов деятельности, так и потенциально возможных в ближайшем будущем из-за стремительного распространения в обществе сервисов интернет коммуникаций и открытых данных.

2. Обзор имеющегося опыта

По теме социального профилирования за последние годы в мире написано немало трудов как технической, так и социально экономической направленности. Рассмотрим часть из них.

Работа «The sociability score: App-based social profiling from a healthcare perspective» [5] авторства Пола Эскеса, Марко Спрута, Сяака Бринкемпера, Якова Ворстмана, Мартиена Дж. Каса описывает метод использования данных смартфона для диагностической оценки социального дефицита. Альфгейр Л. Кристьянссон, Майкл Дж. Манн, Меган Л. Смит и Инга Дора Сигфусдоттир исследуют закономерности в СП подростков, курящих электронные сигареты [6].

Сильвия Росси, Франсуа Ферланд, Адриана Тапас в статье «User profiling and behavioral adaptation for HRI: A survey» [7] рассматривают процесс профилирования пользователей для задачи взаимодействия человека и робота.

Лукаш Д. Качмарек, Мачей Бехке, Тодд Б. Кашдан и другие в научной работе «Smile intensity in social networking profile photographs is related to greater scientific achievements» [8] опубликовали вывод о зависимости между успешностью ученых и частотой улыбок на их фотографиях, полученных из различных открытых источников.

Статья под названием «A social LCA framework to assess the corporate social profile of companies: Insights from a case study» [9] предлагает новую методологию, которая помогает менеджерам в оценке социальных последствий, накапливающихся в повседневной деятельности компании, и в создании сильного социального профиля корпорации.

Целью исследования Хасиба Мустафы, М. Джунайд Анвера и Сайеда Сикандер Али в «Measuring the Correlation between Social Profile and Voting Habit and the moderation effect of Political Contribution for this Relationship» является поиск ответа на вопрос, существует ли сильная корреляция между социальным профилем человека и его избирательной активностью [10].

Существует ряд работ, посвященных изучению различных аспектов процесса построения социального профиля. Лилиан Митроу, Мильтиадис Кандиас, Василис

Ставроу и Димитрис Грицалис в своей работе «Social Media Profiling: A Panopticon or Omniopticon Tool?» поднимают вопросы вторжения в частную жизнь и дискриминации конкретного человека со стороны общества в процессе социального профилирования [11]. Франсиско Рангеля и Паоло Россо описывают новый графовый метод [12] для определения различных признаков, таких как возраст и пол, автора на основе его публикаций. Дэвид Н. Чин и Уильям Р. Райт также занимаются рассмотрением особенностей анализа социальных медиа в целях профилирования людей [13].

Однако наблюдается недостаток научных работ, в которых поднимается вопрос необходимости создания социальных профилей людей из открытых источников сети Интернет на базе технологии Big Data, с описанием конкретных прикладных задач.

3. Сферы применимости социальных профилей

Говоря о процессе социального профилирования и практическом применении его результатов, важно отметить, что классические информационные системы, включающие в свой состав сервер с реляционной СУБД и веб-интерфейсом, малоэффективны с точки зрения производительности и экономии. Обработка данных может занять очень большой промежуток времени, в результате чего результаты потеряют свою актуальность до того, как будут получены, и как следствие вычислительные и энергоресурсы будут потрачены впустую. Внедрение технологии Big Data способно решить эти проблемы за счет более эффективного использования уже имеющихся аппаратных ресурсов посредством распараллеливания информационных потоков между кластерами и применения нереляционных хранилищ данных.

Область возможного применения социальных профилей, построенных на основе открытых источников информации, довольно обширна. Наиболее популярными в настоящее время являются задачи коллаборативной фильтрации [14] и противодействия преступности. Первое направление распространено среди коммерческих организаций и часто используется при анализе социальных сетей с целью выдачи рекомендаций на основе предпочтений отдельных пользователей или их сообществ. Рассмотрим второе направление. Анализ социальных профилей может использоваться в таких видах правоохранительной деятельности, как составление психологического портрета преступника, поиск улик, опрос свидетелей, проверка алиби подозреваемых, вычисление уровня напряженности в обществе и т.д. В частности, для данных целей правоохранительные органы проводят разведку на основе открытых источников (OSINT) [15].

Открытые социальные данные находят свое применение при разработке механизмов искусственного интеллекта и машинного обучения. Вот лишь некоторые известные примеры их практической реализации. Проектирование беспилотных автомобилей [16, 17] - совокупность социальных профилей водителей служит источником метаданных для обучения нейронной сети автопилота. Применение суперкомпьютера IBM Watson в различных сферах деятельности [18], в частности в медицине для диагностики и изучения заболеваний, а также предоставления скорректированных методов лечения пациентов. В этом случае исследуются не только результаты анализов и записи историй болезней пациентов из картотек медицинских учреждений, но и дополняющая информация из других источников об их активности и окружении. Столь глубокое изучение различных факторов позволяет ускорить и повысить точность выявления причин возникновения и особенностей протекания редких и опасных заболеваний [19], что увеличивает шансы на выживаемость и выздоровление больных.

Далее проведем обзор прикладных направлений, которые уже заинтересованы в обработке социальных данных. В скором будущем они могут стать основными потребителями систем социального профилирования.

Сферы юридических и банковских услуг непосредственно работают с людьми. Оценка возможных рисков при сотрудничестве для них очень значима и влияет как на доходы, так

и на репутацию. Социальные профили в деятельности подобных компаний будут выполнять роль резюме и документа-характеристики клиентов. На его основании организация может производить решения: разрешить или отказать персоне в своих услугах, подобрать наиболее подходящий пакет услуг, например, тип банковского вклада и его оптимальный срок.

Еще одной возможной областью применения социальных профилей является администрирование крупных социально-экономических систем. К таким системам в первую очередь относят «Умный город» - подходы, заключающиеся в интеграции информационно-коммуникационных технологий и Интернета вещей (IoT решения), для управления городским имуществом и улучшения качества жизни населения посредством оптимизации урбанистических процессов. Это также относится к индивидуальной подстройке окружения под конкретного человека с использованием «интернета вещей».

С другой стороны, использование социальных профилей в кадровых службах предприятий может упростить процедуру трудоустройства и повысить эффективность работы коллектива. Алгоритм работы кадровых служб с социальными профилями следующий: на первом шаге происходит отбраковка резюме, не соответствующих требованиям организации, далее проводится сортировка кандидатов по ряду оценочных критериев, значения которых добываются из профилей, выдаются вспомогательные рекомендации для принятия решений сотрудникам отдела кадров, и в завершение формируются списки кандидатов на прохождение итогового собеседования.

В образовательных учреждениях составление групповых социальных профилей позволит повысить эффективность образовательно-воспитательной работы. На их основе педагог может вносить коррективы в учебный план, с целью внедрения индивидуального подхода к каждому обучающемуся. Также социальные профили помогут более эффективно предотвращать и разрешать конфликтные ситуации в коллективе обучающихся.

В научной сфере возможно моделирование социальных процессов на основе множества социальных профилей, разработка более гибких механизмов социальной политики. Сюда относятся исследование и прогнозирование поведения людей в различных нестандартных ситуациях, упрощение проведения соц. опросов. Профили могут найти свое применение при написании биографий или оценке исторических событий. Например, чтобы показать отношение современников, провести сравнительную оценку по таким критериям, как уровень жизни людей из одной социальной группы в разные исторические периоды.

Различные организации ведут собственные проекты, связанные с созданием и практическим использованием социальных профилей. Среди них Google Analytics Solutions, Microsoft Social Computing, Usalytics, но публикаций с результатами их исследований недостаточно для объективной оценки. Поэтому текущая работа имеет важность с точки зрения прогнозирования вариантов практического использования результатов социального профилирования. Как отмечалось выше, область применения существующих решений ограничена. Приведем ряд примеров. FindFace – проект для поиска людей по эталонному изображению. Wobot, Klout, IQBuzz – решения [20] в электронном маркетинге и анализе социальных отношений. Охват только популярных социальных сетей и малоэффективность в задаче идентификации человека внутри сети приводит к тому, что приходится обрабатывать крупные объемы данных вручную [4].

Нужно отметить, что закрытость социальных сетей негативно влияет на сбор необходимой для построения социального графа информации. Во многих ресурсах сети Интернет используются технологии Single Page Application, Ajax и DHTML для динамического вывода информации, что вызывает проблемы при работе с парсерами и поисковыми роботами. Основным недостатком генерации рекомендаций является проблема холодного старта — расчёт рекомендации для новых социальных объектов. Другая проблема процесса социального профилирования, препятствующая его массовому

распространению, состоит в различии социальных сетей: главную роль играют семантика связей между социальными объектами и графами разной топологии.

В большинстве случаев для решения конкретной прикладной задачи требуется все разнообразие данных социального профиля. Поэтому аналитикам приходится сталкиваться с проблемой гибкого представления результатов СП (см. рис. 3).

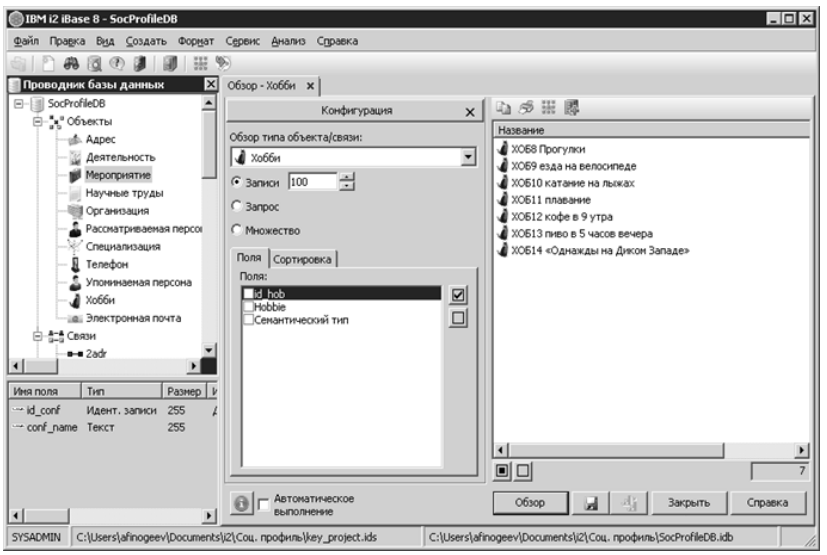


Рис. 3. Реализация графовой базы данных социального профиля

Таким образом можно выделить наиболее очевидных потенциальных потребителей разрабатываемой системы:

- коммерческие организации – для «персонализированного» продвижения товаров и услуг, а также обеспечения лучшего взаимодействия с потребителем;
- муниципалитеты и мэрии – оптимизация и постройка городской инфраструктуры, обеспечение диалога между горожанами и властью, достаточно точное выявление предпочтений народа;
- правоохранительные органы – для поиска преступников и составления карт напряженности общества с целью предупреждения возможных правонарушений;
- кадровые службы предприятий – для упрощения и повышения эффективности поиска и найма сотрудников, организации различных мероприятий, таких как конференции, семинары…;
- НИИ и организации, исследующие поведение социума – различные прикладные исследования, такие как: анализ избирательной активности, определение толерантности общества, медицинская статистика (прогнозирование эпидемий, выявление новых методов лечения, диагностика), исследование психических процессов человека, автоматизация процессов машинного обучения (разработка автопилотов, переводчиков) и т.д.

4. Заключение

Конечными результатами процесса социального профилирования являются база данных СП и графическое представление в виде социального графа. Граф сравнительно прост для визуального анализа и достаточно нагляден. На основе баз данных социального профиля возможно построение статистических диаграмм, получение выборок по конкретным вопросам с помощью применения NoSQL-запросов к ним.

Можно сделать вывод, что область возможного применения социальных профилей широка, но еще недостаточно освоена. Предполагается, что они будут активно использоваться в сфере бизнеса для оценки покупательского спроса и построения успешного диалога с клиентами, правоохранительных органах для предотвращения правонарушений, отделах кадров различных учреждений для более гибкого взаимодействия с рабочим коллективом и подбора новых сотрудников, а также в направлении Интернета вещей.

Литература

[1] Тимонин А.Ю. Разработка социального профиля на основе имеющихся инструментальных средств // Новые информационные технологии и системы: сборник научных статей XII международной научно-технической конференции, - Пенза: Изд-во ПГУ, 2015. С. 221-224.

[2] Тимонин А.Ю. Визуализация данных социального профиля для использования в прикладных задачах // Вестник Пензенского государственного университета. 2017. №3 (19). С. 112-118.

[3] Тимонин А.Ю., Бождай А.С. Уровни математического описания процесса построения социального профиля // Новые информационные технологии и системы: сб. научн. статей XIV Международной научно-технической конференции (г. Пенза, 22 – 24 ноября 2017 г.) – Пенза: Изд-во ПГУ, 2017. С. 353 – 357.

[4] Timonin A. Y., Bershadsky A. M., Bozhday A. S. The Process of Personal Identification and Data Gathering Based on Big Data Technologies for Social Profiles. // First International Conference «Digital Transformation and Global Society (DTGS)», St. Petersburg, Russia, June 22-24, 2016. Volume 674 of the series Communications in Computer and Information Science. Springer, 2016. P. 576-584. DOI=https://link.springer.com/chapter/10.1007/978-3-319-49700-6_57

[5] Eskes P., Spruit M., Brinkkemper S., Vorstman J., Kas M.J. The sociability score: App-based social profiling from a healthcare perspective // Computers in Human Behavior. 2016. Vol. 59. P. 39-48. DOI=https://doi.org/10.1016/j.chb.2016.01.024

[6] Kristjansson A.L., Mann M.J., Smith M.L. Social Profile of Middle School-Aged Adolescents Who Use Electronic Cigarettes: Implications for Primary Prevention // Prevention Science, 2017. P. 1–8. DOI=https://doi.org/10.1007/s11121-017-0825-x

[7] Rossi S., Ferland F., Tapus A. User profiling and behavioral adaptation for HRI: A survey // Pattern Recognition Letters. 2017. Vol. 99. P. 3-12. DOI=https://doi.org/10.1016/j.patrec.2017.06.002

[8] Kaczmarek L., Behnke M., Kashdan T., Kusiak A., Marzec K., Mistrzak M., Włodarczyk M. Smile intensity in social networking profile photographs is related to greater scientific achievements // The Journal of Positive Psychology. 2017. DOI=http://dx.doi.org/10.1080/17439760.2017.1326519

[9] Tsalis T., Avramidou A., Nikolaou I.E. A social LCA framework to assess the corporate social profile of companies: Insights from a case study // Journal of Cleaner Production. 2017. Vol. 164. P. 1665-1676. DOI=https://doi.org/10.1016/j.jclepro.2017.07.003

[10] Mustafa H., Anwer M.J., Ali S.S. Measuring the Correlation between Social Profile and Voting Habit and the moderation effect of Political Contribution for this Relationship // International Journal of Management & Organizational Studies. 2016. Vol. 5, № 1. P. 44-57. URL: http://www.ijmos.net/wp-content/uploads/2016/05/Haseeb-et.-al.pdf (дата обращения: 17.02.2018).

[11] Mitrou L., Kandias M., Stavrou V., Gritzalis D. Social Media Profiling: A Panopticon or Omniopticon Tool? // Proc. of the 6th Conference of the Surveillance Studies Network. 2014. [Электронный ресурс] URL: https://www.infosec.aueb.gr/Publications/2014-SSN-Privacy%20Social%20Media.pdf (дата обращения: 17.02.2018).

[12] Rangel F., Rosso P. On the Multilingual and Genre Robustness of EmoGraphs for Author Profiling in Social Media // Experimental IR Meets Multilinguality, Multimodality, and Interaction. Lecture Notes in Computer Science, 2015. Vol. 9283. Springer, Cham. DOI=https://doi.org/10.1007/978-3-319-24027-5_28
[13] Chin D., Wright W. Social Media Sources for Personality Profiling // UMAP 2014 Extended Proceedings. EMPIRE 2014: Emotions and Personality in Personalized Services, 2014. P. 79-85. URL: http://ceur-ws.org/Vol-1181/empire2014_proceedings.pdf (дата обращения: 17.02.2018).
[14] Rajeswari M. Collaborative Filtering Approach For Big Data Applications in Social Networks // International Journal of Computer Science and Information Technologies, 2015. Vol. 6 (3). P. 2888-2892. URL: https://www.scribd.com/document/289535998/Collaborative-Filtering-Approach-for-Big-Data-Applications-Based-on-Clustering-330-pdf (дата обращения: 17.02.2018).
[15] Richelson J.T. The U.S. Intelligence Community. 7th Edition. Routledge, 2015. 650 p.
[16] Ализар А. Знаменитый хакер Geohot собрал беспилотный автомобиль у себя в гараже [Электронный ресурс] // Geektimes.ru URL: https://geektimes.ru/post/267820/ (дата обращения: 17.02.2018).
[17] Kwon D., Park S., Ryu J.-T. A Study on Big Data Thinking of the Internet of Things-Based Smart-Connected Car in Conjunction with Controller Area Network Bus and 4G-Long Term Evolution // Symmetry. 2017. Vol 9 (152). DOI= http://dx.doi.org/10.3390/sym9080152
[18] Chen Y., Elenee Argentinis J.D., Weber G. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research // Clinical therapeutics. 2016. Vol. 38, № 4. P. 688-701. DOI=http://dx.doi.org/10.1016/j.clinthera.2015.12.001
[19] Bakkar N., Kovalik T., Lorenzini I. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis // Acta Neuropathol. 2018. Vol. 135, № 2. P. 227–247. DOI=https://doi.org/10.1007/s00401-017-1785-8
[20] Изместьева Е. 12 инструментов для мониторинга и аналитики социальных медиа [Электронный ресурс] URL: https://te-st.ru/tools/tools-monitoring-and-analysis-of-social-media/ (дата обращения: 17.02.2018).

Utilization of the Social Profiling Results in the Process of Solving Urgent Social-Economic Tasks

Alexander S. Bozhday, Alexey Y. Timonin

Penza State University

The Internet open data serve as the basis for various applied researches. One of them is the social profiling process. A set of analytical tools and technologies is used to solve this problem. It requires to quickly processing large amounts of heterogeneous data. The final result of the study is a unique social profile or a group of them. They have value in the different fields of socio-economic analytics.

The current article provides a current experience overview of using personalized information related to the social profiling task. Practical examples of personal social profiles application in various socio-economic tasks are given. A conclusion is drawn on the trends basis of information society growth in the near future: the most likely applicability areas of social profiles based on Big Data technologies are submitted.

Keywords: Big Data, data analysis, personal social profile, public data sources, social media, unstructured data