

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАФОВЫХ МОДЕЛЕЙ

(автореферат бакалаврской работы)

студента 4 курса 451 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Донов Дениса Станиславовича

Научный руководитель

кандидат физико-математических наук, доцент

_____ Л.Б. Тяпаев
подпись, дата

Зав. кафедрой

кандидат социологических наук, доцент

_____ И.Г. Малинский
подпись, дата

Саратов, 2020

ВВЕДЕНИЕ

Актуальность исследования. Современное общество, вошедшее в стадию информационного, постиндустриального, характеризует рост ценности информационных технологий и коммуникаций.

Информатизация коснулась большого количества областей жизни общества: от использования в служебных целях для автоматизации основной деятельности до сферы продаж, предоставления услуг (наличный оборот постепенно вытесняют платежные системы), традиционные магазины отдают долю рынка Интернет-системам, широкое распространение также получают государственные услуги в электронном виде. Личная жизнь людей в настоящее время также связана с влиянием информационных технологий посредством социальных сетей.

В настоящее время большая часть человечества не представляет своей жизни без использования мессенджеров и социальных сетей, все больше личной информации становится доступной для просмотра в данных сервисах.

В связи с этим, актуальность приобретает вопрос проведения анализа профилей пользователей с использованием социальных сетей. Данная задача актуальна для служб, работающих непосредственно с гражданами. Роль информатизации в обществе достаточно велика. Актуальность данной проблемы обусловлена тем, что информатизация каждой сферы жизнедеятельности общества приобретает всё большее влияние. Новейшие технологии в сфере информатизации и коммуникации ускоряют процесс развития общества. Один из первых шагов к информатизации общества – использование знаний во всех сферах жизни общества. Информационное общество – это следующая ступень развития цивилизации.

Одним из результатов эпохи технологического прогресса можно назвать использование социальных сетей для виртуального взаимодействия людей друг с другом. Социальная сеть – это уже устоявшийся способ общения в современном обществе, который затрагивает многие формы жизни. К примеру, социальные сети оказывают значительную роль в формировании читательских

предпочтений современной молодежи и являются одним из ресурсов для формирования этих предпочтений. В последнее время одной из основных форм проведения досуга среди молодого поколения стало времяпрепровождение в социальных сетях.

Степень научной разработанности проблемы. Широкий пласт научных статей и исследований посвящаются активно развивающейся тенденции использования социальных сетей в разных целях. Социальные сети исследуются в социологии как способ взаимодействия людей друг с другом, например, в трудах А. В. Щекотурова «Гендерные самопрезентации подростков на страницах социальной сети «ВКонтакте»¹. Родоначальником теории графов считается выдающийся математик, член Петербургской академии наук Леонард Эйлер.

В 1736 году в одном из своих писем он формулирует и предлагает решение задачи о семи кёнигсбергских мостах, ставшей впоследствии одной из классических задач теории графов.

Объект исследования - информационные технологии анализа профиля пользователей социальных сетей.

Предмет исследования - графовые модели как технология анализа профиля пользователей социальных сетей.

Цель данной работы: анализ использования возможностей графовых моделей при изучении профилей пользователей социальных сетей.

Задачи:

- изучение алгоритмов использования графовых моделей;
- анализ алгоритмов анализа данных по профилям пользователей в социальных сетях с использованием графовых моделей;
- оценка функционала существующих решений в области автоматизации сбора данных о пользователях;

¹ Щекотуров, А. В. Гендерные самопрезентации подростков на страницах социальной сети «ВКОНТАКТЕ» // Мониторинг общественного мнения. 2012. № 4 С. 80–88.

- разработка информационной модели задачи анализа профиля пользователей социальных сетей с использованием графовых моделей.

Теоретическую основу бакалаврской работы составляют научные работы как зарубежных, так и отечественных исследователей.

Эмпирическая база бакалаврской работы состоит из контента систем анализа социальных сетей RADIAN6, ALTERIAN SM2, RECORDED FUTURE, «PALANTIR» (КОРПОРАЦИЯ PALANTIR).

Структура работы. ВКР состоит из введения, двух глав по три параграфа, заключения и списка использованных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе «Анализ социальных сетей и графов» рассматривается процесс информатизации общества, история развития данного понятия, особенности, технологии, перспективы. А также появление понятия «Социальная сеть» в современном мире.

В современном обществе определяющую роль играет информация. В повседневную жизнь приходят различные инновации, которые ориентируют человека на получение новой информации, на улучшение навыков и знаний, на повышение квалификации. Развитые страны мира находятся на стадии «информационного» становления социума. Инфраструктуру такого общества формируют методы и средства сбора, хранения, обработки информации. Информация становится одним из основных источников коммуникации.

Информатизация общества – это глобальный социальный процесс, в котором основополагающим видом деятельности в сфере общественного производства является сбор, обработка, анализ, накопление, передача и хранение информации, осуществляемые с помощью использования информационных технологий. Информационные технологии окружают современное поколение повсюду. Во многом, это упрощает повседневные действия человека. Информационные технологии можно рассматривать как элемент современного информационного общества, который регулирует,

поддерживает, сохраняет и совершенствует такое общество. На протяжении многих тысячелетий информация передавалась с помощью традиций и обычаев, культурных стереотипов, различных правил и предписаний, но в настоящее время главную роль занимают технологии. Информационные технологии имеют ряд основных функций:

- позволяют активизировать и эффективно использовать информационные ресурсы общества, которые сегодня являются наиболее важным стратегическим фактором его развития;
- позволяют автоматизировать и оптимизировать информационные процессы, которые в последнее время занимают значительное место в жизнедеятельности современного общества;
- информационные процессы являются важными элементами более сложных социальных или же производственных процессов.

К основоположникам в области теории постиндустриализма и формирования информационного общества следует отнести: М. Маклюэна, Дж. Гэлбрейта, М. Кастельса, Э. Тоффлера, Д. Белла, Е. Масуда. Значимую роль в адаптации положений теории постиндустриального развития к специфике постсоветских стран сыграли работы О. Н. Вершинской, В. Л. Иноземцева, В. П. Руднева, Д. С. Черешкина, В. А. Трапезникова, Г. Л. Смоляна, Н. Н. Моисеева, А. А. Лазаревича.

Мир опутан паутиной всемирной сети. Скорость распространения информации достигла огромных высот. Информация, накопленная за тысячелетия существования человечества, стала доступна любому человеку, который только пожелает ей обладать.

В настоящее время распространение социальных сетей привело к тому, что у подавляющего большинства граждан имеются аккаунты в одной и более социальных сетей.

Основными видами информации, оставляемыми пользователями социальных сетей являются:

- личная информация (ФИО, дата и место рождения, гражданство, место проживания и др.);

- данные об образовании и месте работы;

- данные об увлечениях;

- информация о личной жизни.

Кроме того, социальные сети представляют собой своего рода электронный дневник, в котором пользователи оставляют информацию об актуальных для них проблемах, событиях своей жизни. Также в социальных сетях существуют возможности работы с информацией о коммуникациях пользователя (наличие друзей, членство в группах и др.).

Источником информации при изучении личности пользователя социальных сетей также могут быть:

- загруженные фотографии, видеоматериалы и музыка;

- оставленные комментарии на форумах, в группах;

- переписка.

Изучение личности пользователей посредством социальных сетей в настоящее время может интересовать:

- кадровые службы предприятий - при изучении возможности трудоустройства претендента на ту или иную позицию;

- службы безопасности и кредитные организации банков при принятии решения о предоставлении кредита;

- органы внутренних дел при проведении расследований;

- спецслужбы.

Так, по анализам предпочтений пользователей может быть составлен психологический и социальный портрет личности, в соответствии с которым производится принятие тех или иных решений в его отношении. Например, при приеме на работу кандидат может указать свое отношение к работе и вредным привычкам как необходимо работодателю, а реальная информация, открытая им в социальных сетях, может показать совершенно иное.

Таким образом, при чрезмерной активности пользователя прозрачными могут становиться данные о:

- политических взглядах;
- увлечениях;
- состоянии здоровья (большой объем поставленных отметок "нравится" к медицинской тематике);
- типе темперамента, способности к сохранению конфиденциальной информации;
- склонности к получению внешних поощрений;
- о затратах личного и служебного времени на социальные сети;
- контактах, социальном статусе пользователя;
- проблемах личного характера;
- отношении к вредным привычкам;
- отношению к общественным ценностям, религии, склонности к влиянию со стороны экстремистских организаций.

Во второй главе «Практическое применение модели графов в социальных сетях» речь идет, о математических методах решения задачи. Обработка данных производится с использованием моделей анализа на графах, нейронных сетей и экспертных систем, дискриминантного и корреляционного анализа.

1.Кластерный анализ

При анализе и прогнозировании социально-экономических явлений исследователь довольно часто сталкивается с многомерностью их описания. Это происходит, например, при решении задач сегментирования рынка, построения типологии стран по достаточно большому числу показателей, прогнозирования конъюнктуры рынка отдельных товаров, изучения и прогнозирования экономической депрессии и многих других проблем.

Кластерный анализ — один из методов многомерной статистики — наиболее ярко отражает черты многомерности в процедуре классификации объектов. Название «кластерный анализ» происходит от английского слова

«cluster» — гроздь, скопление. Впервые определил предмет кластерного анализа и дал его описание исследователь Трион (Тгуоп) в 1939 г.

Главное назначение кластерного анализа — разбиение множества исследуемых объектов, характеризуемых совокупностью признаков, на однородные в соответствующем понимании группы (кластеры). Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней. Иными словами, предполагается выделение компактных, удаленных друг от друга групп объектов или отыскание «естественного» разбиения совокупности на области скопления.

Кластерный анализ является одним из направлений статистического исследования социально-экономических процессов, которые связаны с изучением массовых явлений.

Для проведения анализа профиля методом кластерного анализа, кроме проведения сбора информации, необходимо провести классификацию: на какое количество кластеров необходимо провести разделение данных и каким методом провести определение степени сходства в данных. Существует много разных мер сходства, наиболее употребительными из них являются порядка десяти.

2. Факторный анализ

В случае наличия большого числа параметров (более 100) имеет смысл сгруппировать параметры и анализировать уже не каждый параметр в отдельности, а группы параметров как единый комплексный параметр (фактор). В основе факторного анализа лежит идея о том, что за сложными взаимосвязями явно заданных признаков стоит относительно более простая структура, отражающая наиболее существенные черты изучаемого явления, а «внешние» признаки являются функциями скрытых общих факторов, определяющих эту структуру.

Например, для анализа структуры экономического роста России можно проанализировать все макроэкономические параметры, предварительно объединив их в группы. Одним из таких факторов будет являться ВВП [6].

Объединение параметров можно делать вручную, эмпирически, как это сделано с ВВП, а можно с помощью метода факторного анализа. Применение факторного анализа позволяет, во-первых, уменьшать (редуцировать) число рассматриваемых параметров, во-вторых — находить осмысленные группы параметров, каждая из которых будет являться одним самостоятельным параметром. Спецификой этого метода является то, что при объединении параметров в факторы каждый фактор аккумулирует в себе общие закономерности во всех параметрах, отбрасывая особенности каждого параметра в отдельности.

3. Нейронные сети

Нейронные сети имеют свои особенности, но тип конфигурации аналогичен показанному выше. В модели искусственных нейронов существует множество входных сигналов, обозначаемых $x_1, x_2 \dots, x_n$. Данные входы соответствуют сигналам, поступающим в синапсы биологического нейрона. Каждый сигнал умножается на соответствующий вес $W_1, W_2 \dots, W_n$, и поставляется суммирующий блок, обозначенный σ . Каждый вес соответствует "силе" одной биологической синаптической связи. Суммирующий блок, соответствующий телу биологического элемента, алгебраически сочетает в себе взвешенные входы, создавая видимость продаж.

Данную модель можно представить с использованием соотношения:

$$NET = \sum_{i=1}^n w_i x_i + w_0$$

Сигнал w_0 , который называется смещением, является значением функции предельных значений сдвига. Данный сигнал позволяет сместить происхождение функции активации, что дополнительно увеличивает скорость обучения. Этот сигнал добавляется к каждому нейрону, он учится, как и все остальные шкалы, и его особенность в том, что он подключен к сигналу +1, а не к выходу предыдущего нейрона.

Полученный сигнал NET как правило обрабатывается функцией активации и дает выходной нейронный сигнал OUT

В заключении сделаны основные выводы бакалаврского исследования, которые указывают, что в рамках данной работы проведено решение задачи анализа профиля пользователя социальной сети с использованием графового метода. В настоящее время с развитием социальных сетей производит накопление большого количества разнородной информации о пользователях. Анализ массива пользовательских данных позволяет получить множество информации об объекте исследования, характеризующей особенности личности, увлечения, склонности, политические убеждения, социальный статус и многое другое.

Проведение подобного анализа может быть интересно многим - от кадровых служб, банков до сотрудников органов внутренних дел и спецслужб.

В рамках данной работы проведена реализация графового метода анализа профиля пользователя социальной сети - определены задачи автоматизации, проведен анализ математических моделей для решения задач оценки профилей пользователей социальных сетей.

В рамках проектирования системы был определен перечень функциональных и нефункциональных требований к системе, построена математическая модель в соответствии со спецификой поставленной задачи. Проанализированы возможности использования экспертных методов, а также методов использования технологий анализа с помощью графовых методов.

Полученные результаты позволяют проводить анализ личности объекта, результаты которого могут использоваться в различных целях. Полученные результаты позволяют формировать отчетность в формате MS Excel, строить диаграммы, на которых приведены основные личностные характеристики изучаемого объекта. При этом даже отсутствие тех или иных сведений о пользователе может характеризовать его определенным образом.

В ходе анализа предметной области также показано, что социальные сети могут создать человеку как положительный, так и негативный образ. Поэтому, при работе с социальными сетями необходимо соблюдение мер предосторожности, при несоблюдении которых могут наступать негативные

последствия. При этом отказываться от использования социальных сетей в настоящее время не нужно, существует ряд методик создания положительного образа, которым необходимо следовать при работе в социальных сетях.