

СОЦИОЛОГИЯ ЦИФРОВОЙ СРЕДЫ

DOI: 10.24412/1994-3776-2022-3-53-59

УДК: 316.4

А.А. Ахтырский

ЦИФРОВЫЕ МЕТОДЫ В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

A. Akhtirskii. Digital methods in sociological research

Аннотация. Статья посвящена анализу цифровых методов исследования в социологической науке. Основываясь на статистических данных и литературном обзоре, в статье описано влияние «четвертной парадигмы» на социологические методы исследования и определена роль больших данных в социологических исследованиях. Статья будет полезна практикующим социологам, аналитикам и преподавателям.

Ключевые слова: цифровая социология, данные в социологии, цифровизация, искусственный интеллект в социологии, большие данные.

Контактная информация: 193023, Санкт-Петербург, Садовая ул., д.21; тел.: (812) 310-19-04; e-mail: andrei_1893@mail.ru

Abstract. The article is devoted to the analysis of digital research methods in sociological science. Based on statistical data and a literature review, the article describes the influence of the “fourth paradigm” on sociological research methods and defines the role of big data in sociological research.

The article will be useful to practicing sociologists, analysts and teachers.

Keywords: digital sociology, data in sociology, digitalization, artificial intelligence in sociology, big data.

Contact information: 21 Sadovaya street, Saint-Petersburg, 193023; tel.: (812) 310-19-04; e-mail: andrei_1893@mail.ru

Введение

Социальные науки охватывают базовый подход к индивидуальному социальному поведению, но экстраполяция данного феномена выявляет совокупный паттерн общего значения, что позволяет описать существующие социальные связи и процессы. Долгий период времени этому способствовали классические социологические исследования. Вместе с тем современные социологи отмечают сложности при проведении опросов в виду повышенного % отказов участия респондентов. Например В. Федоров отметил: *«Сегодня люди менее склонны отвечать на наши вопросы, причем количество вопросов, которые мы можем им задать, также становится меньше, потому что ценность времени выросла»*.....*«Все стало очень подвижным, переменчивым. И разработать или модифицировать какую-то модель устройства общества становится все сложнее»* [3]. Отказы респондентов от участия в социологических исследованиях увеличивают вероятность смещения данных из-за выпадающих из выборки потенциальных респондентов. Как следствие, это может привести к недостоверным выводам.

Данная проблемная ситуация все чаще подталкивает исследователей к применению междисциплинарного подхода и применению больших данных в работе. Этому способствуют интеллектуальные машины и компьютерные системы, которые выявляют поведенческие модели социальных групп или общества в целом, чтобы на основе результатов предсказывать с высокой точностью будущие тенденции.

Ахтырский Андрей Александрович – кандидат социологических наук, старший преподаватель кафедры социологии и управления персоналом. Санкт-Петербургский государственной экономической университет.

A. Akhtyrskii – PhD in Social Sciences, Senior Lecturer of the Department of Sociology and Human Resource Management, Saint-Petersburg State Economic University.

© Ахтырский А.А., 2022.

Как следствие в данной статье будут проанализированы основные цифровые технологии и концепции, которые могут оказать существенное воздействие на развитие цифровой социологии. Представляется, что это поможет определить роль цифровизации в социологии и будущее направление развития социологических методов исследования. Основанием для анализа послужил литературный обзор зарубежных и отечественных статей на тему цифровой социологии, а также анализ статистических данных Всемирного банка.

Основная часть. Ориентация на данные

Начиная с начала 2000-х, исследователи открыли для себя пользу алгоритмов, кода и платформ, привлекая внимание общественности к этим ценным практикам. Ученые, в первую очередь в области антропологии, коммуникации, права, философии, науки и техники, определили алгоритмы как объект гуманитарных и социальных исследований.

При этом еще в 1985 г. Вулгар исследовал возможность связи между социологией и искусственным интеллектом, основываясь на альтернативном восприятии того, что искусственный интеллект представляет собой «повод для переоценки центральной аксиомы социологии» [25]. Вулгар призывал к развитию социологии машин для анализа того, как системы ИИ функционируют как человеческие социальные акторы, формируя социальные отношения и конструируя социальные реальности. Эта точка зрения напоминает акторно-сетевую теорию и предполагает, что ранние социологические исследования ИИ воспринимались в первую очередь как научно-техническое развитие.

Затем в 1989г. Шварц дал раннее и новаторское представление об искусственном интеллекте как социологическом феномене и предположил, что точность такого современного подхода зависит от «условий знания в современном обществе». По мнению ученого, важной задачей социологов является изучение того, как системы ИИ «проникают и трансформируют социальные институты» [24]. Т. е. ученый ориентировал социологов на исследование данных о социальной реальности.

Данные никогда не бывают нейтральными, вместо этого они несут ценность. Например, данные статистики смертности представляют собой нечто большее, чем совокупное количество умерших лиц. Более века назад Дюркгейм продемонстрировал, что уровень самоубийств соответствует социальной интеграции [5], а высокую младенческую смертность рассматривал как «показатель социального положения» [6]. Как следствие, социологи признают, что то, что считается данными, является социализированным, политизированным и многоуровневым, поскольку данные о людях также часто являются данными о социальных структурах.

В социальных науках «большие данные» стали известны благодаря качеству своего предсказательного потенциала, впервые продемонстрированного крупнейшим американским ритейлером – сетью Walmart, который даже назвал свой проект обработки данных Social Genome. С помощью применения нейросетей и сбора данных о своих клиентах сеть Walmart довела точность прогнозов покупательского поведения до 83%.

Пожалуй, именно большие данные больше всего обсуждается в контексте применения цифровых методов социологических исследований и вызывает самые ожесточенные споры в среде социологов. Ю. Н. Толстова пишет, что «ядром этой технологии является набор достижений компьютерной науки, позволяющих искать и формировать массивы больших, разбросанных, изменяющихся, структурированных и неструктурированных данных» [15].

Современные ученые в области искусственного интеллекта признают существование социального и используют несколько стратегий для его учета в рамках обработки больших данных. Типичный практический подход включает расширенный сбор данных о измерениях социальных миров, например, когда в электронных медицинских картах включается информация о демографических данных пациента (этническая принадлежность, пол,

возраст) и других социальных детерминантах (жилищный статус, стабильность питания, социальная поддержка).

Эти подходы все больше ориентируют ученых на использование интеллектуального анализа больших данных, описывающих социальную реальность, задачи которого связаны с поиском закономерностей в открытых, постоянно пополняемых больших объемах данных.

Появление все больше данных о действиях групп людей является следствием развития технологий и увеличения интернет-аудитории, что открыло новые возможности и ограничения для организации социологических исследований.

Долгий период времени США была страной, в которой Интернет стал неотъемлемым элементом повседневной жизни большинства населения. Рисунок 1, свидетельствуют о том, что до 2010 г. американская аудитория преобладала среди всех стран по количеству пользователей Интернета. Логично, что идеи использования Интернета в качестве инструмента для сбора социологической информации начали прорабатываться именно американскими социологами [4]. Однако с 2012 процентное соотношение интернет-аудитории Китая стало преобладать над аудиторией США и закрепилось по настоящее время.

Стоит обратить внимание на темпы роста пользователей интернета в России: долгий период времени россияне, по сравнению с Китаем и США значительно в меньшей степени пользовались интернетом. Например, в 2006 г. США и Китай насчитывали 70% и 60% пользователей, тогда как в России их было всего 20%. Вместе с тем, с того времени российская интернет-аудитория быстро увеличилась и в 2020 г. сравнялась с развитыми странами по количеству пользователей интернетом (рис. 2). В 2020 году в России 85% населения пользуются интернетом, что на 30% больше общемирового показателя.

Высокие темпы распространения интернета в России сопровождается рост интенсивности его использования: доля самых активных пользователей – тех, кто выходит в Сеть ежедневно, – выросла по сравнению с 2010 г. в 2.2 раза, достигнув 57.7% в 2016 г. [14].

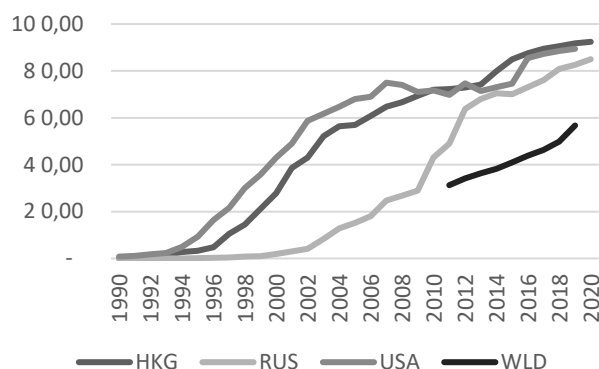


Рисунок 1. Сравнительная характеристика интернет-аудитории России с Китаем и США, в % от всего населения

Составлено автором, данные из [2]

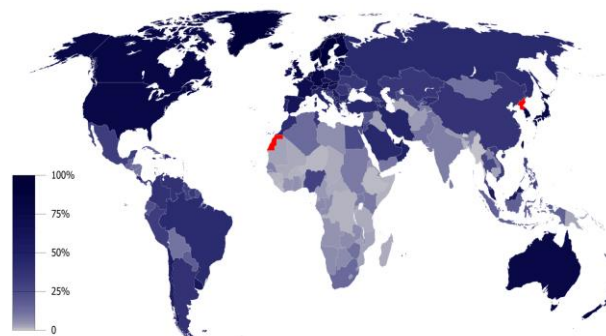


Рисунок 2. Сравнительная характеристика интернет-аудитории в мире

Мировые объемы данных ежегодно увеличиваются. Согласно оценкам и прогнозам только за последние 2 года было создано 90% всех данных в мире и этот объем будет ежегодно увеличиваться на 40% [9]. Поэтому социологи все чаще используют данные с различных мобильных приложений людей, для определения их социальной и интернет-активностей. В ходе «цифровизации» социальной жизни пользователи добровольно выкладывают большие массивы личной информации. В перспективе такая прозрачность обеспечивает пространство слежки за огромным множеством людей. В том числе различные мобильные приложения собирают данные с мобильного телефона:

- ip-адреса и ID мобильного устройства (например, mac-адрес, imei);
- данные об устройстве (например, язык использования, вид операционной системы, тип браузера);
- cookie и аналогичные технологии;
- местоположение на основе ip-адреса;
- точные данные о геолокации (gps, с вашего согласия);
- данные о использовании мобильных сервисов (например, метрики использования приложений и взаимодействия с ними).

Кроме этого, в качестве основных источников больших данных выступают оперативная с предприятий, логистики, интернета вещей, информация о взаимодействии людей и информация о положении в мире интернета, а также данные, полученные в ходе научных исследований и т.д. [18].

Несомненно, обилие такой информации может вскружить голову исследователю. Например, вот что написал Д. Кинг в своей работе «Меняющаяся доказательная база социальных исследований»: *«Вместо того чтобы пытаться каждые два года извлечь мнения о политике у нескольких тысяч активистов путем искусственно созданной ситуации разговора в виде опросного интервью, мы можем использовать новые методы и получить десятки миллионов политических мнений, которые появляются ежедневно в блогах. Так же как вместо того, чтобы изучать влияние контекста на взаимодействия людей, спрашивая респондентов об их последних контактах, мы можем собрать информацию за длительный промежуток времени об их телефонных звонках, письмах и сообщениях. При отсутствии официальной статистики мы можем судить об экономическом развитии и росте населения, основываясь на информации со снимков спутника об освещенности, расположении дорог и других объектов инфраструктуры.»* [23].

Данные ситуации создают предпосылки серьезной конкуренции цифровых методов над классическими анкетными методами в социологии, формируя все больше оснований развития датаизма среди исследователей. Датаизм – образ мышления, основанный на предании большим данным (Big Data) первостепенного значения. По мнению представителей данной парадигмы, применение больших данных в исследовании имеют безграничные возможности. Исследователи отмечают формирование своего рода фетишизма больших данных и самообучающихся алгоритмов.

Однако ученые утверждают, что последствия широкого и повсеместного использования цифровых технологий с позиции последствий для общественных отношений не определены [10]. По оценкам экспертов завышенные ожидания в отношении систем искусственного интеллекта, необоснованно подогреваемые не только СМИ, но и крупными цифровыми проектами российских чиновников, могут обрушиться, что повлечет за собой еще большее снижение доверия к тем организациям, которые их используют» [12].

Эти противоречия служат основанием для обсуждения среди исследователей возможности применения цифровых технологий в социологических исследованиях.

Обсуждение

Большие данные, одно из современных направлений исследований, в том числе и в социальных науках, являются частью более масштабного и глубокого тренда трансформации научного знания — формирования так называемой «четвертой парадигмы» (или в другой терминологии e-науки (e-science — электронной науки) либо науки, интенсивно использующей данные (data-intensive science) [16].

В рамках четвертой парадигмы в истории науки формируется исследовательское поле на пересечении социальных и компьютерных, информационных наук, получившее название «вычислительные социальные науки» (computational social science). Неслучайно в

исследовательской литературе отмечается, что «вычислительная социальная наука — это комплексное, междисциплинарное исследование социальных систем как структур, обрабатывающих информацию, и при помощи передовых вычислительных систем» [19].

К. Чиоффи-Ревилла различает пять основных методов, которые используются сегодня в вычислительных социальных науках: автоматизированное извлечение информации, анализ социальных сетей (SNA), геопространственный анализ (социальные ГИС), моделирование сложности и социальное моделирование [19].

Российский социолог и специалист по методологии социологического исследования Ю.Н. Толстова определяет несколько компьютерных технологий, которые активно применяются в социальных науках:

- большие данные (Big Data);
- интеллектуальный анализ данных (Data Mining);
- цифровые гуманитарные науки (Digital Humanity);
- наука о данных (Data Science) [15].

Интеллектуальный анализ как правило относится к сфере искусственного интеллекта и подразумевает несколько особенностей использования данной технологии социологами:

1. социолог нацелен не на определение типологии респондентов, а на фиксацию в большой массе данных закономерностей, описывающих ранее не известные структуры связей и т.п.;

2. при выявлении закономерностей в больших объемах данных социологи обрабатывают, комбинируют и преобразуют их таким образом, чтобы проверить интересующих исследователя гипотез о связях и закономерностях;

3. data mining имеет широкий набор инструментов визуализации больших данных, что дает возможность работать с ними социологам без дополнительной математической подготовки.

Как следствие, исследователи отмечают следующие положительные стороны применения технологий больших данных и интеллектуального анализа в социологии:

- увеличение предсказательного потенциала в сравнении с массовыми опросами, проблемы которых начали активно обсуждаться с 2011 г. в России и других странах;
- возможность сплошного обследования генеральной совокупности позволяет точнее формулировать крупные, эмпирических обоснованные выводы;
- большие данные постоянно обновляются и увеличиваются за счет интеграции новых массивов, в том числе неструктурированных данных [7].

Несомненно, это описывает высокий потенциал эвристического подхода больших данных над классическими «малыми данными» в социологии. Вместе с тем, описываемые технологии не лишены недостатков. Критика концепции больших данных в социологической литературе включает в себя две основные группы аргументов: технические и этические. Если к этическим проблемам как правило относят нарушение приватности при отслеживании деятельности поведения человека, используя его «цифровые следы» и «цифровое неравенство», [16] то с технической стороны больше сложностей. К техническим проблемам применения больших данных в социологии относят:

- сложности в определении генеральной и выборочной совокупностей и анализе интернет-данных. Например, при сборе данных активности пользователей и их комментариев в социальных сетях, сложно автоматически определить какие комментарии написаны людьми, а какие написаны ботами;

- сложности сбора и регистрации данных в силу их закрытого характера. Например, А.В. Одицов пишет, что «проблемы закрытости баз данных социологических исследований не позволяет не только эффективно интегрировать уже полученные и имеющиеся в распоряжении социологии данные, но и вторично их использовать» [7].

– из-за большого количества переменных при первичном анализе и поиске корреляций (data-mining) алгоритмы могут исказить результат.

Данные ограничения все больше побуждают социологов к применению дополнительного программного обеспечения при проведении исследований. Как следствие можно сказать, что современному социологу необходимо овладеть компетенциями извлечения данных из различных виртуальных источников и их скачивания для будущего анализа. Вместе с тем, несмотря на наличие большого массива количественной и качественной информации об «отношениях» между человеком и машиной и о взаимодействии людей «через» машины, социологи мало занимаются данной проблематикой, отдавая ее на откуп уже не кибернетикам, но специалистам в области computer science [11].

Как следствие развитие темы больших данных в социологии с новой силой напомнило о «методологических травмах социологов», т. е. о ситуациях «растерянности исследователей перед обилием социологических теорий, методологий, методов и в процессе принятия решений о выборе средств познавательной деятельности [13].

В этой связи предлагаются различные стратегии для социологов по отношению к большим данным: критика и поиск уязвимостей Big Data; сотрудничество с Big Data; изучение Big Data как особой техники; игнорирование Big Data [8].

Однако в большинстве случаев социологи предпочитают сотрудничать с Big Data, что подтверждается возрастающим использованием больших данных для анализа социальных институтов, практик, норм и т.д. Например, исследование В.В. Волкова, Д.А. Скугаревского и К. Д. Титаева, в котором на основе анализа базы данных решений судов РФ рассматривается влияние социального статуса подсудимого на решение суда [1]. Другим интересным примером применения технологий больших данных является исследование Г. А. Николаенко «цифровых следов» на примере научной социальной сети Research Gate [6]. Так и зарубежные ученые часто проводят исследования с использованием больших данных, например, исследование городских трущоб на основе данных сотовых операторов [21], анализ сообщений Twitter и прогнозирование на их основе социальных настроений [22], анализ читательских предпочтений на основе чтения данных с электронных устройств для чтения книг [17].

Заключение

На основании представленного материала можно сделать некоторые выводы относительно развития цифровых технологий в социологических исследованиях:

– ежегодное увеличение объемов данных и интернет-пользователей будет все чаще подталкивать социологов к использованию Big Data и предиктивной аналитики в социологических исследованиях;

– для эффективного использования больших данных при исследовании общества, необходимо на институциональном уровне развивать культуру «открытых данных», что в свою очередь повысит качество и точность цифровых методов в социологических исследованиях;

– использование больших объемов данных может сформировать новые риски, связанные с конфиденциальностью, этикой, безопасностью и т.п. Как следствие, необходимо создавать условия для использования цифровых технологий, с учетом возможных последствий;

– на рынке труда все чаще будет предъявляться новые требования к цифровым компетенциям социологов. К ним можно отнести: знание что такое большие данные и умение их извлекать; умение собирать, обрабатывать и визуализировать большие объемы данных;

владеть компьютерными программами и языками программирования, например SPSS, R или Python.

Литература

1. Волков В. В., Скугаревский Д. А., Титаев К. Д. Проблемы и перспективы исследований на основе Big Data (на примере социологии права) // Социологические исследования. 2016. № 1. С. 48–58.
2. Всемирный банк. Электронный ресурс. Режим доступа <https://data.worldbank.org/indicator/IT.NET.USER.ZS>. Дата обращения 12.04.2022
3. ВЦИОМ. Россияне стали реже участвовать в опросах. Электронный ресурс. Режим доступа: <https://wciom.ru/sobytie/rossijane-stali-rezhe-uchastvovat-v-oprosakh-pozhalovalsja-gendirektor-wciom>. Дата обращения: 12.04.2022
4. Докторов Б. Из XVII столетия в наступивший век: к становлению постгэллаповских опросных технологий // Телескоп. 2003. № 2.
5. Дюркгейм, Эмиль. [1897] 2006. О самоубийстве. Лондон: Пингвин.
6. Николаенко Г. А. Перспективы использования цифровых следов исследователей для анализа их коммуникативных стратегий (на примере социальной сети Research Gate) // Социология науки и технологий. 2019. Т. 10. № 2. С. 93–109.
7. Одинцов А. В. Открытость баз данных как условие формирования «больших данных» в социологии // Научно-методический электронный журнал «Концепт». – 2017. – № 12 (декабрь). – 0,4 п. л. Электронный ресурс. Режим доступа: <http://e-koncept.ru/2017/173020.htm>. Дата обращения 17.05.2022
8. Одинцов А. В. Социология общественного мнения и вызовы Big Data // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 3. С. 30–43.
9. Организация объединённых наций. Электронный ресурс. Режим доступа: <https://www.un.org/ru/global-issues/big-data-for-sustainable-development/> Дата обращения: 12.04.2022
10. Потемкин, В. К. Цифровая технологическая экспансия в системе управления персоналом / В. К. Потемкин // Управление персоналом. Социальные измерения эффективности и качества труда. Вызовы цифровой экономики : сборник научных трудов. – Санкт-Петербург : Санкт-Петербургский государственный экономический университет, 2020. – С. 5-10.
11. Резаев А.В., Трегубова Н.Д. «Искусственный интеллект», «онлайн-культура», «искусственная социальность»: определение понятий // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 6.
12. *Социодиггер. 2020. Октябрь. Том 1. Выпуск 3: Цифровизация и искусственный интеллект.* Электронный ресурс. Режим доступа: https://wciom.ru/fileadmin/file/nauka/podborka/wciom_sociodigger_102020.pdf (дата обращения 17.06.2022)
13. Татарова Г. Г. Методологическая травма социолога. К вопросу интеграции знания // Социологические исследования. 2006. № 9. С. 3–12.
14. Тенденции развития интернета в России: аналитический доклад / Г.И. Абдрахманова, Н. В. Бондаренко, К. О. Вишневский, Л. М. Гохберг и др.; Координационный центр национального домена сети Интернет, Нац. исслед. ун-т «Высшая школа экономики». – М.: НИУ ВШЭ, 2018. – 184 с
15. Толстова, Ю. Н. Социология и компьютерные технологии / Ю. Н. Толстова // Социологические исследования. – 2015. – № 8(376). – С. 3-13.
16. Щекотин Е.В. Цифровые технологии в социальных науках: предмет и метод цифровой социологии. Социология и право. 2020;(1):49-59. Электронный ресурс. Режим доступа <https://doi.org/10.35854/2219-6242-2020-1-49-59> Дата обращения: 12.04.2022
17. Alter A. Your E-book Is Reading You // WSJ. 2012. June 29.
18. Chen M., Mao S., Zhang Y., Leung V. C. Big Data. Related Technologies, Challenges, and Future Prospects. Springer, 2014. 100 p.
19. Cioffi-Revilla C. Computational Social Science // Wiley Interdisciplinary Reviews: Computational Statistics. 2010. Vol. 2. No. 3. P. 259–271.
20. Du Bois, W.E.B. 1906. The Health and Physique of the Negro American. Atlanta, GA: Atlanta University Press.
21. Eagle N. Big Data, Global Development and Complex Systems // Santa Fe Institute. 2010. 5 May.
22. Golder, Scott A. Diurnal and Seasonal Mood Vary with Work, Sleep and Day length Across Diverse Cultures // Science. Vol. 333, no. 6051. September 30. 2011. p. 1878–1881.
23. King G. (2009). The Changing Evidence Base of Social Science Research // King G., Scholzman K., Nie N. (eds.). The Future of Political Science: 100 Perspectives. New York: Routledge. P. 91–93.
24. Schwartz, R. D. (1989). Artificial intelligence as a sociological phenomenon. Canadian Journal of Sociology, 14(2), 179–202.
25. Woolgar, S. (1985). Why not a sociology of machines? The case of sociology and artificial intelligence. Sociology, 19(4), 557–572.