

Методология познания цифрового общества

УДК 303.4 DOI 10.26425/2658-347X-2020-2-17-26

Получено 21.05.2020 Одобрено 26.06.2020 Опубликовано 03.07.2020

Мещерякова Наталия Николаевна

Д-р социол. наук, профессор, ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», г. Томск, Российская Федерация

ORCID: 0000-0001-7658-7993

E-mail: natalia.tib@mail.ru

АННОТАЦИЯ

Цифровая социология – вычислительная социальная наука, использующая современные информационные системы и технологии, – уже сложилась. Однако до сих пор не разрешен конфликт с традиционной социологией и ее методами исследований. Этот конфликт можно преодолеть, если помнить, что есть общая цель – познание явлений и процессов общественной жизни, – первичная по отношению к методам, о которых предстоит договориться. Цифровая трансформация социологии необходима, поскольку традиционные социологические методы не решают задачи предоставления объемных, надежных эмпирических данных качественно и в короткие сроки; востребован переход от контактных методов исследований к бесконтактным.

Адаптация четырех современных информационных технологий: облачные вычисления, большие данные, интернет вещей и искусственный интеллект, – для целей социологии обеспечивает качественный переход в методологии познания цифрового общества. Облачные вычисления предоставляют

исследователям инструменты, большие данные – сырье, интернет вещей – технологию, направленную на сбор показателей (получение сигналов) в большом объеме, реальном времени, в качестве прямых, а не опосредованных свидетельств человеческого поведения. Развитие технологии «искусственный интеллект» расширяет возможности получения обработанных сигналов качества работы социальной системы без построения предварительной гипотезы, в короткие сроки и на большом объеме обработанных данных.

Цифровая трансформация социологии не означает отказ от использования традиционных методов социологического анализа, но предполагает расширение компетенций социолога, что требует пересмотра учебных программ вузов. Вместе с тем совмещение функций эксперта по предмету (социолога) и аналитика данных в одном специалисте оценивается как неперспективное, предлагается коллаборация их профессиональных компетенций в работе над едиными исследовательскими проектами.

Ключевые слова

Большие данные, интернет вещей, искусственный интеллект, неинвазивные методы, облачные вычисления, специалист по данным, транзакционные данные, цифровая социология, цифровая трансформация, эксперт по предмету.

Цитирование

Мещерякова Н.Н. Методология познания цифрового общества // Цифровая социология. 2020. Т. 3. № 2. С. 17–26.



Methodology for cognition of digital society

DOI 10.26425/2658-347X-2020-2-17-26

Received 21.05.2020

Approved 26.06.2020

Published 03.07.2020

Meshcheryakova Nataliya

Doctor of Sociological Sciences, Professor, National Research Tomsk Polytechnic University, Tomsk, Russia

ORCID: 0000-0001-7658-7993

E-mail: natalia.tib@mail.ru

ABSTRACT

Digital sociology is a computational social science that uses modern information systems and technologies, has already formed. But the conflict with traditional sociology and its research methods has not yet been resolved. This conflict can be overcome if we remember that there is a common goal – the knowledge of the phenomena and processes of social life, which is primary in relation to the methods to be agreed upon. Digital transformation of sociology is essential, since 1) traditional sociological methods do not solve the problem of providing voluminous, reliable empirical data qualitatively and in a short time; 2) the transition from contact research methods to unobtrusive ones is in demand.

The adaptation of four modern information technologies—cloud computing, big data, the Internet of things and artificial intelligence – for the purposes of sociology provides a qualitative transition in the methodology of knowledge of the digital society.

Cloud computing provides researchers with tools, big data – research materials, Internet of things technology aimed at collecting indicators (receiving signals) in large volume, in real time, as direct, not indirect evidence of human behavior. The development of “artificial intelligence” technology expands the possibility of receiving processed signals of the quality of the social system without building a preliminary hypothesis, in a short time and on a large volume of processed data.

Digital transformation of sociology does not mean abandoning the use of traditional methods of sociological analysis, but it involves expanding the competence of a sociologist, which requires a revision of University curricula. At the same time, combining the functions of an expert on the subject (sociologist) and data analyst in one specialist is assessed as unpromising, it is proposed to combine their professional competencies in working on unified research projects.

Keywords

Artificial intelligence, big data, cloud computing, data specialist, domain specialist, digital sociology, digital transformation, Internet of things, transactional data, unobtrusive methods.

For citation

Meshcheryakova N.N. (2020) Methodology for cognition of digital society. *Digital sociology*. Vol. 3, no 2, pp. 17–26. DOI: 10.26425/2658-347X-2020-2-17-26



ВВЕДЕНИЕ

Цифровая социология или вычислительная социальная наука, не в споре о более точном названии суть этой статьи, но эта наука уже родилась [Lazer et al., 2009]. И речь не только о применении цифровых технологий в социологии, но и о том, что используются они для познания нового типа общества – цифрового общества; новой социальной реальности – современных методах познания.

Если обратиться к гегелевской категории триады, то мы у порога выработки понятия, которое объединяет и выражает внутреннее единство двух противоположностей: традиционной социологии, опирающейся прежде всего на опросные методы исследований и вычислительной социологии, применяющей программные инструменты обработки и анализа больших данных. То что отношения между двумя этими направлениями развития социологии непростые, подробно и обоснованно раскрывает в своих работах Е.Ю. Журавлева [Журавлева, 2015]. У нее упоминаются и перемещение социальных исследований из академических институтов в корпоративные лаборатории, и то, что аналитики больших данных подменяют собой зачастую экспертов-социологов, и противопоставление «естественных» данных, полученных на основе анализа поведения в социальных сетях, якобы «искусственным», добытым путем опросов и наблюдений [Журавлева, 2015].

Кажется, что эти подходы не совместимы, и их конкурентные отношения Р. Берроуз и М. Севидж еще в 2007 г. назвали «назревающим кризисом эмпирической социологии» [Savage, Burrows, 2007]. С этой публикации прошло десятилетие и, по словам тех же авторов, конфликт до сих пор не разрешен и консенсус не найден.

В чем упрекают социологи сторонников новых методов¹ и какие эвристические возможности они упускают, М. Бэрроуз и Р. Севидж показывают на примере проекта «Большое британское обследование классов» (англ. The Great British Class Survey; GBCS, ББОК) [Бэрроуз, Севидж, 2016]. Это гибридный проект, в котором нашлось место традиционным опросным методикам и сетевой платформе, предоставленной Би-Би-Си, на которой проходил опрос респондентов в интерактивном формате. Какие преимущества для исследователей дало привлечение дигитальных данных? Во-первых, масштабность – ни одна традиционная выборка не предполагает такое количество единиц наблюдения. Во-вторых, исследование стало аттрактором для притяжения и самоопределения некой социальной группы (образованные и зажиточные), получившей новое «зеркало», в котором узрела

¹ Какие возможности раскрывают перед дисциплинарным проектом социология большие данные и методы их анализа полно характеризует в своих работах К. Губа [Губа, 2018].

самое себя и задумалась о самоопределении. В-третьих, интерактивное исследование в большей степени оказывает влияние на изучаемый объект, провоцируя феномены, которые в свою очередь требуют изучения: реакция респондентов на первые результаты и новая волна заполнения анкет, критические отклики в твитах и прочие перспективы использования обратной связи, в том числе привлечение техники краудсорсинга. Проект стал уходить в длительную незавершенную перспективу [Бэрроуз, Севидж, 2016].

За что же критикуется это исследование ортодоксальными социологами? Прежде всего, за неверифицированные методы сбора данных и нерепрезентативность. На стороне критиков сила авторитета классической науки, фундаментальность теоретического мышления, умение увидеть слабости нарождающихся методов [Бэрроуз, Севидж, 2016]. Но, возможно, в ходе этой обоснованной критики вместе с грязной водой выплескивают и ребенка, не позволяя проникнуть духу обновления в классическую науку, оставляя перспективу расширения исследовательских возможностей в руках необществоведов.

В свою очередь, аналитики данных, опьяненные перспективами, которые открывает перед ними использование алгоритмов машинного анализа больших данных, но не имеющие фундаментальной теоретической подготовки, которая есть у классических социологов, не получая их экспертной поддержки, усматривают причинно-следственные связи там, где возможна только статистическая корреляция. Некоторое время назад большой интерес вызывало исследование, устанавливающее взаимосвязь между запросами о симптомах и лечении гриппа и его сезонной эпидемией [Ginsberg et al., 2009]. Отрезвление наступило в 2013 г., когда Google Flu Trends (GFT) ошибся с определением пика эпидемии на 140 %². Причина ошибки – непрозрачная методика анализа данных, отсутствие тематической модели в ее основе. Построение корреляций между случайными величинами без экспертного анализа их возможных отношений может привести к ошибке индейки, которая вплоть до Дня благодарения и последовавшего разочарования была уверена, что хозяин ее любит, поскольку холит и лелеет³.

Причина отсутствия до сих пор консенсуса между классическими социологами и аналитиками данных не в том, что он невозможен, а в том, что его не ищут, исходя из глубокого недоверия друг к другу. Между тем цифровизация социологии сегодня – это даже не свободный выбор, а осознанная необходимость. Это

² Lazer D., Kennedy R. (2015). What we can learn from the epic failure of Google flu trends // Wired. Режим доступа: <https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/> (дата обращения: 15.05.2020)

³ Н. Талеб использовал эту аналогию в своей книге «Черный лебедь. Под знаком непредсказуемости» для описания события, которое невозможно предсказать.

я попытаюсь доказать в своей статье. А также последовательно раскрою, как каждая из четырех ведущих современных технологий: облачные вычисления, большие данные, интернет вещей и искусственный интеллект, – находят или скорее могут находить себя в социологии, расширяя ее познавательные возможности.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Доказательство неизбежности цифровизации социологии будет строиться вокруг двух постулатов. Во-первых, общество изменилось, и опросные методы исследований, разработанные для индустриального общества, перестали быть столь же надежными в информационном обществе. Во-вторых, общество изменилось, и теперь и востребован переход от контактных методов исследований к бесконтактным.

Необходимо переходить к новым методам и технологиям постижения общества. И эта мера – не только сулящая потенциальные возможности, но и вынужденная, поскольку традиционные методы неэффективно работают. До сих пор в социологии сохраняется «священное заклинание» – репрезентативность. Если выборочная совокупность обследуемых элементов не репрезентативна по отношению к генеральной совокупности, об объективности полученных результатов мы говорить не можем, а следовательно, и на научность претендовать тоже. Все мониторинговые основания оценок той или иной деятельности органов власти в Российской Федерации построены на репрезентативных исследованиях. Но в реальности полной репрезентативности социология обеспечить никогда не могла. Всегда были члены генеральной совокупности респондентов, до которых интервьюер добраться не мог. Поэтому изначально определялась доверительная вероятность.

Советская социология, неразрывно связанная с именем В.А. Ядова, находилась в тесных отношениях с социальной психологией, учитывала особенности личности респондента⁴. Но нельзя не признать, что диспозиция личностных установок современного социального актора изменилась по сравнению с установками респондентов, для которых разрабатывались традиционные методы массовых опросов. Эти методы соответствуют индустриальному обществу, в котором большинство населения уже грамотное, но не считает возможным отказать в ответах человеку с верительными грамотами, даже если у него лишь статус члена университетской исследовательской группы. Сегодня такой принудительной силой еще обладают участковые, маячащие за спиной переписчиков. Но к каждому опросу их не приставишь. Если пол века назад социологи могли опросить всех,

⁴ Например, в рамках диспозиционной теории, иерархизирующей личностные установки [Ядов и др., 2013]

до кого могли добраться, а добраться они фактически не могли до тех, кто за колючей проволокой и в казармах, то сегодня они опрашивают только тех, кто соглашается дать ответы. А отказываются массово. Где модель социотипа того, кто идет на контакт с интервьюером? И можно ли исследование, построенное только на опросе тех, кто согласен отвечать, считать репрезентативным?

Некоторые исследователи полагают, что компромиссным вариантом может быть технология САТИ (англ. Computer Assisted Telephone Interview – система компьютеризированного телефонного интервью) [Рогозин, Ипатов, Галиева, 2018]. Но в телефонных интервью также наблюдаются массовые отказы от ответов, плюс добавляются ограничения по формулировкам вопросов и вариантов закрытий, воспринимаемых только на слух.

Еще одно соображение по поводу того, что прежнее понимание репрезентативности не актуально. Если оставить в стороне уже упомянутую перепись и регулярное обследование домохозяйств, которое проводит Росстат, иные мониторинговые исследования становятся все более специфичными, где не все население выступает носителями искомых признаков. Всегда ли объективно мнение среднего человека по вопросам коррумпированности правительства, органов государственной безопасности и прочих структур, к которым он не имеет никакого непосредственного отношения? Не является ли их мнение проекцией позиций средств массовой информации и отдельных лидеров мнений? Или насколько взвешенным, основанным на полной информации является точка зрения всего населения по вопросам реализации государственных программ на территории регионов: по управлению государственным имуществом, развитием и использованием минерально-сырьевой базы территории и т.п.? Между тем на эти реальные исследования тратятся миллионы бюджетных денег. Господствующие до сих пор подходы к мониторинговым мероприятиям, результаты которых потом ложатся в основу принятия управленческих решений, изжили себя. Они неэффективны и чрезвычайно затратны.

Предлагаемая статья написана весной 2020 г., в эпоху становления «новой нормальности»; грядут перемены, которые мы даже не можем пока предугадать в полном объеме. Но об отдельных частностях мы говорить можем. Как уходят в прошлое рукопожатия при встрече из-за новых норм гигиены человеческих взаимодействий, так контактные формы опроса (подомовые опросы, интервью face-to-face (рус. личные интервью) могут оказаться невостребованными. Если до сих пор перед интервьюером захлопывали дверь из-за нежелания тратить на опрос свое личное время, то теперь отказ может объясняться соображениями безопасности. Конечно, вирус уйдет, но исчезнут ли с этой угрозой новые формы отношений,

способов коммуникации, норм гигиены и безопасности? Бесконтактные или неотрузивные [Бэрроуз, Севидж, 2016] методы исследований⁵ актуализируются. И под ними подразумевается не простое перенесение опросных листов на электронные платформы с автоматическим подсчетом средних величин.

Мы живем в эпоху цифровой трансформации. Т. Сибел в своей книге «Digital Transformation: Survive and Thrive in an Era of Mass Extinction» [Siebl, 2019], обращает внимание на то, что цифровая трансформация – это не просто серия поколенческих изменений в информационных технологиях или перенос данных организации на цифровую платформу. Предтечи цифровой трансформации – цифровизация и сеть «Интернет» (далее – Интернет), но это еще не цифровая трансформация как таковая. Влияние Интернета и предшествовавшей ему волны цифровизации заключалось прежде всего в оцифровке существующих компетенций. Оцифровка – это использование цифровых инструментов для автоматизации и улучшения существующего способа работы, не изменяя его фундаментально. Автор считает, что мы на пороге переломного момента, когда облачные вычисления, большие данные, интернет вещей и искусственный интеллект сошлись, чтобы управлять сетевыми эффектами и создавать экспоненциальные изменения [Siebl, 2019].

Эти экспоненциальные изменения, по мнению профессоров Массачусетского технологического института Э. Бриньольфсона и Э. Макафи [Brynjolfsson, McAfee, 2016] связаны с тем, что компьютеры, долгое время хорошо следовавшие инструкциям, теперь способны учиться. Эта способность будет иметь фундаментальные последствия для мира.

До сих пор и в социологии цифровизация и Интернет означали оцифровку данных о социальных фактах, что облегчало и ускоряло их обработку, хранение, использование и передачу, и перенос опросных листов на электронные платформы. Публичное облако дает возможности краудсорсинга⁶, что особенно важно для российской науки, как отмечают С.Е. Егеров и С.А. Захарова [Егеров, Захарова, 2015], в силу постоянного дефицита в ней исследовательских и инженерных кадров. Но большие данные и искусственный интеллект сулят и качественно новые возможности.

⁵ Неотрузивные – простая транскрипция англоязычного термина unobtrusive, который традиционно подразумевал методы исследований, исключаящие контакт исследователя с объектом исследования дабы избежать искажающего влияния на его поведение и мнение. К таким методам относится наблюдение с закрытой позицией наблюдателя, контент-анализ и многие другие. Современные коммуникационные технологии и большие данные качественно расширяют возможности неотрузивных методов.

⁶ Краудсорсинг исследовательской деятельности – совместное выполнение проектов исследователями или экспертами, использующими коммуникационные технологии для удаленной работы вне рамок традиционных лабораторий.

Первое и очевидное преимущество больших данных – возможность поиска неизвестных корреляций. Корреляции требуют разработок моделей отношений между переменными. И такие модели создаются. Например, Е.В. Бродовская и А.Ю. Домбровская [2018], отмечая корреляционность больших данных, конструируют цифровые маркеры политических процессов с применением онлайн-сервисов для мониторинга социальных медиа. Расширяются возможности проведения контент-анализа, который по словам Б.А. Сусларова и Е.С. Кундашевой [2018], является типичным примером прикладного информационного анализа текста. По сути, это тот же метод, которым Г. Лассуэл препарирует газету «Истинный американец», но оснащенный прикладным программным обеспечением и используемый на больших данных⁷. Даже в таком виде обнаружение значимых корреляций традиционными для социологии методами математического моделирования социальных явлений и процессов дает большие возможности. Если мы видим связь одного параметра с другим, то мы как минимум можем попытаться повлиять на один из параметров и пронаблюдать изменение другого, либо попытаться сломать корреляцию. Но даже, если мы не имеем четкой модели отношений между факторами, методы машинного обучения, основанные на индуктивном обучении, позволяют обнаружить эмпирические закономерности в данных. С развитием этой технологии мы ожидаем, что компьютер обучится делать что-то, что человек делать не умеет. И мы должны быть готовы грамотно поставить перед ним задачу.

Цифровизация, как оцифровка данных, и всемирная система объединенных компьютерных сетей – предтечи цифровой трансформации. Облачные вычисления (инструменты) и большие данные (сырье) – это накопление количественных изменений, интернет вещей и искусственный интеллект – переход в новое качество, согласно законам диалектики. Попасть в русло этого процесса – вот наша задача. Выиграют те, кто использует цифровые технологии и передовую аналитику для повышения ценности конечных результатов.

Значение феномена больших данных заключается не столько в размере набора данных, к которому мы обращаемся, сколько в его полноте и отсутствии ошибки выборки. С вычислительной и накопительной мощностями, обычно доступными сегодня, мы можем получить доступ, хранить и обрабатывать весь набор данных, связанных с рассматриваемой проблемой. В идеале, когда набор данных достаточно полон, чтобы мы могли обрабатывать его весь, это меняет все в вычислительной парадигме, позволяя нам решать

⁷ Не хотелось бы здесь спорить, что можно считать большими данными, а что под это определение не попадает, одно бесспорно, что эмпирическая база подобных исследований неуклонно возрастает.

большой класс проблем, которые ранее были неразрешимы, в том числе разработать механизмы прогнозирования развития того или иного социального процесса. Без понимания того, что есть большие данные для нас, социологов, и как они увеличиваются, мы не найдем применения современным технологиям.

Возможность фиксировать, хранить, обрабатывать и анализировать данные любого размера, скорости и формы закладывает основу для широкого внедрения и применения методов компьютерной обработки информации, в том числе алгоритмов машинного обучения. Данные, генерируемые повсеместно в организации, могут иметь огромную ценность для социологов. Для их обозначения можно использовать термин «транзакционные» [Бэрроуз, Севидж, 2016]. Это данные больших коммерческих и государственных баз данных. Название связано с тем, что они являются побочным продуктом обычных транзакций⁸ между гражданами, организациями как коммерческими, так и государственными и муниципальными. Понятие более широкое, чем имеющее хождение в отечественной статистике «административные данные», поскольку это не только используемая при формировании официальной статистики документированная информация, но и обращения граждан в различные организации через систему предусмотренной обратной связи, ответы, комментарии и др. Общее в этих данных их дигитальность, то есть они сохранены в цифровом формате.

Договориться, что есть большие данные в социологии, классифицировать их, научиться собирать на единой масштабируемой цифровой платформе, создать программы для анализа возможных ошибок, связанных с вводом данных, – необходимый, но только подготовительный этап грядущей цифровой трансформации. Какое место занимают в ней интернет вещей и искусственный интеллект? На первый взгляд может показаться, что технология интернета вещей притягивается в социологию «за уши». Докажем, что она так же необходима для развития дисциплины, как и в других сферах деятельности.

Основная идея технологии интернета вещей заключается в том, чтобы подключить любое устройство, оснащенное соответствующими возможностями обработки и связи, к Интернету, чтобы оно могло отправлять и получать сигналы практически в режиме реального времени. Развитие этой технологии расширяет объем больших данных. Как выглядит процесс и результат этого расширения в производственной сфере очевидно: сенсоры размещаются на любом участке предмета (подключенный объект) и передают показания об объекте, например, значение напряжения в электрической сети, так что все устройства в цепочках

⁸ Здесь под транзакцией имеется в виду разовое законченное действие.

становятся удаленно управляемыми машинами в реальном или почти реальном времени. Такие датчики создают возможности мониторинга состояния объекта, определение его местоположения, производительности и прочее. Как создать аналогичные технологические возможности поступления значимых для социальных исследований данных, и платформу, способную агрегировать, нормализовать и анализировать большие объемы разрозненных оперативных данных в реальном времени? Социологи не могут разместить датчики на своих объектах информации. Так как растут базы данных в нашем случае?

Это нас возвращает и к вопросу о том, что есть данные для сегодняшней социологии, и к вопросу их фиксации, агрегации и хранения для последующего использования. Это данные, которые мы назвали транзакционными, и так называемый «цифровой выхлоп» [Журавлева, 2015] – цифровой побочный результат применения коммуникационных технологий. Эти неотгузивные данные по словам Р. Бэрроуза и М. Севиджа [2016] бросают вызов традиционной чувствительности социологов, поскольку опросные данные – это всегда дискурсивные отчеты о некоторых предшествующих действиях, которые не всегда совместимы с отчетами цифрового отслеживания. Проще говоря, то, что мы говорим, куда мы ходили и с кем встречались не всегда совпадает с тем, что показывает прикладная программа на нашем смартфоне.

Мы все отдаем себе отчет, что проблема упирается в вопрос конфиденциальности личных данных. Чрезвычайная ситуация весны 2020 г. уже привела к тому, что этот вопрос отодвинули на задний план, применив технологию распознавания лиц фактически за слежки за населением, нарушающим условия самоизоляции. Но решать проблему надо не простым игнорированием рисков ограничения личных свобод или отказом использования подобных технологий, а принципиально, путем обсуждений, юридических экспертиз и с учетом мнения общественности. Безопасность информации, приватность и конфиденциальность будут становиться все более важными. Как показывает практика мы достаточно легко жертвуем ими в угоду обеспечения физической безопасности и здоровья, но почему бы ни поставить вопрос безопасности информации как первый принцип дизайнера любого прикладного программного продукта, не отказываясь от самой идеи расширения наших знаний о социуме через сбор и научно-обоснованный анализ дигитальных данных?

Социология всегда была основана на сборе данных. Это основной и самый трудоемкий этап работы, отнимающий много времени. Исследователи вынужденно обращались к статистической информации, полученной на небольших выборках, чтобы сделать выводы о всей совокупности. Из-за небольших размеров выборки и погрешностей при сборе данных

приходилось тратить значительные усилия и время на ремонт выборки, взвешивание, чтобы устранить погрешности, которые потенциально могли исказить результаты анализа. Сегодня эластичное облако предоставляет неограниченные вычислительные и запоминающие возможности. Появляется программное обеспечение, предназначенное для обработки данных в больших масштабах. Больше нет необходимости в выборке или в кураторстве данных⁹. Есть возможности делать выводы почти в реальном времени на основе всей доступной эмпирики. Эта возможность является основополагающей для последних достижений в области искусственного интеллекта. Ее обеспечивают алгоритмы машинного обучения (подмножество алгоритмов искусственного интеллекта), глубокого обучения (подмножество машинного обучения, использование искусственных нейронных сетей с большим количеством нейронов, слоев и взаимосвязей). Одна из перспектив применения алгоритмов искусственного интеллекта связана с тем, что отпадает необходимость в экспертной гипотезе. Они могут делать выводы о поведении сложных систем непосредственно из данных, генерируемых этими системами.

Машинное обучение широкое подмножество алгоритмов искусственного интеллекта, которые учатся на примерах и опыте (представленных наборами входных/выходных данных), а не полагаются на жестко закодированные и предопределенные правила, характеризующие традиционные алгоритмы¹⁰ [Pyle, San José, 2015]. Это называется индуктивное обучение или обучение по прецедентам. В отличие от точно определенных правил традиционных алгоритмов, алгоритмы машинного обучения математически анализируют любое разнообразие данных (изображения, текст, звуки, временные ряды и т.д.) и их взаимосвязи для того, чтобы делать выводы. Алгоритмы искусственного интеллекта используют другой подход, чем традиционные алгоритмы, основанные на логике. Они основаны на идее, что вместо того, чтобы кодировать компьютерную программу для выполнения задачи, проектируйте программу, чтобы учиться непосредственно из данных. Это сулит перспективы развитию

⁹ Здесь присутствует большое допущение. Как доказывают в своей статье Л.А. Василенко и В.В. Зотов [2020], электронный документооборот, который является потенциальным источником транзакционных данных из государственных и муниципальных управленческих структур, пока неэффективен. Это связано с частым изменением состава показателей в отчетных формах, стремлением вышестоящих инстанций решить проблему ввода объемных первичных данных за счет нижестоящих иерархических уровней, несовместимостью информационных ресурсов, нехваткой необходимых цифровых компетенций у государственных и муниципальных служащих и пр. Эти недостатки снижают качество эмпирики.

¹⁰ Pyle D., San José C. (2015). An executive's guide to machine learning // McKinsey Quarterly. Режим доступа: <https://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning> (дата обращения: 15.05.2020).

и социологического знания. При этом никто не отрицает дедуктивное обучение, формализующее знания экспертов. Сочетании того и другого сулит большие перспективы. Например, социолог может использовать неконтролируемые алгоритмы для кластерного анализа данных о группах населения с более-менее устойчивыми паттернами поведения в вопросах здоровьесбережения, мобильности, приверженности источникам информации и пр. Глубокое обучение интересно именно тем, что оно потенциально может быть применено к любой задаче – от прогнозирования электорального поведения до выявления коррупционных схем – с гораздо меньшим вмешательством исследователя по сравнению с другими методами машинного обучения из-за значительно уменьшенной или исключенной потребности в разработке характеристик. Но стоит ли полагаться только на это и превращать социолога только в интерпретатора? С этим также необходимо разобраться.

Интеллектуальные системы сулят нам невероятные возможности, область применения искусственного интеллекта постоянно расширяется. Ряд факторов ускоряют этот процесс: постоянное снижение стоимости вычислений и хранения данных, а также совершенствование аппаратного обеспечения и инновации. По мере того, как вычислительная техника становится все более мощной и менее дорогостоящей, искусственный интеллект можно применять ко все большему и более разнообразным наборам данных для решения все большего числа задач и принятия решений в реальном времени. Социологам необходимо освоить прогнозную аналитику и алгоритмы искусственного интеллекта для принятия правильных управленческих решений. Последняя обозначенная задача ставит перед нами вопрос о достаточном наборе компетенций современного социолога.

В силу специфики образования сами социологи сегодня не способны самостоятельно обеспечить цифровую трансформацию методологии социологических исследований. Вопрос состоит в том, нужно ли принципиально пересмотреть квалификационные требования социологов, чтобы максимально приблизить их компетенции к аналитикам данных, или пойти путем коллаборации двух видов деятельности? Сразу несколько соображений, почему я сторонник второго пути. Невозможно определить, какими инструментами должен владеть социолог, чтобы работать с большими наборами данных и выполнять прогнозную аналитику с использованием единого образа данных, поскольку сами эти инструменты постоянно трансформируются. Возможно, следует пойти по пути кооперации компетенций специалиста по предмету и специалиста по данным. Но, чтобы специалист по данным разработал алгоритм машинного обучения для поиска маркеров того или иного социального явления или процесса, тем более

для того, чтобы сделать последующий анализ полученных результатов, необходим социолог с его компетенцией моделирования отношений между переменными.

Необходимо тесное сотрудничество специалистов по данным и экспертов по предмету¹¹, чтобы определить существенные данные, которые влияют на результат (например, неэффективность конкретной программы преобразований). Алгоритм машинного обучения обучается путем перебора тысяч (или миллионов) исторических случаев, корректируя относительную важность (вес) каждого из признаков до тех пор, пока он не сможет сделать вывод о причинах неудовлетворенности населения как можно точнее. Результатом работы алгоритма машинного обучения становится набор весов, которые могут быть использованы для вывода. И в то время как алгоритм определяет веса, аналитики определяют особенности. Тем не менее пересмотр квалификационных требований при подготовке социологов в цифровом обществе кажется необходимым.

ЗАКЛЮЧЕНИЕ

В начале этой статьи было заявлено, что консенсус между традиционными методами социологии и современными, основанными не просто на компьютерной обработке больших данных, но применении алгоритмов искусственного интеллекта, возможен. Главный аргумент здесь один: цель первична, методы вторичны. Если помнить, что социология – наука, направленная на познание явлений и процессов общественной жизни и создание управленческих проектов, опирающихся на полученное знание, то о методах мы договоримся.

Все в современной социальной реальности, даже ситуация с пандемией COVID-2019, говорит нам, что что просто следовать тенденциям социальных изменений недостаточно для того, чтобы выжить

¹¹ Деление специалистов на data scientists (аналитики данных) и domain experts (специалисты в предметной области) подсмотрено у Т. Сибела [Siebl, 2019]. Кроме них для успешной цифровой трансформации деятельности он считает необходимым привлекать professional services providers (поставщики профессиональных услуг, которые обеспечивают облачные хранилища, цифровые платформы, прикладное программное обеспечение и пр.).

и в прямом физическом и широком экзистенциальном смысле. Социологии, как и любой другой отрасли необходимо пройти фундаментальную цифровую трансформацию, чтобы быть успешной в своем деле. Цифровая трансформация будет иметь глубокие последствия, и не все из них мы можем сейчас предсказать. Но без принятия рисков и поощрения постоянных инноваций, не стоит даже начинать. Необходимо решить следующие задачи: создание единой федеральной базы социальных данных, обеспечение безопасности сохраняемых данных, выбор платформы и определение модели информационно-технологической инфраструктуры, разработка, установка, тестирование прикладного программного обеспечения, развертывание моделей машинного обучения и ряд других задач. Также необходимо помнить, что все методы, используемые сегодня, устареют через 5-10 лет. Архитектура должна быть масштабируемой и обеспечивать возможность замены любых компонентов их продвинутыми аналогами следующих поколений, а также включать любые новые инновации в области программного обеспечения. Это не значит, что следует отказываться от традиционных, наработанных социологией методик. Там, где объем генеральной совокупности респондентов невелик и они доступны: в лонгитюдных исследованиях, безусловно, в экспертных опросах, – сохраняется пространство для традиционных подходов. Но в исследованиях социально-экономической, тем более политической направленности, где особенно ярко проявляется спираль молчания и другие искажающие информацию феномены общественного мнения; в исследованиях мониторингового типа, где необходимо обеспечить возможность постоянного наблюдения за процессами для оценки их состояния и прогнозов развития, применение методик анализа больших данных расширяет познавательные возможности дисциплины.

Это страшно, менять что-то вроде работающее на нечто рискованное и недоказанное. Но в этом суть нашего исторического момента, этой «новой нормальности». Говоря словами Т. Сибела, «никогда ставки не были столь высоки, как в смысле риска исчезновения, так и возможностей получения выгоды» [Siebl, 2019, p. 76].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Берроуз Р., Севидж М.* (2016). После кризиса? Big data и методологические вызовы эмпирической социологии // Социологические исследования. № 3. С. 28–35.
- Бродовская Е.В., Домбровская А.Ю.* (2018). Большие данные в исследовании политических процессов: учебное пособие. М.: Изд-во Московского педагогического государственного университета. 88 с.
- Василенко Л.А., Зотов В.В.* (2020). Цифровизация публичного управления в России: риски, казусы, проблемы // Цифровая социология. № 2. С. 4–16.

- Губа К. (2018). Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. Т. 17. № 1. С. 213–236. doi: 10.17323/1728-192X-2018-1-213-236
- Егереv С.Е., Захарова С.А. (2015). Краудсорсинг в науке // Социологический альманах. № 6. С. 311–322.
- Журавлева Е.Ю. (2015). Социология в сетевой среде: к цифровым социальным исследованиям // Социологические исследования. № 8. С. 25–33.
- Рогозин Д.М., Ипатова А.А., Галиева Н.И. (2018). Стандартизированное (телефонное) интервью. М.: Пункт. 416 с.
- Сулаков Б.А., Кундышева Е.С. (2015). Математическое моделирование контекстных высказываний при социологических опросах // Вестник МНЭПУ. Т. 7. С. 318–327.
- Ядов В.А. и др. (2013). Саморегуляция и прогнозирование социального поведения личности: диспозиционная концепция. 2-е, расширенное изд. М.: ЦСПиМ. 376 с.
- Brynjolfsson E., McAfee A. (2016). *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. NY; London: W.W. Norton & Company. 336 p.
- Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S., Brilliant L. (2009). Detecting influenza epidemics using search engine query data // *Nature*. Vol. 457. No. 7232. Pp. 1012–1014.
- Lazer D., Pentland A., Adamic L., Aral S., Barabasi A-L., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D., Van Alstyne M. (2009). Computational social science // *Science*. Vol. 323. No. 5915. Pp. 721–723.
- Savage M., Burrows R. (2007). The coming crisis of empirical sociology // *Sociology*. No. 41 (5). Pp. 885–899.
- Siebl T. (2019). *Digital transformation: survive and thrive in an era of mass extinction*. NY: Rosetta Books. 256 p.

REFERENCES

- Brodovskaya E.V. and Dombrovskaya A.Yu. (2018), *Big data in the study of political processes [Bol'shie dannye v issledovanii politicheskikh protsessov]*, Izd-vo Moskovskogo pedagogicheskogo gosudarstvennogo universiteta, Moscow, Russia. [In Russian].
- Brynjolfsson E. and McAfee A. (2016), *The second machine age: work, progress, and prosperity in a time of brilliant technologies*, W.W. Norton & Company, New York, London.
- Burrows R. and Savage M. (2016), “After the crisis? Big data and the methodological challenges of empirical sociology” [“Posle krizisa? Big data i metodologicheskie vyzovy empiricheskoi sotsiologii”], *Sociological Studies [Sotsiologicheskie issledovaniya]*, no 3, pp. 28–35.
- Egerev S.E. and Zakharova S.A. (2015), “Crowdsourcing in science” [“Kraudsorsing v nauke”], *Sotsiologicheskii almanakh*, no. 6, pp. 311–322.
- Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S. and Brilliant L. (2009), Detecting influenza epidemics using search engine query data, *Nature*, vol. 457, no. 7232, pp. 1012–1014.
- Guba K. (2018), “Big data in sociology: new data, new sociology?” [“Bol'shie dannye v sotsiologii: novye dannye, novaya sotsiologiya?”], *Russian Sociological Review [Sotsiologicheskoe obozrenie]*, vol. 17, no. 1, pp. 213–236.
- Lazer D., Pentland A., Adamic L., Aral S., Barabasi A-L., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D. and Van Alstyne M. (2009), Computational social science, *Science*, vol. 323, no. 5915, pp. 721–723.
- Rogozin D.M., Ipatova A.A. and Galieva N.I. (2018), *Standardized (telephone) interview [Standartizirovanoe (telefonnoe) interv'yu]*, Punkt, Moscow, Russia. [In Russian].
- Savage M. and Burrows R. (2007), The coming crisis of empirical sociology, *Sociology*, no. 41 (5), pp. 885–899.
- Siebl T. (2019), *Digital transformation: survive and thrive in an era of mass extinction*, Rosetta Books, New York.
- Suslakov B.A. and Kundysheva E.S. (2015), “Mathematical modeling of context statements during conducting sociological poll” [“Matematicheskoe modelirovanie kontekstnykh vyskazyvaniy pri sotsiologicheskikh oprosakh”], *Vestnik MNEPU*, no. 7, pp. 318–327.
- Vasilenko L.A. and Zotov V.V. (2020), “Digitalization of public administration in Russia: risks, incidents, problems” [“Tsifrovizatsiya publichnogo upravleniya v Rossii: riski, kazusy, problemy”], *Digital Sociology [Tsifrovaya sotsiologiya]*, no. 2, pp. 4–16.
- Yadov V.A. et al. (2013), *Self-regulation and forecasting of a person's social behavior: a dispositional concept [Samoregulyatsiya i prognozirovaniye sotsial'nogo povedeniya lichnosti: dispozitsionnaya kontseptsiya]*, 2-e rasshirennoe izd. TsSPiM, Moscow, Russia. [In Russian].
- Zhuravleva E.Yu. (2015), “Sociology in digital environment: towards digital social research” [“Sotsiologiya v setevoi srede: k tsifrovym sotsial'nym issledovaniyam”], *Sociological Studies [Sotsiologicheskie issledovaniya]*, no. 8, pp. 25–33.

TRANSLATION OF FRONT REFERENCES

¹What opportunities big data and methods of its analysis reveal before the disciplinary project “sociology”, K. Guba [Guba, 2018] fully characterizes in her works.

²Lazer D. and Kennedy R. (2015), What we can learn from the epic failure of Google flu trends, *Wired*. Available at: <https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/> (accessed 15.05.2020).

³Nassim Taleb used this analogy in his book the “Black Swan. Under the sign of unpredictability” for description of the event that cannot be predicted.

⁴For example, within the framework of the dispositional theory that hierarchizes personal attitudes [Yadov et. al., 2013].

⁵Non-Intrusive – a simple transcription of the English-language term unobtrusive, which traditionally meant research methods that exclude contact between the researcher and the object of research in order to avoid distorting influence on his behavior and opinion. These methods include observation with a closed observer position, content analysis, and many others. Modern communication technologies and big data qualitatively expand the possibilities of non-Intrusive methods.

⁶Research crowdsourcing is the collaborative execution of projects by researchers or experts who use communication technologies to work remotely outside of traditional laboratories.

⁷I would not like to argue here what can be considered big data, and what does not fall under this definition, but there is no doubt that the empirical base of such research is steadily increasing.

⁸Here a transaction is a one time completed action

⁹There is a big assumption here. As L.A. Vasilenko and V.V. Zotov prove in their article [2020], electronic document management, which is a potential source of transaction data from state and municipal administration structures, is still ineffective. This is due to frequent changes in the composition of indicators in reporting forms, the desire of higher authorities to solve the problem of entering large primary data at the expense of lower hierarchical levels, incompatibility of information resources, lack of necessary digital competencies for state and municipal employees, and so on. These disadvantages reduce the quality of empirics.

¹⁰Pyle D. and San José C. (2015), An executive’s guide to machine learning, *McKinsey Quarterly*. Available at: <https://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning> (accessed 15.05.2020).

¹¹The division of specialists into data scientists (data analysts) and domain experts (specialists in the subject area) was seen in Thomas Siebel [Siebel, 2019]. In addition, for the successful digital transformation of activities, he considers it necessary to involve professional services providers (professional service providers that provide cloud storage, digital platforms, application software, etc.).