

ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.26794/2226-7867-2022-12-3-36-40
УДК 303(045)

Социологические исследования в цифровую эпоху: формирование базы знаний вычислительной социологии

К.Ф. Рафикова

РАНХиГС при Президенте РФ, Москва, Россия

АННОТАЦИЯ

Непрерывный рост больших массивов данных и достижения в области компьютерных наук способствуют развитию такой отрасли знания, как вычислительная социология. У исследований, выполненных с использованием вычислительных методов для анализа больших данных, имеется высокий потенциал. Однако для создания теоретической базы вычислительной социологии необходимо провести множество фундаментальных работ. В данной статье предлагается создать базу знаний для вычислительной социологии. В ней собраны исследования, выполненные с применением вычислительных и математических методов, а также связанные с моделированием, машинным обучением и анализом социальных сетей. Существует ряд ограничений для использования таких методов в социальных науках, однако они значительно расширяют исследовательское поле.

Ключевые слова: вычислительные социальные науки; вычислительная социология; большие данные; моделирование; машинное обучение; социальные сети; цифровые следы

Для цитирования: Рафикова К.Ф. Социологические исследования в цифровую эпоху: формирование базы знаний вычислительной социологии. *Гуманитарные науки. Вестник Финансового университета*. 2022;12(3):36-40. DOI: 10.26794/2226-7867-2022-12-3-36-40

ORIGINAL PAPER

Sociological Research in the Digital Age: Forming the Knowledge Base of Computational Sociology

K.F. Rafikova

RANEPA, Moscow, Russia

ABSTRACT

The continuous growth of big data and developments in computational sciences have contributed to the development of such an area of knowledge as computational sociology. Research performed using computational methods to analyse big data have great potential. However, a lot of fundamental work needs to be done to establish the theoretical basis of computational sociology. This paper aims to build a knowledge base for computational sociology. It brings together research done using computational and mathematical methods, as well as related to modelling, machine learning, and social networks analysis. There are several limitations of employing such methods in the social sciences, but they significantly expand the research field.

Keywords: computational social sciences; computational sociology; big data; simulation; machine learning; social networks; digital footprints

For citation: Rafikova K.F. Sociological research in the digital age: Forming the knowledge base of computational sociology. *Gumanitarnye Nauki. Vestnik Finansovogo Universiteta = Humanities and Social Sciences. Bulletin of the Financial University*. 2022;12(3):36-40. (In Russ.). DOI: 10.26794/2226-7867-2022-12-3-36-40

Основной целью вычислительной социологии является описание, объяснение, моделирование и прогнозирование сложных масштабных социальных взаимодействий, процессов и явлений с помощью достижений компьютерных наук. Это междисциплинарное поле, основанное на данных социальных наук, статистики, информатики и инженерии. Благодаря такой интеграции появляется возможность применять математические методы к большим источникам цифровых данных.

Вычислительная социология имеет несколько особенностей. Во-первых, при любом исследовании предполагается работа с большими массивами данных, которые обладают рядом специфических свойств. Описание этих свойств представлено ниже. Во-вторых, вычислительная социология отталкивается от идеи, что общество — сложная система с эмерджентными свойствами [1]. Кроме того, нельзя забывать о моделировании (в частности, агентном моделировании, построении искусственных обществ) как об одном из основных инструментов при работе в рамках вычислительной социологии.

Исследования с использованием вычислительных методов дают возможность зафиксировать взаимосвязи между изначально неструктурированными данными и сделать детализированный обзор. Но надо понимать, что выявленные закономерности часто требуют дополнительной проверки с привлечением других массивов данных или теоретических экспликаций.

Задача статьи — охарактеризовать, опираясь на последние исследования в области вычислительных социальных наук, где и каким образом в социальных науках применяются вычислительные (математические) методы.

Большие массивы данных, которые собираются и обобщаются сегодня, обладают своей спецификой, связанной с цифровой эпохой. Для начала важно выделить разные средства их сбора. Доступными для исследователей являются данные из социальных сетей. Это может быть фиксация активности пользователей, регистрация изменений, информация о частоте использования или цифровые следы в целом. Данные также могут собираться с помощью различных цифровых устройств, в том числе посредством системы видеонаблюдения.

Данные датчиков в различных предметах (носимые биометрические гаджеты, личные

автомобильные навигаторы) также являются нетривиальным источником информации [2, с. 53]. История посещения сайтов, использование мобильных приложений и регистрация пользовательских действий образуют еще одну группу средств сбора больших массивов данных [3].

Большие данные обладают рядом неочевидных свойств, что выделяет их на фоне других собираемых массивов. Их огромный объем обеспечен высокой скоростью производства. Данные накапливаются и непрерывно увеличиваются. При этом они остаются гибкими, т.е. несут в себе свойства расширяемости и масштабируемости. Такие данные часто дают исследователям возможность провести подробный анализ, так как сами они обладают высоким уровнем детализации и внутренним многообразием [3].

Проанализировав средства сбора и свойства больших данных, можно отметить, что научные исследования адаптируются к изменениям.

Применение технологий в рамках вычислительной социологии (в контексте разговора о социальных науках) происходит не так, как в других областях знания. Разрабатываются гибридные методологии, а вычислительные (математические) методы сочетаются с уже устоявшимися и традиционными [4]. Некоторые методы анализа разрабатываются специально для работы с большими данными.

Попробуем описать, как используются вычислительные методы. Для этого обратимся к современным исследованиям в области вычислительной социологии. С помощью анализа цитирования выявим журналы, темы публикаций, а также авторов, которые активно работают в этой сфере.

Для удобства сгруппируем следующие тематические блоки: область знания, связанная с моделированием (например, имитационное моделирование), применение методов машинного обучения (анализ естественного языка, компьютерное зрение, анализ вкраплений слов), и то, каким образом исследователи работают с данными, полученными из социальных сетей (цифровые следы, поисковые запросы). Исследования в данной отрасли довольно тесно связаны с прогнозированием, поэтому в конце обзора обозначим, какие практические выводы возможно сделать.

Моделирование можно рассматривать, опираясь на различные подходы, например формализацию классических и современных социологических теорий (верификация взаимодействия

теоретических категорий «габитус» П. Бурдьё и «структура» Э. Гидденса [5]). Построение таких имитационных моделей дает возможность описать достаточно сложные процессы и увидеть противоречия, недоказательность, погрешности проанализированных теорий. Не все теории и их положения поддаются формализации, но в любом случае это довольно продуктивное направление исследования [6].

Построение моделей и формализация могут быть также использованы при проверке теорий среднего уровня (например, стратификация [7], социальное управление [8], солидарность [9]). Проведение таких операций требует качественной подготовки в области математических и статистических методов анализа данных.

Агентное моделирование представляет собой процесс построения модели на основе индивидуальных решений агентов и является уточнением имитационной модели. Благодаря таким моделям можно оценить влияние различных факторов на принятие решения (имитационная модель избирательной кампании [10]). На каждом следующем шаге агент (субъект) может изменить свое мнение, а соответственно, и повлиять на поведение системы в целом.

Завершая разговор о моделировании как методе вычислительной социологии, отметим, что вопросы социальной сложности тесно связаны с нелинейными структурами и взаимодействием социальных агентов. Имитационные модели являются хорошим инструментом в решении таких задач, но есть ряд проблем: иногда невозможно разработать функционирующую модель социологических теорий, поскольку основные концепты плохо поддаются операционализации. Помимо этого, существует риск несоответствия между имитационной моделью и реально действующим механизмом социальной жизни.

Применение машинного обучения оказалось полезным для науки в целом. Машинное обучение для социальных наук — особый инструмент, требующий обдуманного подхода. Здесь важны прозрачность (алгоритм не должен восприниматься как черный ящик или участвовать в создании и воспроизводстве неравенства [4]), этичность моделей и прогнозов, анализ последствий [11].

Используя методы машинного обучения, можно прогнозировать ненаблюдаемые результаты (например, в вопросах социального неравенства) [12]. Методика «вкрапления слов» позволяет оценить изменения касательно дискриминации,

отношения к женщинам и этническим меньшинствам на заданном промежутке времени. Анализируя соотношение текстовых данных с данными переписи населения, можно увидеть общественно значимые изменения (демографические, национальные, профессиональные) во времени [13].

Машинное обучение предлагает исследователям возможность анализа субъективного благополучия, основываясь на количественном анализе естественного языка, и тем самым представляет большой интерес для социологов [14].

Социальные сети — это один из традиционных и наиболее активно исследуемых предметов в вычислительной социологии. Пользователи постоянно генерируют новые данные (текстовые сообщения, изображения, видео), создавая довольно обширный источник информации для исследователей. С помощью цифровых следов и поисковых запросов можно выявить множество неочевидных закономерностей социального мира.

Исследования цифровых следов тесно связаны с предсказанием поведения личности [15]. Например, можно понять влияние социального взаимодействия на принятие экономических решений [16] или построить индекс социальной связности на основе дружеских связей в социальной сети [17]. В проблеме растущего количества отказов людей участвовать в опросах социальные сети тоже могут сыграть свою роль. Так, данные из социальной сети Twitter являются хорошей альтернативой, о чем свидетельствуют проведенные исследования [18].

Цифровые следы позволяют сделать выводы о здоровье населения [19], исследовать наркополитику страны [20], гендерное неравенство [21].

Анализ поисковых запросов — еще один мощный инструмент, который позволяет описать и объяснить распространение тех или иных процессов и явлений. Например, это может быть систематическое исследование распределения культурных вкусов, потребительских предпочтений и интересов по всему миру [22]. Поисковые запросы позволяют оценить межгрупповые предрассудки, а также восприимчивость определенных этнических групп к радикализации [23].

Приведенные примеры во многом свидетельствуют о практической эффективности развития вычислительной социологии. Отношение к прогностической составляющей исследования постепенно меняется, тем не менее прогнозирование в сложных социальных ситуациях требует лучшей стандартизации и описания.

Благодаря современным исследованиям формируется и дополняется представление о социальной науке как «идущей от данных», а не подтверждающей идеи эмпиризма [3]. Любые данные производятся и собираются в строго заданном контексте, что нужно учитывать в ходе исследования. Отсутствие специальных знаний (социальной «системы координат») может привести к тому, что полученные данные могут быть интерпретированы в неверном ключе.

По-прежнему важно придерживаться существующей научной методологии и критически относиться к производству знания исключительно на основе больших данных (например, при условии охвата генеральной совокупности).

В цифровую эпоху становится значимым тот факт, что далеко не все выявленные закономерности — содержательные и обоснованные. В конечном счете каждый конкретный способ сбора больших данных и их адаптации к различным исследовательским задачам должен быть тщательно обдуман.

Вычислительная социология формируется и открывает перед исследователями возможности выявления новых закономерностей функционирования обществ и отдельных процессов. Междисциплинарное сотрудничество, выраженное в качественной математической подготовке и знаниях социальной теории, имеет многообещающие научные перспективы.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Tornberg P, Uitermark J. For a heterodox computational social science. BIG DATA & SOCIETY. URL: <https://journals.sagepub.com/doi/10.1177/20539517211047725>
2. Гринфилд А. Радикальные технологии: устройство повседневной жизни. М.: Издательский дом «Дело» РАНХиГС; 2018.
Greenfield A. Radical technology: The structure of everyday life. Moscow: Publishing house "Delo". RANEPa; 2018. (In Russ.).
3. Китчин Р. Большие данные, новые эпистемологии и смена парадигм. Пер. с англ. *Социология: методология, методы, математическое моделирование (Социология:4М)*. 2017;(44):111–152.
Kitchin R. Big Data, New Epistemologies and Paradigm Shifts. Translated from English. *Sociology: Methodology, Methods, and Mathematical Modeling (Sociology:4M)*. 2017;(44):111–152. (In Russ.).
4. Edelmann A. et al. Computational Social Science and Sociology. *Annual Review of Sociology*. 2020;46(4):61–81.
5. Fararo T., Butts C. Advances in generative structuralism: Structured agency and multilevel Dynamics. *The Journal of Mathematical Sociology*. 1999;24(1):1–65.
6. Давыдов А.А. Математическая социология: обзор зарубежного опыта. *Социологические исследования*. 2008;(4):105–111.
Davydov A.A. Mathematical Sociology: Review of Foreign Experience. *Sociological Studies*. 2008;(4):105–111. (In Russ.).
7. Fararo T., Kosaka K. Generating Images of Stratification: A Formal Theory. N.Y.: Kluwer Academic Publishers; 2003.
8. McClelland K., Fararo T. Purpose, Meaning and Action: Control Systems Theory in Sociology. London.: Palgrave-Macmillan; 2006.
9. Doreian P., Fararo T. The Problem of Solidarity: Theories and Models. N.Y.: Gordon and Breach; 1998.
10. Евсеев Е.А., Ефимова Е.Д. Имитационная модель избирательной кампании. URL: <https://cyberleninka.ru/article/n/imitatsionnaya-model-izbiratelnoy-kampanii>
Evseev E.A., Efimova E.D. Simulation Model of Election Campaign. URL: <https://cyberleninka.ru/article/n/imitatsionnaya-model-izbiratelnoy-kampanii>. (In Russ.).
11. Wallach H. Computational Social Science = Computer Science + Social Data. *Communications of the ACM*. 2018;61(3):42–44.
12. Garip F. What Failure to Predict Life Outcomes Can Teach Us. *Proceedings of the National Academy of Sciences*. 2020;117(15):8234–8235.
13. Garg N., Schiebinger L., Jurafsky D., Zou J. Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes. *Proceedings of the National Academy of Sciences*. 2018;115(6):3635–3644.
14. Hills T.T., Proto E., Sgroi D., Seresinhe C.I. Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books. *Nature Human Behaviour*. 2019;3(12):1271–1275.
15. Azucar D., Marengo D., Settanni M. Predicting the Big 5 Personality Traits From Digital Footprints on Social Media: A Meta-Analysis. *Personality and Individual Differences*. 2018;(124):150–159.

16. Bailey M., Cao R., Kuchler T., Stroebel J. The Economic Effects of Social Networks: Evidence From the Housing Market. *Journal of Political Economy*. 2018;126(6):2224–2276.
17. Bailey M., Cao R., Kuchler T., Stroebel J., Wong A. Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*. 2018;32(3):259–280.
18. Baghal T.A., Sloan L., Jessop C., Williams M.L., Burnap P. Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. *Social Science Computer Review*. 2020;38(5):517–532.
19. Chae D.H., Clouston S., Martz C.D., Hatzenbuehler M.L., Cooper H.L. F., Turpin R., Stephens-Davidowitz S., Kramer M.R. Area Racism and Birth Outcomes Among Blacks in the United States. *Social Science & Medicine*. 2018;(199):49–55.
20. Enghoff O., Aldridge J. The Value of Unsolicited Online Data in Drug Policy Research. *International Journal of Drug Policy*. 2019;(73):210–218.
21. Garcia D., Mitike Kassa Y., Cuevas A., Cebrian M., Moro E., Rahwan I., Cuevas R. Analyzing Gender Inequality Through Large-Scale Facebook Advertising Data. *Proceedings of the National Academy of Sciences*. 2018;115(27):6958–6963.
22. Bail C.A., Brown T.W., Wimmer A. Prestige, Proximity, and Prejudice: How GoogleSearch Terms Diffuse across the World. *American Journal of Sociology*. 2019;124(5):1496–1548.
23. Bail C.A., Merhout F., Ding P. Using Internet Search Data to Examine the Relationship Between Anti-Muslim and Pro-ISIS Sentiment in U.S. Counties. *Science Advances*. URL: <https://www.science.org/doi/10.1126/sciadv.aao5948>

ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR

Ксения Фуатовна Рафикова — аспирант философско-социологического факультета Института общественных наук РАНХиГС, преподаватель кафедры теоретической социологии и эпистемологии философско-социологического факультета Института общественных наук РАНХиГС при Президенте РФ, Москва, Россия
Ksenia F. Rafikova — Postgraduate student in the Department of Philosophy and Sociology, Institute for Social Sciences, RANEPA; professor of the Department of Theoretical Sociology and Epistemology, Faculty of Philosophy and Sociology, Institute for Social Sciences, RANEPA, Moscow, Russia
<https://orcid.org/0000-0002-9876-015X>
KseniyaRafikova@yandex.ru

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.
Conflicts of Interest Statement: The author has no conflicts of interest to declare.

Статья поступила 30.03.2022; принята к публикации 27.04.2022.
Автор прочитала и одобрила окончательный вариант рукописи.
The article was received on 30.03.2022; accepted for publication on 27.04.2022.
The author read and approved the final version of the manuscript.