

НЕКОТОРЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ МАТЕМАТИКО-СТАТИСТИЧЕСКИХ МЕТОДОВ ПРИ ПРОВЕДЕНИИ АНАЛИЗА СОЦИАЛЬНО-ГУМАНИТАРНЫХ ТЕКСТОВ

УДК 311.2

Заира Магомедовна Алиева, аспирант кафедры Информационных технологий и Электронных библиотек Московского Государственного Института Культуры
Эл. почта: Zaira1991@yandex.ru

В статье анализируются вопросы применения математико-статистических методов при анализе социально-гуманитарных текстов. Раскрывается сущность математических и статистических методов, приводятся варианты использования их при исследовании гуманитарных и социальных явлений. Рассмотрены ключевые проблемы, с которыми сталкивается эксперт при применении математико-статистических методов в социально-гуманитарной сфере, включая наличие устойчивого противопоставления социально-гуманитарных наук и математики; сложности выделения объекта, который является носителем проблемы; необходимости использования вероятностного подхода. Сделан вывод по результатам исследования.

Ключевые слова: социально-гуманитарный текст, математика, метод, статистический, вероятность.

Zaira M. Alieva
postgraduate student, Department of Information technologies and electronic libraries Moscow State Institute Of Culture
E-mail: Zaira1991@yandex.ru

SOME ASPECTS OF THE USE OF MATHEMATICAL-STATISTICAL METHODS IN THE ANALYSIS OF SOCIO-HUMANISTIC TEXTS

The article analyzes the application of mathematical and statistical methods in the analysis of socio-humanistic texts. The essence of mathematical and statistical methods, presents examples of their use in the study of Humanities and social phenomena. Considers the key issues faced by the expert in the application of mathematical-statistical methods in socio-humanitarian sphere, including the availability of sustainable contrasting socio-humanitarian Sciences and mathematics; the complexity of the allocation of the object that is the bearer of the problem; having the use of a probabilistic approach. The conclusion according to the results of the study.

Keywords: Humanities and social text, mathematics, method, statistics, probability.

В современном быстроразвивающемся обществе объективный анализ явлений человеческой жизни требует использования актуальных методологии и методики научного исследования. При этом наблюдается синтез знаний, а также постоянный процесс интеграции идей и методов самых разных аспектов человеческого познания. Данный процесс сопровождается постоянным увеличением и накоплением научных знаний, требующих овладения способами и приемами использования культурного наследия мировой цивилизации.

Одним из проявлений синтеза знаний называют математизацию социально-гуманитарных текстов, что является признаком появления потребности в количественном измерении закономерностей, используя математико-статистический алгоритм. При этом в науке существует мнение [4, с. 124], что принципиально нематематических научных направлений вообще не существует. Однако при этом необходимо учитывать степень математизации и этап эволюции научной дисциплины, требующей обязательного использования элементов математизации и статистического инструментария [4, с. 124].

Основная цель проведения анализа социально-гуманитарного текста с использованием математико-статистического инструментария состоит в нахождении связей количественной и качественной стороны социальных процессов как методологической основы количественных методов. Кроме того, анализ направлен также на определение объективных социальных и научных предпосылок включения в арсенал научно-исторической методики математико-статистических методов.

Использование математических и статистических методов способствует получению содержательных выводов за счет возможности анализа больших массивов информации и учета значительного количества факторов. Фактически без применения математического аппарата трудно обойтись при решении любой практически значимой задачи анализа социально-гуманитарных явлений. Математика предоставляет возможность пользоваться интеллектуальными достижениями, которые накопило человечество для изучения конкретных объектов [2].

Текст представляет собой образующую единое целое последовательность знаков. Согласно мнению М. Хайдеггера, «естественнонаучному эксперименту соответствует в историко-гуманитарных науках критика источников», то есть текстов разной природы [7, с. 98]. При этом отсутствие текста, не позволяет выделить объект исследования для социально-гуманитарных наук. Это наглядно проявляется в таких гуманитарных науках, как лингвистика, филология, искусствоведение и так далее. Для гуманитарных наук, по мнению М.М. Бахтина, их предмет исследования, то есть «дух (и свой, и чужой) не может быть представлен как вещь (прямой объект естественных наук), а только в знаковом выражении, реализации в текстах и себя самого, и другого» [5, с. 84]. Потому для ученых-гуманитариев необходимо глубокое и тонкое понимание текста, а также количественная интерпретация с использованием математико-статистических методов.

Необходимо отметить, что математическая наука является своеобразной коммуникативной наукой при анализе текстов. Она способна органически сливаться с другими науками. Так, при анализе социально-гуманитарной информации стала уже очевидной ее синтезирующая роль. Традиционный анализ массовых информационных источников учеными-обществоведами направлен на выявление сути анализируемых материалов, а также ведущих тенденций общественного развития, нахождению содержательных сторон исследуемого материала.



Рис. 1. Процедура контент-анализа

Отметим, что под статистическими методами понимается совокупность методов и моделей математической статистики, которая используется в науке при сборе, обработке, анализе, моделировании и сопоставлении данных исследований и характеристик объектов [5, с. 85].

Прежде всего, к количественным методам исследований относят способы получения информации об изучаемом объекте, которые позволяют выявить его количественные характеристики. Среди основных методов выделяют контент-анализ, который представлен техникой выведения заключения, которое производится благодаря объективному и систематическому выявлению характеристик текста в соответствии с задачами исследования.

Контент-анализ – это статистическая семантика, которая позволяет выявить частоту появления терминов и определить их характеристику во внеязыковой реальности. Исследователь производит квантификацию материала, что позволяет дополнить математические методы качественными характеристиками.

Процедура контент-анализа представлена на рисунке 1.

Все эти этапы контент-анализа должны базироваться на рабочей программе, которая разрабатывается по стандартным правилам прикладного статистического исследования. Разработанные принципы и процедуры позволяют провести анализ как печатных источников, так и аудио, видео, телевизионных материалов, произведений искусства, расширяя, тем самым, область научных изысканий. Многообразие источников влияет на многообразие тем, приемов и концептуальных подходов.

При этом необходимо выполнение нескольких условий:

1. Наблюдаемые явления повторяются неограниченное число раз,

или сразу осуществляется наблюдение за одинаковыми событиями в большом количестве. Однако провести наблюдение за большим числом одинаковых событий возможно только при изучении массовых источников, массовых совокупностей однородных документов.

2. Требуется наличие критерия независимости событий. Применительно к социально-гуманитарным текстам нельзя говорить о независимости фактов, между ними постоянно присутствует причинно-следственная связь.

3. Наличие постоянных условий при создании источниковой базы [5, с. 132].

Рассмотрим несколько примеров. Так, В. Альтман в 2008 году проанализировал Балладу Гете «Лесной царь» с учетом количественных характеристик ритма, звуковых структур, лексики, смыслов и некоторых грамматических свойств и, таким образом,

получил всеобъемлющую, но далеко не полную статистическую картину текста. Основной задачей являлось выявление закономерностей в тексте. Еще одну попытку анализа текста в 2011 году предпринял Овербек. Он стремится проанализировать текст с использованием количественных и качественных методов, чтобы доказать наличие определенного стиля итальянских оперных либретто [11].

Текст является более проблематичным для анализа, так как статистические методы не могут быть применены в полной мере, и необходимо работать с текстом, чтобы определить теоретические концепции. Затем можно использовать статистические методы для определения появления концепций в отношении друг к другу, определения вероятности, но на начальном этапе анализа требуется ручная обработка текста с использованием информационных технологий, к примеру, с помощью Пакета качественного анализа Atlas.ti [10].

Кроме того, существуют и другие специальные программы для анализа текста, представленные в табл. 1.

На сегодняшний день существуют десятки автоматизированных количественных методов анализа гуманитарного текста. Для выбора конкретного инструментария исследователю необходимо обратить внимание на следующие моменты:

Таблица 1

Информационные программы для анализа текста

Программа	Описание
САТРАС	Читает текстовые файлы и производит различные выходы, начиная от простой диагностики (например, слова и буквенные частоты) в краткое изложение основных идей в тексте. Это раскрывает закономерности словоупотребления и позволяет проводить систематизацию и кластерный анализ.
Concordance 2.0	Эта программа позволяет сделать полный анализ согласований с текстами любого размера. Ограничивается только доступным дисковым пространством и памятью.
Diction 5.0	Содержит ряд встроенных словарей, которые ищут текстовые документы по пяти основным семантическим признакам (активность, оптимизм, уверенность, реализм и общность) и 35 суб-функциям (в том числе, упорство, вина, амбивалентность, движение и связи). После того как текст пользователя анализируется, программа сравнивает результаты для каждой из 40 категорий словаря с «нормальным диапазоном оценки».
DIMAP	Программа включает в себя различные лексические инструменты с использованием компьютерной лингвистики и обработки естественного языка
MECA	Содержит 15 процедур для анализа текста. Многие из этих процедур основываются на когнитивном картировании с акцентом на обеих концепциях и взаимосвязях между ними.

Источник: Quantitative Text Analysis Programs.

1. Вычислительные и статистические сложности (например, сводная статистика, выпуклая оптимизация, латентная переменная обучения);

2. Количество доменных предположений в качестве входных данных (например, ручная сборка материала);

3. Сложность языковой репрезентации, используемой в анализе (например, слова и фразы, лица, мнения) [8].

Однако анализ социально-гуманитарных текстов при использовании математико-статистических методов сопряжен с рядом существенных проблем, которые стоят перед исследователями.

Во-первых, количественные и качественные методы, используемые в социальных науках, зачастую контрастируют с аналитическим, критическим, и спекулятивным методами, используемыми в гуманитарных науках, которые сосредотачиваются на состоянии человека, проявляющегося, например, в литературе, религии, философии, и визуальных и исполнительских искусствах, а также коммуникации, культурной антропологии, истории, лингвистики и могут быть изучены с использованием гуманистического или социального научных методов [9].

Во-вторых, необходимо учитывать, что любой источник предполагает двойную информацию, которая напрямую выражена и скрыта. Так, скрытая информация представлена взаимосвязями социальных процессов, которые выражены через качественные признаки, не позволяющие выявлять при социальном познании основополагающие особенности содержания информационных массивов. Данная задача может быть реализована с использованием количественных методов в социально-гуманитарном тексте [3, с. 14].

Во-третьих, на уровне обыденного сознания на сегодняшний день сохраняется устойчивое противопоставление социально-гуманитарных наук и математики, существует мнение об их несовместимости, что не исключает между ними довольно успешного сотрудничества.

В-четвертых, на практике часто возникают ситуации, когда довольно сложно выделить объект, который

является носителем проблемы и соответственно использовать его как источник информации. Как правило, такие ситуации связаны с попытками прогнозировать изменение социальных явлений, процессов; а также существует необходимость представить состояние интересующего предмета через один, два, пять лет и более или дать объективную характеристику таким сторонам деятельности, по которым их самооценка может оказаться искаженной.

В-пятых, существует проблема достоверности полученных количественных результатов анализа социально-гуманитарной сферы, что измеряется понятием вероятности. Под ней понимается объективная категория, которая выступает мерой возможности конкретного результата, характеризуемого с количественной определенностью возможности появления события.

В-шестых, математико-статистические методы пришли в микроэкономический анализ из экономической статистики. Методы изучения связей связаны с большим количеством оговорок и недопущений, среди которых, к примеру, невозможность повтора требуемого явления и события в целях формирования совокупности. Аналитик должен исключительно четко представлять себе всю условность количественных оценок, которые получены с помощью таких методов, а не абсолютизировать их [6].

Однако первоочередным условием правильного, научно обоснованного использования математико-статистических методов является соответствие ее средств и аппарата реальным отношениям, существующим в объектах исследования. При этом требуется наличие определенных закономерностей и связей, допускающих математическую обработку и выражение на языке современной математики.

Таким образом, использование математико-статистических методов при проведении анализа социально-гуманитарных текстов представляет собой современный подход к анализу ранее не подвергавшихся количественной обработке явлений. При этом применение подобных методов сопряжено с проблемами и оговорками, представленными в данном исследовании. Данные

оговорки должны учитываться при проведении анализа. Они имеют решающее и принципиальное значение для объективности и актуальности полученных выводов.

Литература

1. Бородкин Л.И. Методологические проблемы применения математических методов в историко-гуманитарных исследованиях // Математизация современной науки: предпосылки, проблемы, перспективы. – М., 1986. С. 130–139.

2. Жолудева В.В., Панарский Н.С. Математико-статистические методы и модели в социологическом исследовании. [Электронный ресурс]. Режим доступа: http://www.edit.muh.ru/content/mag/trudy/05_2011/14.pdf

3. Грехов А.В. Единство квантификационного и традиционного методов исследования как методологическая проблема исторического познания. Автореферат дис. на соискание учен. степени доктора филос. наук. 2005.

4. Громов Г. Р. Очерки информационной технологии. – М.: ИнфоАрт, 1993.

5. К философии поступка // Философия и социология науки и техники. Ежегодник. 1984–1985. – М., 1986.

6. Математико-статистические методы изучения связей: корреляционный, регрессионный анализ, кластерный анализ, дисперсионный анализ. [Электронный ресурс]. Режим доступа: <http://economicsforall.usoz.ru/>

7. Хайдеггер, М. Статьи и работы разных лет / Пер., сост. и вступ. ст. А. В. Михайлова. – М.: Гнозис, 1993.

8. Brendan T. O'Connor. Statistical Text Analysis for Social Science. CMU-ML-14-101. 2014.

9. Generic Definitions for Area Distributions (HU, NS, SS, MSA, CE, ID). [Электронный ресурс]. Режим доступа: <https://www.lsa.umich.edu/students>

10. Tom Wilson. Information science' and research methods. [Электронный ресурс]. Режим доступа: <http://www.informationr.net/tdw>

11. Vivien Altmann, Gabriel Altmann: Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen. RAM-Verlag, Lüdenscheid. – 2008.