

УДК 001.53

***ПРОВЕРКА ГИПОТЕЗ В БОЛЬШИХ ДАННЫХ С МОДЕЛИРОВАНИЕМ
В КОГНИТИВНЫХ СИСТЕМАХ***

Кудрина А.А.

студент,

Российский университет транспорта

Москва, Россия

Шмаль В.Н.

к.т.н., доцент,

Российский университет транспорта

Москва, Россия

Аннотация

В статье рассматривается актуальный вопрос, касающийся философии науки с точки зрения математической логики в современном мире. В процессе рассуждения, авторами определяются термины «наука», «ненаука», «этика», раскрываются их взаимосвязи между собой, указываются важные факторы, которые отделяют эти термины друг от друга. Описана важность и перспектива применения современных инструментов и способов обработки большого количества информации в инновационных сферах деятельности человека.

Ключевые слова: Наука, знания, гипотеза, информация, модель, данные, большие данные, методы обработки, когнитивные системы.

***HYPOTHESIS TESTING IN BIG DATA WITH MODELING IN COGNITIVE
SYSTEMS***

Kudrina A. A.

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

student,

Russian University of Transport

Moscow, Russia

Shmal V. N.

Ph. D., Associate Professor,

Russian University of Transport

Moscow, Russia

Abstract

The article deals with a topical issue concerning the philosophy of science from the point of view of mathematical logic in the modern world. In the process of reasoning, the authors define the terms "science", "non-science", "ethics", reveal their relationship to each other, indicate the important factors that separate these terms from each other. The article describes the importance and prospects of using modern tools and methods of processing a large amount of information in innovative areas of human activity.

Keywords: Science, knowledge, hypothesis, information, model, data, big data, processing methods, cognitive systems.

Гипотеза — это предлагаемое объяснение явления или, альтернативно, аргументированное предложение, предполагающее возможную корреляцию между набором явлений. В более широком смысле это можно определить как объяснение набора произошедших событий; или, другими словами, объяснение модели или процесса создания модели.

Гипотеза — это просто модель события, которая делает предсказание о том, что произойдет в будущем. Другими словами, предлагаемая модель не

предсказывает, как именно произойдет событие. Скорее, он дает отчет о том, что могло произойти, давая некоторое общее представление о том, как будет разворачиваться следующее событие.

Размышление о том, есть ли у нас подходящая гипотеза о процессе, — это первый шаг к тому, чтобы сделать этот процесс более строгим вместо того, чтобы сосредоточиться на неточно проверяемых предсказаниях. Проблема, возникающая из-за того, что этого не делается, заключается в необходимости личного эксперимента для подтверждения вашей предложенной теории. Один из способов подумать об этой проблеме - провести различие между простой проверкой гипотез в нашем индивидуальном сознании и тестами, значимыми в смысле получения точных ответов.

В науке мы больше, чем просто проверяем гипотезы; мы проводим эксперименты, которые раскрывают правду о нашей реальности. В широком смысле мы можем сказать, что любой эксперимент, который может проверить модель, является научной гипотезой.

Когда мы пытаемся проверить наши гипотезы, они часто зависят от входных данных, которые у нас есть в данный момент. Другими словами, если мы собираемся сделать прогноз относительно процесса, стоит понять, на что похожа внешняя среда.

Если мы хотим проверить научную гипотезу, полезную для нашего мира, имеющую реальные последствия для нас, мы должны понять, на что похожа реальность, как с точки зрения пользы для нас, так и для других людей.

Научные эксперименты включают в себя прогнозирование желаемого результата и проверку этих прогнозов. С помощью этого процесса мы можем построить модель, которая дает нам точный отчет о том, как устроен мир, и точность этого отчета зависит от точности гипотез.

Причина тщательного тестирования заключается в том, что оно помогает создавать модели, актуальные для мира, в котором мы живем. Если наши знания — это просто наши личные представления о мире, тогда мы не

сможем получить самое точное представление о реальности. Мы могли бы потратить больше времени на проверку наших гипотез, чем на проверку наших теорий о том, как работает реальность. Наше знание реальности может зависеть от нашего знания наших личных представлений о реальности.

В последние годы большие данные в науке стали дорогостоящим занятием, равно как и разработкой других технологий, изменяющих мир. Большие данные в науке стали сложным междисциплинарным предметом, включающим исследования и разработки в широком диапазоне областей. Компонент больших данных, или наука о данных, в научных исследованиях и в научных продуктах сейчас очень важен для понимания почти всех аспектов нашей повседневной жизни и окружающей среды. Спрос на ученых, способных работать в этой области, также растет.

Термин «большие данные» может относиться к множеству методов и технологий, связанных с обработкой данных и управлением большими объемами данных для различных приложений. Например, большие данные обычно относятся к процессам и технологиям, связанным с компьютерными сетями. Технология больших данных в настоящее время применяется во многих областях.

Системы информационных технологий: такие процессы, как обслуживание баз данных, конфигурация сети, управление серверами, а также мониторинг и обслуживание веб-сайтов, становятся все более сложными. Технологии анализа больших данных обычно сочетаются с системами больших данных, чтобы обеспечить быстрый интеллектуальный мониторинг, контроль и управление такими ИТ-системами.

Финансы: анализ больших данных позволяет менеджерам кредитного риска и инвесторам быстро реагировать на все более сложные рыночные условия, выявляя возможности в своих данных и понимая последствия принимаемых решений. Все большее число организаций, таких как агентства

кредитных рейтингов и банки, используют анализ больших данных для понимания кредитного поведения и рыночной активности.

Социальные и поведенческие науки: аналитика больших данных станет неотъемлемой частью повседневной социальной науки, где исследователи будут пытаться решать сложные социальные проблемы, используя большие данные и идеи, полученные из социальных сетей, чтобы понять, как общество влияет на окружающую среду, систему образования, здоровье и экономическую деятельность. Это позволит исследователям понять отношение людей к различной социально-экономической и политической политике.

Наука: исследователи работают над способами получения важной и надежной научной информации с помощью аналитики больших данных. Аналитика больших данных может позволить ученым собирать большие наборы данных, анализировать данные и получать критическую информацию для понимания таких проблем, как изменение климата, производство энергии, токсичные вещества и влияние использования воды.

Современная медицина: используя большие данные, исследователи разрабатывают решения важных и сложных клинических проблем и помогают практикующим врачам принимать более эффективные медицинские решения. Медицинские исследователи используют большие данные для сбора широкого спектра медицинской и научной информации, а также для анализа данных и разработки новых лекарств для решения сложных проблем со здоровьем, таких как сердечно-сосудистые заболевания. Например, исследователи собирают ценные данные о здоровье и разрабатывают сложные методы анализа данных для решения этих сложных клинических проблем.

Статистика: анализ больших данных используется для улучшения традиционных статистических инструментов, таких как регрессионный анализ, логистическая регрессия и многовариантный анализ. Использование анализа больших данных для анализа статистики обеспечивает более точные

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

прогнозы и предоставляет своевременную, точную и надежную статистику. Например, специалисты по статистике используют статистические методы для понимания и решения важных статистических проблем в банковском деле и финансах, здравоохранении, мониторинге окружающей среды, изменении климата и экономической статистике.

Одной из составляющей больших данных являются системы когнитивных вычислений, которые все еще находят свое место и нишу в мире, даже несмотря на битву с лидерами ИИ.

Одним из примеров является безопасность, лежащая в основе систем когнитивных вычислений. IBM Cognitive Security Systems, приобретенная Symantec, является одним из таких примеров. Cognitive Security Systems использует искусственный интеллект для обнаружения внутренних угроз и подозрительной деятельности, например незаконных финансовых транзакций, секретных документов и использования незащищенных мобильных устройств для связи, а также других угроз.

Однако ретроспективный анализ и предсказание - не самые полезные функции когнитивной безопасности. Когнитивные системы действительно оказывают влияние, выявляя закономерности в конце воронки. По словам Макса Шутца, директора по инновациям и компьютерной безопасности в IBM, конец воронки — это когда человек получает представление об угрозе.

Например, IBM Watson предлагает, чтобы пользователь был вовлечен в определенные действия, чтобы его можно было пометить как потенциальный риск, например, скачивание приложений или скачанных документов в течение последних шести месяцев. Но когнитивные системы не думают о сценариях «что, если».

Машинное обучение используется для построения системы классификации. IBM также использует то, что Шутц назвал «точным машинным обучением», чтобы помочь разработчикам улучшить систему.

Помимо когнитивных систем в больших данных применяются принципы математической кибернетики, которая фокусируется на факторах информации, взаимодействии частей в системах и структуре систем, математических свойствах информации и алгоритмах, которые могут управлять такими системами. В центре внимания математической кибернетики также нелинейная структура систем и связи между частями. На фокус математической кибернетики сильно повлияли работы Клода Элмера-Дьюара и, совсем недавно, Адриана МакКинти, которые разработали структурную кибернетику. В структурной кибернетике связи между частями делаются явными, что упрощает оценку систем с помощью математических методов.

Математическая кибернетика - относительно новая область, возникшая относительно недавно. Во многом это результат объединения идей и методов математической физики и кибернетики с нелинейными системами, статистическими алгоритмами и теорией информации. Математическая кибернетика выросла вместе со структурной кибернетикой и охватывает все информационные и математические аспекты этих двух областей. Математические модели информационных систем, структурная кибернетика и математическая кибернетика могут применяться к таким процессам, как информационный поток, технологическая эволюция, а также к процессам принятия решений. Вычислительные ресурсы, такие как классические компьютеры и базы данных, можно комбинировать с математическим моделированием для моделирования решений.

Задачи математической кибернетики и структурной кибернетики: выявление и понимание процесса, лежащего в основе любой системы; понимание структуры любой системы; выбор математической модели и ее оптимизация.

Теория информации — это раздел математики, тесно связанный с обработкой информации, изучением математики данных и информации.
Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

Многие методы, используемые в математической кибернетике, были разработаны в теории информации, поэтому терминология часто пересекается. Например, реляционные базы данных изначально назывались ассоциативными таблицами. Концепция статистической модели, введенная Сергеем Боренштейном, является еще одним совпадением. Этот раздел в основном посвящен нереляционным моделям, но также следует использовать статистические модели.

Информационные и статистические модели можно сформулировать по-разному. Информационные модели - это модели информации, а статистические модели - это модели того, как система реагирует на последовательность входных данных. Обычно предполагается, что информация генерируется определенным образом, в то время как статистические модели требуют, чтобы входные данные и ответные данные были случайными. Также предполагается, что статистическая модель является симметричной, так что каждый элемент в наборе данных с одинаковой вероятностью будет релевантным. С математической точки зрения статистические модели состоят из математических операций, которые можно выразить математическим правилом. В информационной модели каждый элемент в наборе данных считается случайным, тогда как в статистической модели один элемент может быть релевантным, а другие не могут быть актуальными одновременно. Элементы можно описать с помощью математического правила, так что каждый элемент связан с определенным шаблоном взаимодействий. Концепции статистической механики и хаоса описывают как случайность взаимодействий между предметами, так и характер этих взаимодействий.

Все статистические модели определены и предсказаны математически, что позволяет обрабатывать данные автоматически. Это похоже на использование математической модели для управления роботизированной или механической системой. Одним из преимуществ математического Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

моделирования перед биологической или технологической моделью является автоматическая реализация поведения, которое часто описывается математически. Однако есть и недостатки, такие как невозможность описать совершенно неизвестное поведение. Статистические и математические модели представляют особый интерес, поскольку они собирают информацию, которая является более абстрактной, чем информация, которую можно описать с помощью биологических и технологических моделей.

В математическом контексте информация — это особый тип математического объекта, который может быть представлен как распределение любой конкретной математической функции. Информацию также можно определить как распространение распределения. Информация — это особый тип распределения, который похож на статистику, но отличается тем, что предполагает определенную взаимосвязь между входом и выходом. С математической точки зрения информационные модели — это модели, которые пытаются объяснить взаимосвязь между вводом и выводом. Информационная модель состоит из набора объектов, по одному для каждого входа и по одному для каждого выхода, а также набора математических функций. Набор объектов описывает вход, а набор функций описывает выход. В результате получается набор уравнений, описывающих соответствующее поведение.

К задаче моделирования информации можно подойти двумя способами. В одном подходе считается, что информация указана точно, а выходные данные - нет. При таком подходе информационные модели чаще всего разрабатываются в математической теории информации. Другой подход, называемый теоретико-информационным моделированием, предполагает, что все соответствующие пары входов и выходов указаны наиболее точно. Если это предположение верно, информационные модели можно описать математически. Важность ввода также определяется сложностью вывода. Очень сложный результат, который невозможно описать математически, Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

обычно считается непригодным для дальнейшего моделирования, хотя это может иметь большее значение в биологическом или технологическом контексте. Информационные модели часто описываются математически. Информационные модели можно определять по-разному. Например, в интеллектуальном анализе данных информационные модели обычно определяются как математические структуры, которые являются моделями входных и выходных данных, с использованием математических операций, которые могут быть определены в терминах математических правил.

Таким образом философия науки смотрит на математическую логику, лежащую в основе научного метода, на то, что отделяет науку от ненаучности, и на этику, которая подразумевается в науке. Один из основополагающих принципов философии науки состоит в том, что вы должны исследовать природу научных вопросов. Если вы хотите быть хорошим философом науки, недостаточно иметь хорошее представление о природе науки; вы также должны иметь хорошее представление о природе научных вопросов.

Библиографический список:

1. Светлов В.А. Философия и методология науки: Учебное пособие // В.А. Светлов, И.А. Пфаненштиль. - М.: Инфра-М, 2019. - 288 с.
2. Прокофьева Е.С., Егоров П.А., Бородина Е.В. Подготовка специалистов с использованием современных методов обучения // Железнодорожный транспорт. 2019. №7. С. 22-25.
3. Редько В.Г. Моделирование когнитивной эволюции: На пути к теории эволюционного происхождения мышления / В.Г. Редько. - М.: Ленанд, 2019. - 264 с.
4. Шмулевич М.И. Технология блокчейн и перспективы её применения в транспортно-логистических системах // Транспорт: наука, техника, управление. Научный информационный сборник. 2019. №2. С. 3-7.

5. Баксанский О.Е. Моделирование в науке: Когнитивные модели и интеллект/ О.Е. Баксанский. - М.: Ленанд, 2019. - 304 с.
6. Резер С.М., Шмулевич М.И., Резер А.В. Создание единой цифровой платформы для информационной транспортной системы // Транспорт: наука, техника, управление. Научный информационный сборник. 2019. №11. С. 3-6.
7. Бородина Е.В., Ефимов Р.А., Сергиенко В.К. Повышение качества подготовки студентов // Железнодорожный транспорт. 2019. №9. С. 23-25.
8. Кузнецов Б.Г. Философия оптимизма: Перспективы науки и философские основы прогноза / Б.Г. Кузнецов. - М.: Ленанд, 2019. - 360 с.
9. Арбиб М. Мозг, машина и математика / М. Арбиб. - М.: Главная редакция физико-математической литературы издательства "Наука", 2019. - 224 с.
10. Маликова С. Big Data: тенденции развития, опасности и перспективы // Экономика и жизнь. 2018. №17-18(9733).

Оригинальность 100%