

ПРОБЛЕМЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

В.А. Герасимов, аспирант первого года обучения кафедры информационных технологий и управляющих систем,

Научный руководитель С.В. Шайтура, к.т.н., доцент кафедры информационных технологий и управляющих систем, Государственное бюджетное образовательное учреждение высшего образования Московской области «Технологический университет имени дважды Героя Советского Союза, летчика-космонавта А.А. Леонова» г.о. Королев, Московская область

А.Д. Емельянов, аспирант первого года обучения кафедры информационных технологий и управляющих систем,

Научный руководитель Ю.В. Стрелянок, д.т.н., профессор кафедры информационных технологий и управляющих систем, Государственное бюджетное образовательное учреждение высшего образования Московской области «Технологический университет имени дважды Героя Советского Союза, летчика-космонавта А.А. Леонова» г.о. Королев, Московская область

В данной статье рассмотрена проблема обработки больших объемов данных. В работе кратко описаны как древние, так и современные способы хранения информации, также причины увеличения объемов и скорости её накопления. Особое внимание уделено современному представлению информации, получившего название «нечеткие данные», трудностями их обработки и необходимости использования технологий искусственного интеллекта для выполнения этой задачи

Способы хранения информации, нечеткие данные, технологии искусственного интеллекта

APPLICATION OF ARTIFICIAL TECHNOLOGIES FOR BIG DATA PROCESSING

V.A. Gerasimov, graduate first year of the Department of Management, **Scientific adviser S.V. Shaytura**, Candidate of Technical sciences, Associate professor of the Department of Information technologies and control systems, State Budgetary Educational Institution of Higher Education of the Moscow Region «Technological University named after twice Hero of the Soviet Union, pilot-cosmonaut A.A. Leonov», Korolev, Moscow region

A.D. Emelianov, graduate first year of the Department of Information technologies and control systems,

Scientific adviser Y.V. Strenalyuk, Doctor of Technical Sciences, professor of the Department of Information technologies and control systems, State Budgetary Educational Institution of Higher Education of the Moscow Region «Technological University named after twice Hero of the Soviet Union, pilot-cosmonaut A.A. Leonov», Korolev, Moscow region

This article considers the problem of processing large amounts of data. The paper briefly describes both ancient and modern ways of storing information, as well as the reasons for the increase in the volume and speed of its accumulation. Particular attention is paid to the modern presentation of information, called "fuzzy data", the difficulties of processing it and the need to use artificial intelligence technologies to perform this task. Key words...

Information storage methods, fuzzy data, artificial intelligence technologies

Начиная с древних времен и по настоящее время, информация является одним из важнейших ресурсов, позволяющих человеку верно оценивать окружающую его действительность, находить наиболее выгодное решение поставленной задачи, сделать обдуманый выбор. Наличие и отсутствие определенных данных напрямую влияло и влияет на развитие различных сфер государства: экономики, политики, науки. В связи с этим, одним из ключевых процессов как древнего, так и современного общества стало постоянное накопление подобного ресурса для дальнейшего использования.

Изначально информация представляет из себя какой-либо набор данных, цифр, фактов, которые не имеют какую-то логику, структурированность. Для того, чтобы информацию использовать – ее необходимо обработать. Средства и способы обработки информации – процесс динамический и изменяется в зависимости от технологического прогресса. Человечество проделало путь от запоминания лично информации и устный пересказ или записи на пергамент до использования массивных алгоритмов сбора и обработки информации в современности. Основные способы хранения информации, расположенные в хронологической последовательности указаны в Таблице 1

Таблица 1 – Способы хранения информации

Способ	Описание
Человеческая память	Информация хранилась в умах людей и передавалась в устной форме между поколениями и представляя собой социальные обычаи, важные исторические события, личный опыт
Бумажные носители информации	Изобретение письменности позволило использовать бумагу в качестве физического носителя знание и таким образом исключить проблему искажения информации при передачи её в устной форме, а также начать создание специализированных хранилищ: архивов и библиотек
Цифровые системы хранения данных	Увеличение количества источников и объемов хранимой информации из-за постоянного развития различных сфер общества требовали новых способов хранения, превосходящие бумажные по скорости и удобства доступа. Решение подобной проблемы стало возможно благодаря достижениям в области науки и техники, которые способствовали как созданию специализированных комплексов, так и новому формату хранения данных – цифровому,

Помимо технологического развития человечества на методы обработки влияло и сам объем данных, который окружал человека. Если несколько веков назад человеку хватало, занести тот или иной фрагмент данных на бумагу, то сегодня для того, чтобы информация была доступна, достоверна и достаточна для человека – необходимы большие человеческие затраты и использования автоматизированных систем.

Современный мир переживает переизбыток получаемой информацией благодаря тому, что человек получил много инструментов для ее получения. Ярким примером подобного инструмента является глобальная сеть Интернет, предоставляющая людям

разностороннюю информацию: начиная от текстовых статей до состояния загруженности дорог в режиме реального времени. Информация сегодня – динамическая сущность. В одну минуту один и тот же факт может измениться до нескольких раз, дополниться или быть опровергнут. И для человека необходимо все это обработать и предоставить достоверную информацию [11].

К основным методам применяемых человеком при обработке следует отнести:

1. Табличный метод - это метод представления данных в виде таблицы. Данные располагаются в ячейках таблицы и могут быть сгруппированы по разным параметрам. Табличный метод удобен для сравнения значений, вычисления и анализа данных.

2. Статистический метод - это метод анализа данных, использующий статистические показатели, такие как среднее значение, медиана, мода и другие. Статистический метод позволяет делать выводы на основе большого объема данных и определять закономерности в данных.

3. Метод корреляции - это метод анализа данных, который позволяет определить связь между двумя или более параметрами. Корреляционный анализ может быть использован для определения взаимосвязи между продажами и рекламными расходами, например.

4. Метод сценарного анализа - это метод, который используется для прогнозирования будущих событий или результатов на основе различных сценариев. Сценарный анализ позволяет оценить возможные последствия различных решений и выбрать наилучший путь действий.

5. Метод решения проблем - это метод, который используется для решения проблем или поиска оптимальных решений. Этот метод включает в себя анализ причин и последствий, определение целей и выбор наилучшего решения.

Сегодня стал ясен факт, что ручная обработка информации при таком объеме и условиях динамичности, является тупиковым путем развития. Необходим колоссальный объем человеческого ресурса, наряду с монотонностью работы, что может приводить к фатальным ошибкам в обработке информации и породить огромные цепочки ошибок из-за неправильной обработанной информации. Поэтому на помощь пришли алгоритмы и автоматизированные системы, которые структурируют и помогают обработать данных, что дает уменьшение шанса ошибки в обработке. Но это только поверхность проблемы современной информации. Объемы передаваемых данных с каждым годом только увеличиваются. По прогнозам аналитической платформы Statista общий объем информации, располагаемой в сети к 2025 г может достигнуть 163 трлн Гб. Подобная тенденция послужила причиной возникновения такого термина как Big Data, формированию отдельных профессий, к примеру дата-инженер и разработки более совершенных инструментов, способных обработать весь этот объем [1].

Дополнительно влияет факт, что современная информация зачастую представляет собой набор нечетких данных, требующих дополнительные процессы обработки, для возможности хранения, и предоставления информации пользователям.

Нечеткие данные – это данные, которые не имеют четкого значения или определения. Если представить бинарные данные, которые можно описать 0 или 1, то нечеткие данные принимают значения между 0 и 1 [5, с. 92].

Стандартные, даже сложные алгоритмы обработки информации не способны эффективно работать с такими данными, а человеческий ресурс уже не рассматривается. Одним из наиболее распространенных методов для обработки огромных массивов данных, которые в большинстве представляют собой набор нечетких данных - применение технологий искусственного интеллекта.

Искусственный интеллект - это компьютерные системы, которые могут выполнять задачи, которые ранее могли выполнять только люди. Они способны обрабатывать большие объемы данных и выявлять скрытые взаимосвязи в них. Искусственный интеллект может использоваться для обработки

1. Числовые данные: числа, которые могут быть представлены в виде дискретных или непрерывных значений, такие как данные с датчиков, статистические данные, финансовые данные и т.д.

2. Текстовые данные: текстовые данные, которые могут быть представлены в виде слов, фраз, предложений или документов, такие как новостные статьи, отзывы, твиты и т.д.

3. Графические данные: изображения, которые можно использовать для распознавания образов, классификации, детектирования объектов, сегментации и т.д.

4. Аудио данные: аудио-записи, которые можно использовать для распознавания речи, анализа звука и т.д.

5. Видео данные: видео-записи, которые можно использовать для распознавания действий, классификации объектов, анализа движения и т.д.

6. Другие типы данных: такие как генетические данные, данные с датчиков IoT, данные социальных сетей и т.д [9, с. 24].

Для обучения искусственной нейронной сети необходимо иметь большой объем данных, которые будут использоваться для тренировки модели. Обычно, чем больше данных, тем лучше результаты работы нейронной сети. Поэтому важно иметь надежный источник данных, чтобы создать точную и эффективную модель.

, анализа данных, прогнозирования, классификации и многого другого. Среди наиболее распространенных технологий ИИ, применяемых в настоящее время, следует выделить искусственные нейронные сети и экспертные системы.

Разберем возможность внедрение каждой технологии на примерах:

- Определение аномального поведения параметров объекта испытаний с помощью искусственной нейронной сети

Процесс контроля параметров объекта является одним из ключевых на протяжении всего периода испытаний. Наблюдая контролируемые параметры, специалисты делают выводы о соответствии характеристик объекта тактико-техническим требованиям, о протекающих во время работ процессах и выявляют неисправности. По этой причине особое внимание уделяется степени достоверности информации, поступающей с изделия. В случае резкого изменения поведения одного-групп параметров, изображенных на рисунке 1 и 2, выражающихся в виде скачкообразной перемены значения от минимального до максимального или перехода в состояние, символизирующее неработоспособность измерительного канала, необходимо оперативно выяснить считается ли это поведение аномальным.

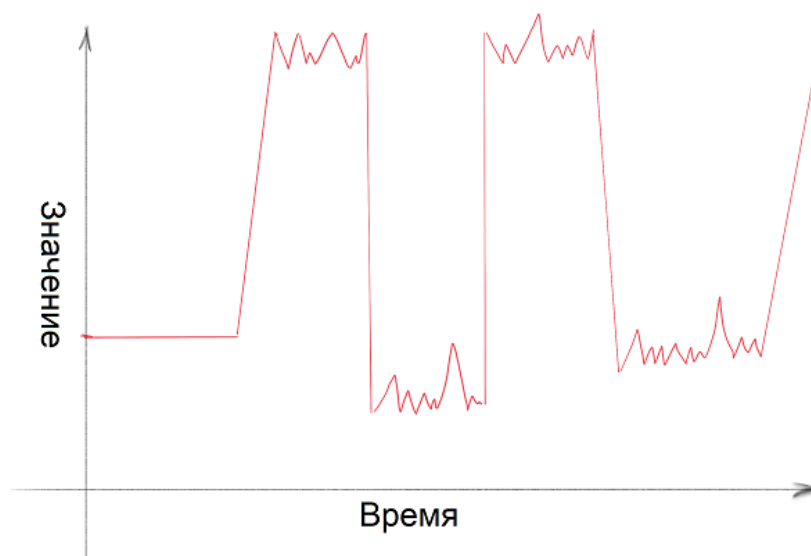


Рисунок 1 – скачкообразное изменение значения параметра

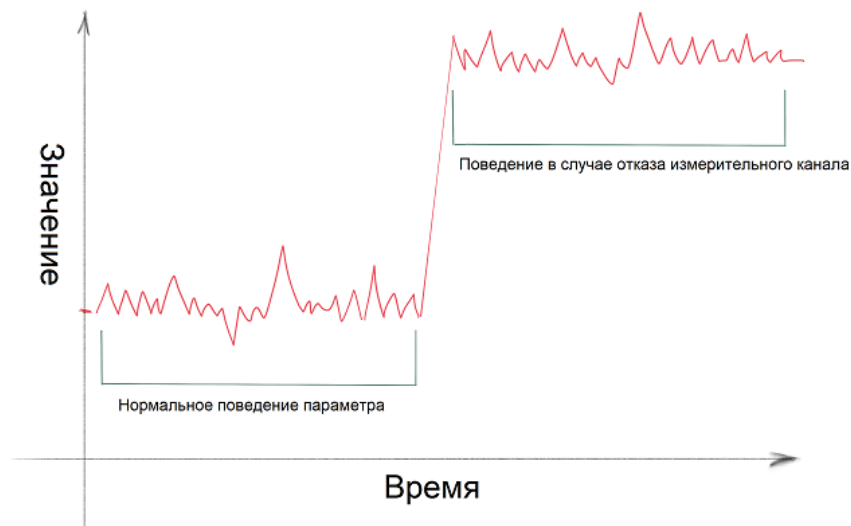


Рисунок 2 – изменение значения параметра в случае отказа измерительного канала

Основная сложность заключается в том, что для установки: нормальное это поведение параметров или нет, необходимо произвести анализ огромного массива информации, который также требует включения человека в данный процесс. Использование программной автоматизации усложняется тем, что для понимания какое поведение параметра считается аномальным нужно определить его стандартное состояние. В условиях стенда это довольно сложная задача, требующая учета многих факторов, которые, в некоторых случаях, невозможно предусмотреть, что может потребовать, как дополнительной доработки программы: оптимизацию кода, отладку для работы с новыми данными, так и создание её с нуля.

Альтернативным решением подобной проблемы является применение искусственных нейронных сетей, которые за счет своей способности обучаться на известных данных и использовать полученные знания для работы с новыми данными, которые заранее неизвестны. Таким образом, подготовив тренировочные данные, характеризующие нормальное состояние параметров объекта испытаний в условиях стенда и, обучив с их помощью ИНС, получится добиться существенной экономии времени, которое необходимо при поиске аномальных параметров.

- Применение экспертных систем для определений возможных причин отказа измерительного канала.

Наряду с использованием нейронных сетей для обработки больших данных, целесообразно применение экспертных систем. Большие данные, даже в учете мониторинга и сбора информации с датчиков стенда, представляют собой процесс обработки многоспектральной информации. В свою очередь обработка затруднена вопросами определения принципов совместной обработки данных, так как из-за различных способов представления, данных, несоответствия точности – порождает проблему обработки. Фактически проблема является аспектом представления знания, которую способна решить экспертная система. Главная особенность экспертной системы – способность накапливать и хранить знания, а также возможность обучаться или самообучаться на основе созданных ею или экспертом знаний и принимать решения для выполнения той или иной задачи [8, с. 67]. При комплексе обработки больших данных экспертная система способна объединить информацию, так как объединение основывается на определенных правилах, которые уже хранятся в базе знаний, а также на анализе достоверности данных, которые поддерживаются нейронной сетью. Далее экспертная система обрабатывает полученный объем и выдает в результат структурированное знание, которое дальше может быть использовано в той или иной предметной области, а также запоминает полученное знание и сможет использовать его для решение других задач.

В качестве одного из возможных способов применения экспертных систем в стендовых испытаниях изделий, можно рассмотреть её использование при установках причин возможных отказов измерительных каналов. Так, Инженер-испытатель, указав необходимую информацию: протяженность линии, способ подключения, время на которой произошел отказ измерительного канала, может получить список возможных причин, именно для этого случая, а также событий спровоцировавший отказ, например, вследствие сильных вибраций из-за которых произошел обрыв контактов кабеля, через которые осуществлялась подключение каналов к измерительной системе.

Литература

1. Анализ рынка искусственного интеллекта в 2021 году // Научно-технический центр ФГУП URL: <https://rdc.grfc.ru/2021/11/artificial-intelligence-market-analysis/> (дата обращения: 07.04.2023).
2. Бабич, А.В. Эффективная обработка информации. Mind mapping для студентов и профессионалов: Учебное пособие / А.В. Бабич. - М.: Бином, 2014. - 223 с.
3. Богомолова, О.Б. Обработка текстовой информации Практикум / О.Б. Богомолова. - М.: Бином. Лаборатория знаний, 2016. - 150 с.
4. Гальцов В.П. Особенности современных нейронных сетей // Вестник науки. 2019. №6 (15).
5. Глотова, М.Ю. Математическая обработка информации: Учебник и практикум / М.Ю. Глотова, Е.А. Самохвалова. - Люберцы: Юрайт, 2016. - 344 с.
6. Гречин И. В. Приобретение знаний экспертными системами // Известия ЮФУ. Технические науки. 2000. №2.
7. Зверева, В.П. Обработка отраслевой информации: Учебник / В.П. Зверева. - М.: Academia, 2018. - 48 с.
8. Ивкин А.Н. Анализ средств проектирования и разработки экспертных систем // Современные технологии поддержки принятия решений в экономике: сборник трудов III Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых / Юргинский технологический институт. – Томск: Изд-во Томского политехнического университета, 2016. – 247 с
9. Информационные технологии и вычислительные системы: Обработка информации и анализ данных. Программная инженерия. Математическое моделирование. Прикладные аспекты информатики / Под ред. С.В. Емельянова. - М.: Ленанд, 2015. - 104 с.
10. Катина Алена Михайловна, Шидловский Станислав Викторович Экспертная система оценки знаний // Доклады ТУСУР. 2006. №5 (13).
11. Обработка и хранение данных: от древности до ЦОДов // Хабр URL: <https://habr.com/ru/companies/safedata/articles/317574/> (дата обращения: 06.04.2023).
12. Остроух А.В. Введение в искусственный интеллект : монография / А.В. Остроух. – Красноярск: Научно-инновационный центр, 2020. – 250 с.
13. Остроух А.В. Интеллектуальные информационные системы и технологии: Монография / А.В. Остроух, Н.Е. Суркова. – Красноярск: Научно-инновационный центр, 2015. – 370 с
14. Остроух, А.В. Ввод и обработка цифровой информации: учебник / А.В. Остроух. - М.: Academia, 2017. - 400 с.