

**ОБЗОР И АНАЛИЗ СОВРЕМЕННЫХ МЕТОДОВ ИЗВЛЕЧЕНИЯ ЗНАНИЙ (OLAP)***Аннотация*

Статья посвящена истории, обзору направлений исследований и классификации и анализу методов извлечения знаний, даны краткие характеристики особенностей разработки современных приложений и приведены характерные примеры эффективности применения моделей и приложений в конкретных областях реальной жизни.

## □ ВВЕДЕНИЕ

Прежде чем обсуждать современные проблемы обработки знаний, приведем классическое определение процесса извлечения знаний из Википедии.

«Извлечение знаний (англ. *knowledge extraction*) — создание знаний из структурированных (реляционных баз данных, XML) и неструктурированных источников (тексты, документы, изображения). Полученное знание должно иметь формат, позволяющий компьютерный ввод, и должно представлять знания так, чтобы облегчить логические выводы.»

В последние десятилетия искусственный интеллект (как он понимается в научных исследованиях) вторгается во все сферы деятельности, становится средством интеграции наук. Программные средства, базирующиеся на технологии и методах искусственного интеллекта, получили значительное распространение в мире. Интенсивные исследования по созданию единого информационного пространства, создающего условия для совместной дистанционной работы на основе баз знаний, сегодня начали проводить все развитые в экономическом отношении страны.

Глобально, методы извлечения знаний подобны методам извлечения информации в области естественного языка и методам преобразования и хранения данных для хранилищ данных, однако главным критерием получения результата является создание структурированной информации или преобразование информации в реляционную схему. Это требует либо преобразования существующего формального знания (повторного использования идентификаторов или онтологий), либо генераций схемы, основанной на исходных данных.

Современные методы извлечения знаний возникают, главным образом, при изучении проблем извлечения знаний из разнородных оцифрованных данных. Прежде всего, необходимо сформулировать набор основных метрических характеристик, применимых к сравнению предметных областей по формальным признакам. Широкое обсуждение в литературе и в сети Интернет технологий разработки искусственного интеллекта, используемых для классификации информационных ресурсов по предметным областям, а также анализ того, насколько они эффективны в конкретных областях применения, дает возможность сделать некоторый сравнительный обзор современных методов обработки знаний.

*Неотвратимость появления технологий обработки знаний.*

Разработка прикладных систем, работающих с базами знаний, невозможна без изменения моделей данных, но ни одно из имеющихся средств хранения данных не может считаться инструментом для работы со знаниями. Для объекта знания требуется *новый подход* к представлению формализованной информации.

Необходимость формального представления сведений о предметной области возникла еще на заре исследований в области искусственного интеллекта: был сформулирован критерий, разграничивающий понятия «данные» и «знания», согласно этому критерию, знания — это «данные плюс метаданные». Любая обработка знаний осуществляется на основе

- метаданных, описывающих способы преобразования,
- вспомогательные закономерности,
- форматы хранения информации и так далее.

Изменение выше указанных составляющих позволяет модифицировать правила обработки информации, не затрагивая сами алгоритмы, что позволяет двигаться вперед, *не разрабатывая всю систему с нуля*.

Классическая обработка данных подразумевает, что вся логика работы жестко фиксирована на уровне алгоритмов, и для ее изменения необходимо модифицировать исходный код программы. В этой схеме метainформация отсутствует, однако всегда фигурирует понятие *формата* данных, который играет роль метainформации и дает исчерпывающее описание структуры обрабатываемой информации. Любая постановка задачи всегда включает в себя форматы входных и выходных данных.

Как учат еще в азах программирования, хороший стиль написания программ подразумевает описание форматов ввода/вывода в виде наборов констант и пользовательских типов данных, что позволяет в процессе разработки поддерживать незначительные изменения формата входной или выходной информации. В этом случае можно считать, что метаданные являются фиксированными и хранятся внутри исходного кода. При обработке знаний метainформация не фиксирована, хранится отдельно от исходного кода и, как любые другие данные, имеет свой формат представления во внешней памяти. Он должен учитываться в ходе разработки программного обеспечения и, следовательно, неявно используется в алгоритмах обработки знаний, а означает, что знания всегда предусматривают формат хранения метainформации и потому являются частным случаем данных.

С одной стороны, основную часть знаний аналитики получают в результате сравнения, анализа и синтеза информации из разрозненных фактов, размещенных в текстах. При работе с большими потоками документов

*процесс автоматического структурирования текстовой информации* заменяет экспертный процесс выделения фактов и объектов, выполняемый вручную, и ориентированного для хранения в больших хранилищах.

С другой стороны, задачи автоматического анализа, распознавания объектов на изображениях и видео, восстановления структуры 3D сцены — основные задачи компьютерного зрения и искусственного интеллекта.

*Компьютерное зрение* относится к области создания искусственных интеллектуальных систем, которые получают информацию из изображений и на ее основе формируют знания и делают заключения. Изображения регистрируются с помощью видео датчиков (видеокамер) в различных частотных диапазонах, и могут быть представлены с помощью двумерных матриц яркости, откликов двумерного сенсора.

*Искусственный интеллект: предмет, история развития*

*Intellectus* (с лат.) означает ум, рассудок, разум, мыслительные способности человека. Под термином *искусственный интеллект* (ИИ) понимается область информатики, предметом которой является разработка аппаратных и программных средств, позволяющих пользователю решать задачи, традиционно считающиеся интеллектуальными. Теорию искусственного интеллекта составляет наука о знаниях, о том, как их добывать, представлять в искусственных системах, перерабатывать внутри системы и использовать для решения практических задач. Технологии с использованием ИИ применяют сегодня во многих прикладных областях.

Термин *искусственный интеллект* был введён известным специалистом в области информатики Джоном МакКарти<sup>1</sup> на Дартмутском семинаре 1956 года. Это была конференция по Искусственному Интеллекту, которая собрала ведущих специалистов того времени по вопросам моделирования функций человеческого разума и естественного интеллекта.

Начало исследований в области ИИ (конец 50-х годов XX в.) связывают с работами Ньюэлла<sup>2</sup>, Саймана<sup>3</sup> и Шоу<sup>4</sup>, исследовавших процессы решения различных задач. Результатами их работ явились такие программы как "ЛОГИК-ТЕОРЕТИК", предназначенная для доказательства теорем в исчислении высказываний, и "ОБЩИЙ РЕШАТЕЛЬ ЗАДАЧ". Эти работы положили начало первому этапу исследований в области ИИ, связанному с разработкой программ, решающих задачи на основе применения разнообразных эвристических методов.

*Эвристический метод* решения задачи при этом рассматривался как свойственный человеческому мышлению "вообще", для которого характерно возникновение догадок о пути решения задачи с последующей проверкой их. Эвристическому методу противопоставлялся используемый в ЭВМ *алгоритмический метод*, который интерпретировался как механическое осуществление заданной последовательности шагов, детерминированно приводящей к правильному ответу. Трактовка эвристических методов решения задач как сугубо человеческой деятельности и обусловила появление и дальнейшее распространение термина ИИ. Первые направления, в которых стали решаться задачи, связанные с разработкой методов исследования в рамках искусственного интеллекта, оказались диаметрально противоположны по применяемым принципам.

*Нейрокибернетика* ориентирована на аппаратное моделирование структур, подобных структуре мозга. Физиологами давно установлено, что основой человеческого мозга является большое количество связанных между собой и взаимодействующих нервных клеток — нейронов. Поэтому усилия нейрокибернетики были сосредоточены на создании элементов, аналогичных нейронам, и их объединении в функционирующие системы. Эти системы принято называть нейронными сетями, или нейросетями. В последнее время нейрокибернетика снова начала развиваться из-за скачка в развитии ЭВМ. Появились нейрокомпьютеры, транспьютеры.

В настоящее время используются три подхода к созданию нейросетей:

- *аппаратный*: создание специальных компьютеров, плат расширения, наборов микросхем, реализующих все необходимые алгоритмы,
- *программный*: создание программ и инструментариев, рассчитанных на высокопроизводительные компьютеры. Сети создаются в памяти компьютера, всю работу выполняют его собственные процессоры;
- *гибридный*, то есть комбинация первых двух: часть вычислений выполняют специальные платы расширения (сопроцессоры), часть — программные средства.

<sup>1</sup> Джон Маккарти (англ. John McCarthy; 4 сентября 1927, Бостон — 24 октября 2011, Стэнфорд) — американский информатик, автор термина «искусственный интеллект» (1956), изобретатель языка Лисп (1958), основоположник функционального программирования, лауреат премии Тьюринга (1971) за огромный вклад в область исследований искусственного интеллекта. Член Национальной академии наук США (1989).

<sup>2</sup> Аллен Ньюэлл (англ. Allen Newell, 19 марта 1927 года, Сан-Франциско, США — 19 июля 1992 года, Питтсбург, США) — американский учёный в области когнитивной психологии и искусственного интеллекта. Работал в исследовательском центре RAND и Университете Карнеги — Меллон, участвовал в разработке языка программирования IPL и двух самых ранних программ искусственного интеллекта — Logic Theory Machine (1956) и General Problem Solver (1957) (совместно с Гербертом Саймоном).

<sup>3</sup> Герберт Александер Саймон (англ. Herbert A. Simon; 15 июня 1916, Милуоки — 9 февраля 2001, Питтсбург, США) — американский учёный в области социальных, политических и экономических наук, один из разработчиков гипотезы Ньюэлла — Саймона. Член Национальной академии наук США (1967) и Американской академии искусств и наук (1959). Лауреат премии по экономике памяти Альфреда Нобеля (1978) и премии Тьюринга (1975). Герберт Саймон оказал заметное влияние на развитие теории организации, менеджмента и управленческих решений. Его работы в области вычислительной техники и искусственного интеллекта оказали существенное влияние на развитие кибернетики.

<sup>4</sup> Джон Клиффорд Шоу (23 февраля 1922 - 9 февраля 1991) был системным программистом в корпорации RAND. Савтор первой программы искусственного интеллекта (Logic Theorist), один из разработчиков General Problem Solver (универсальной машины для решения проблем) и Information Processing Language (языка программирования 1950-х годов). Считается настоящим «отцом» языка JOSS. Одним из наиболее значительных событий, произошедших в программировании, стала разработка концепции обработки списков Алленом Ньюэллом, Гербертом А. Саймоном и Клиффом Шоу во время разработки языка IPL-V. Он изобрел связанный список, который остается фундаментальным во многих областях современных вычислительных технологий. Клифф Шоу - [https://ru.qaz.wiki/wiki/Cliff\\_Shaw](https://ru.qaz.wiki/wiki/Cliff_Shaw).

### Кибернетика «черного ящика»

В основу кибернетики "черного ящика" лег принцип, противоположный нейрокибернетике: не имеет значения, как устроено "мыслящее" устройство, главное, чтобы на заданные входные воздействия оно реагировало так же, как человеческий мозг. Это направление искусственного интеллекта было ориентировано на поиски алгоритмов решения интеллектуальных задач на существующих моделях компьютеров.

Исследования в области искусственного интеллекта прошли долгий и тернистый путь:

- первые увлечения (1960 год),
- лженаука (1960-65),
- успехи при решении головоломок и игр,
- разочарование при решении практических задач,
- первые успехи при решении ряда практических задач, массовое коммерческое использование при решении практических задач.

Но основу коммерческого успеха по праву составляют экспертные системы и, в первую очередь, экспертные системы реального времени. Именно они позволили искусственному интеллекту перейти от игр и головоломок к массовому использованию при решении практически значимых задач.

*Основные направления исследований, ведущиеся в области искусственного интеллекта.*

#### 1. Представление знаний и разработка систем, основанных на знаниях (*Knowledge-based systems*).

Это направление можно подразделить на более мелкие направления, а именно:

- создание специальных моделей и языков для представления знаний, а также аппаратных и программных средств для их преобразования;
- создание баз знаний, образующих ядро экспертных систем;
- методы и модели извлечения и структурирования знаний (*knowledge engineering*).

#### 2. Разработка программного обеспечения искусственного интеллекта (*software engineering for AI*):

- разработка специальных языков для решения интеллектуальных задач, в которых преобладает логическая и символьная обработка над вычислительными процедурами (LISP, REFAL, SMALLTALK, PROLOG);
- создание пакетов прикладных программ, ориентированных на промышленную разработку интеллектуальных систем;
- создание «пустых» экспертных систем, экспертных оболочек — БД, которые можно заполнять извне.

#### 3. Планирование целесообразности поведения, интеллектуальные роботы (*robotics*).

Исследования по созданию методов формирования целей и решение задач планирования действий автоматического устройства, функционирующего в сложной внешней среде. Каждый из этапов, упомянутых ниже, также может представлять самостоятельное направление исследований:

- роботы с жесткой схемой управления (манипуляторы) — первое поколение;
- адаптивные роботы с сенсорными устройствами;
- самоорганизующиеся, или интеллектуальные роботы — конечная цель робототехники. Здесь рассматриваются также проблемы: восприятия внешней информации, машинное зрение (трёхмерное).

4. Общение с ЭВМ на естественном языке и обработка естественного языка (*natural language processing - NLP*), системы машинного перевода. Пожалуй, это одно из первых направлений исследований искусственного интеллекта, и востребованного и развивающегося до сих пор. Среди составляющих этого направления можно выделить следующие:

- популярные исследования — компьютерная лингвистика и машинный перевод;
- создание языковых средств, позволяющих эффективно взаимодействовать с ЭВМ непрограммируемому (так называемому, проблемному) пользователю. В настоящее время в системах машинного перевода используют языки-посредники: *язык оригинала* → *язык смысла* → *язык перевода*. Используется ассоциативный поиск аналогичных фрагментов текста и их перевода и специальных текстовых ресурсов в базах данных (создание корпусов параллельных текстов). Структурный подход: включает в себя анализ и синтез естественных языковых сообщений:
  - морфологический анализ слов в тексте;
  - синтаксический анализ (разбор состава предложения и формирование связей);
  - семантический анализ — анализ смысла составных частей любого предложения на основе некоторой предметно-ориентированной базы знаний (на данное время не решена);
  - прагматический анализ — анализ смысла предложений в реальном контексте на основе собственной базы знаний (пока также считается нерешаемой проблемой).

5. Машинное самообучение (*machine learning*) — разработка моделей и алгоритмов, ориентированных на автоматическое накопление и формирование знаний на основе анализа и обобщения данных. В последнее время к этому направлению примыкают *data mining* — раскопка знаний, поиск и *knowledge discovery* — поиск закономерностей в БД.

6. *Pattern recognition* — распознавание образов — исследования по восприятию зрительной информации, ее обработке, формированию ответных решений на воздействие внешней среды и способов адаптации систем к среде путем обучения.

Основной подход — описание классов объектов через определенные значения значимых признаков. Любому объекту ставится в соответствие матрица признаков, по которой происходит распознавание. Этот подход тесно связан с нейрокибернетикой. Основная идея в том, что единственный объект, способный мыслить — это человеческий мозг, поэтому надо воспроизвести его структуру для мыслящего устройства. В основе подхода лежат:

- структура нейрона, образующие нейронную сеть;

- архитектура взаимосвязи нейронов в сети (топология);
- способ обучения нейронной сети.

## 7. Новая архитектура компьютера (*hardware and architecture*).

*Основные задачи, решаемые в области искусственного интеллекта.*

1. *Представление знаний и разработка систем, основанных на знаниях*: разработка моделей представления знаний, создание баз знаний, образующих ядро экспертных систем (ЭС). В последнее время это направление исследований включает в себя модели и методы извлечения и структурирования знаний и сливается с инженерией знаний. В области искусственного интеллекта наибольшего коммерческого успеха достигли экспертные системы и средства для их разработки.

2. *Игры и творчество*: разработка игровых интеллектуальных задач (шахматы, шашки, го). В основе лежит один из ранних подходов (лабиринтная модель плюс эвристики).

3. *Разработка естественно-языковых интерфейсов и машинный перевод*: перспективы управления голосом, перевод с языка на язык. Первой программой в этой области стала программа перевода с английского языка на русский. Причем, первая идея, положенная в основу метода — пословный перевод, оказалась непродуктивной. В настоящее время используется более сложная модель, включающая анализ и синтез естественно-языковых сообщений, которая состоит из нескольких блоков (которые упоминались ранее на странице 3 п.4), а именно: *морфологического анализа, синтаксического анализа, семантического анализа и прагматического анализа*. Синтез включает аналогичные этапы, но несколько в другом порядке,

4. *Распознавание образов* — это традиционное направление искусственного интеллекта, одна из первых задач, с которой начались исследования в направлении искусственного интеллекта. Каждому объекту ставится в соответствие матрица признаков, по которой происходит его распознавание. Это направление близко к машинному обучению и тесно связано с нейрокибернетикой.

5. *Новые архитектуры компьютеров* — направление представляет разработку новых аппаратных решений и архитектур, направленных на обработку *символьных* и *логических* данных (языки Пролог и Лисп-машины, компьютеры V и VI поколений). Последние разработки посвящены компьютерам баз данных и параллельным компьютерам.

6. *Интеллектуальные роботы* — направление представляющее разработку специализированных электромеханических устройств, предназначенных для автоматизации человеческого труда. Со времени создания первых роботов сменилось уже несколько их поколений. Так разрабатывались:

- *роботы с жесткой схемой управления*. Практически все современные промышленные роботы принадлежат к первому поколению. Фактически это программируемые манипуляторы.
- *адаптивные роботы с сенсорными устройствами*. Есть образцы таких роботов, и осуществляется внедрение их в промышленность.
- *самоорганизующиеся, или интеллектуальные, роботы*. Это важнейшая цель развития робототехники в различных областях современной науки и техники. Основные проблемы при создании интеллектуальных роботов — проблема машинного зрения, создание качественных и уникальных 3D моделей, распознавание речи и управление.

7. *Специальное программное обеспечение* — представляет направление

- разработки специальных языков для решения задач лингвистического направления (LISP, PROLOG, SMALLTALK, РЕФАЛ и другие функциональные и логические языки),
- создания пакетов прикладных программ, ориентированных на промышленную разработку интеллектуальных систем, экспертных оболочек.

8. *Обучение и самообучение* — это разработка моделей, методов и алгоритмов, ориентированных на автоматическое накопление знаний на основе анализа и обобщения данных. Такой подход включает обучение по примерам (или индуктивное), а также традиционные подходы распознавания образов.

*Система знаний: основные определения.*

Рассмотрим более детально понятия, касающиеся собственно обработки знаний.

*Данные* – отдельные факты, характеризующие объекты, процессы и явления в предметной области.

*Знанием* является информация, которая отражает объективные свойства и связи некоторых объектов, явлений, процессов, сущностей и отношениями между ними как в субъективном, так и научном (объективном) выражении.

Под *системой знаний* понимается совокупность знаний, образующих целостное описание некоторой проблемы с доступной и достаточной степенью точности. Целостность означает, что между отдельными сведениями, входящими в знание, существуют связи и, следовательно, одни сведения могут быть выводимы из других.

В системах искусственного интеллекта знания являются основным объектом формирования, обработки и исследования. База знаний, наравне с базой данных, - необходимая составляющая программного комплекса искусственного интеллекта. Машины, реализующие алгоритмы искусственного интеллекта, называют *машинами, основанными на знаниях*, а подраздел теории искусственного интеллекта, связанный с построением экспертных систем, называют *инженерией знаний*.

Знания можно разделить на *процедурные* и *декларативные*. Исторически первичными были процедурные знания, то есть знания, "растворенные" в алгоритмах. Алгоритмы управляют данными, для изменения разного типа данных требуется изменять программы. Однако с развитием искусственного интеллекта приоритет данных постепенно изменялся, и все большая часть знаний сосредоточивалась в структурах данных

(таблицах, списках, абстрактных типах данных), то есть увеличивалась роль декларативных знаний. Сегодня знания приобрели чисто декларативную форму, то есть знаниями считаются предложения, записанные на языках представления знаний, приближенных к естественному, и понятных неспециалистам. Для решения задач некоторой предметной области, необходимо создание её *концептуальной* модели, а для представления знаний в ЭВМ они должны быть дополнительно *формализованы*.

*Представление знаний* — это соглашение о том, каким образом описывается реальный мир. Основная цель представления знаний: получить *математические модели реального мира*, с целью получения на их основе, необходимых решений.

В системах искусственного интеллекта используются следующие универсальные модели представления знаний:

- семантические сети;
- фреймы;
- логическая модель;
- продукционная модель.

Рассмотрим более подробно перечисленные модели представления знаний.

1. *Семантические сети*. *Семантическая сеть* — это ориентированный граф, вершинами которого являются понятия, а дугами — отношения между ними. Понятиями обычно выступают абстрактные или конкретные объекты, а отношения — это связи типа: "это" ("is"), "имеет часть" ("has part"), "принадлежит", "любит".

Можно ввести несколько классификаций семантических сетей. Например, по количеству типов отношений:

- *однородные* (с единственным типом отношений);
- *неоднородные* (с различными типами отношений).

Наиболее часто в семантических сетях используются следующие отношения:

- связи типа "часть-целое" ("класс-подкласс", "элемент-множество" и тому подобное);
- функциональные связи (определяемые обычно глаголами "производит", "влияет"...);
- количественные (больше, меньше, равно...);
- пространственные (далеко от, близко от, за, под, над...);
- временные (раньше, позже, в течение...);
- атрибутивные связи (иметь свойство, иметь значение...);
- логические связи (и, или, не) и другие.



Рис.1. *Фрагмент семантической сети, описывающей предметную область «животные».*

Виды отношений: 1—«часть-целое», 2—«подкласс-класс», 3—«имеет свойство», 4—«умеет».

Проблема поиска решения в базе знаний типа семантической сети сводится к задаче поиска фрагмента сети, соответствующего некоторой подсети, соответствующей поставленному вопросу.

*Преимущества* сетей:

- 1) модель адекватно отображает представления о мире;
- 2) позволяет работать не с полностью определенными знаниями;

*Недостатки*: чем сложнее модель, тем больше сеть, тем сложнее поиск.

Модель, представленная на рисунке 1, ориентирована на использование языка LISP.

2. *Фреймовая модель*. *Frame* (рамка) является единицей представления информации об объекте, который можно описать некоторой совокупностью понятий и сущностей. Фрейм имеет внутреннюю структуру, состоящую из множества элементов, называемыми слотами. Слоты могут быть разного типа (числа, тексты, математические соотношения) и содержать не только конкретное значение, но и имя процедуры, позволяющей вычислить её по некоторому алгоритму.

В слоте может размещаться ссылка на другой фрейм, что позволяет строить иерархическую структуру фреймов и обеспечивать наследование свойств родовых объектов.

```
<имя_фрейма>
<имя_слота 1> (<значение_слота 1>)
<имя_слота 2> (<значение_слота 2>)
...
<имя_слота n> (<значение_слота n>)
```

Имеются 2 типа фреймов: фреймы-прототипы (пустые фреймы, описывающие понятия) и фреймы-экземпляры (описывают конкретные объекты).

Например, фрейм-прототип «студент» может быть задан следующим образом:

студент  
ФИО (<значение\_слота 1>)  
группа (<значение\_слота 2>)  
предмет (<значение\_слота 3>)  
оценка (<значение\_слота 4>)  
Фрейм-экземпляр, описывающий конкретного студента:  
студент  
ФИО ()  
группа (401ФФ)  
предмет (ОИИ)  
оценка (отл.)

Важнейшим свойством теории фреймов является заимствованное из теории семантических сетей наследование свойств. И во фреймах, и в семантических сетях наследование происходит по АКО-связям (A-Kind-Of= это). Слот АКО указывает на фрейм более высокого уровня иерархии, откуда неявно наследуются, то есть переносятся, значения аналогичных слотов.

Для работы с фреймами был разработан язык FRL (*Frame Representation Language*).

*Преимущества* фреймовских систем:

- фреймовское представление позволяет описывать и управлять сложными знаниями больших объемов на основе описания концептуальных объектов;
- в целях увеличения гибкости системы, декларативные и процедурные знания концептуальных объектов комбинируются;
- при решении сложных проблем используют комбинации управления выводом.

*Недостатки*: громоздкость, сложность модификации вида родовых отношений.

**3. Продукционная модель** — это модель, основанная на правилах, что позволяет представить знания в виде предложений типа: «*Если* (условие), *то* (действие)»: «если  $A_1, A_2, \dots, A_n$ , то  $B$ », где  $A_1, A_2, \dots, A_n$  — факты. Под условием понимается некоторое предложение-образец, по которому осуществляется поиск в базе знаний, а под действием — действия, выполняемые при успешном исходе поиска (они могут быть промежуточными, выступающими далее, как условия, и терминальными или целевыми, завершающими работу системы).

При использовании продукционной модели база знаний состоит из набора правил, Программа, управляющая перебором правил, называется *машиной вывода*. Чаще всего вывод бывает прямой (от данных к поиску цели) или обратный (от цели для ее подтверждения — к данным). Данные — это исходные факты, на основании которых запускается машина вывода — программа, перебирающая правила из базы.

Продукционная модель чаще всего применяется в промышленных экспертных системах. Она привлекает разработчиков своей наглядностью, высокой модульностью, легкостью внесения дополнений и изменений и простотой механизма логического вывода.

*Преимущества* продукционных систем:

- подавляющая часть человеческих знаний может быть представлена в виде продукций;
- системы продукции являются модульными;
- при необходимости системы продукции могут реализовывать сложные алгоритмы;
- прозрачность системы (легко проследить логику и объяснить полученный результат).

Продукционные модели имеют два *недостатка*:

- при большом числе продукций ( $> 1000$ ) проверка непротиворечивости становится сложнее;
- неоднозначность выбора из фронта готовых продукций.

Языком, использующим продукционную модель, является язык ПРОЛОГ.

#### **4. Логическая модель.**

В основе описания лежит формальная система с четырьмя элементами:

$M = \langle T, P, A, B \rangle$ , где  $T$  — множество базовых элементов различной природы с соответствующими процедурами;  $P$  — множество синтаксических правил.

С их помощью из элементов  $T$  образуют синтаксически правильные совокупности. Процедура  $P(P)$  определяет, является ли эта совокупность правильной;

$A$  — подмножество множества  $P$ , называемых аксиомами. Процедура  $P(A)$  дает ответ на вопрос о принадлежности к множеству  $A$ ;

$B$  — множество правил вывода. Применяя их к элементам  $A$ , можно получить новые синтаксически правильные совокупности, к которым можно применить эти правила снова. Процедура  $P(B)$  определяет для каждой синтаксически правильной совокупности, является ли она выводимой.

**5. Экспертные системы (ЭС)** — это сложные программные комплексы, аккумулирующие знания специалистов в конкретных предметных областях и тиражирующие этот эмпирический опыт для консультаций менее квалифицированных пользователей.

Цель исследования экспертных систем — разработка программ, которые при решении задач из некоторой предметной области, получают результаты, не уступающие по качеству и эффективности результатам,

полученным экспертами. Экспертные системы предназначены для решения неформализованных, практически значимых задач. Использовать экспертные системы следует только тогда, когда их разработка является возможной и целесообразной.

Факты, свидетельствующие о необходимости разработки и внедрения экспертных систем:

- нехватка специалистов, расходующих значительное время для оказания помощи другим;
- потребность в многочисленном коллективе специалистов, поскольку ни один из них не обладает достаточным знанием;
- низкая производительность, поскольку задача требует полного анализа сложного набора условий, а обычный специалист не в состоянии просмотреть (за отведенное время) все эти условия;
- наличие конкурентов, имеющих преимущество в том, что они лучше справляются с поставленной задачей.

По *функциональному* назначению экспертные системы можно разделить на ниже перечисленные типы:

1. Мощные экспертные системы, рассчитанные на узкий круг пользователей (системы управления сложным технологическим оборудованием, экспертные системы ПВО). Такие системы обычно работают в реальном масштабе времени и являются очень дорогими.
2. Экспертные системы, рассчитанные на широкий круг пользователей. К ним можно отнести системы медицинской диагностики, сложные обучающие системы. База знаний этих систем стоит недешево, так как содержит уникальные знания, полученные от специалистов экспертов. Сбором знаний и формированием базы знаний занимается специалист по сбору знаний — инженер-когнитолог.
3. Экспертные системы с небольшим числом правил и сравнительно недорогих. Эти системы рассчитаны на массового потребителя (системы, облегчающие поиск неисправностей в аппаратуре). Применение таких систем позволяет обойтись без высококвалифицированного персонала, уменьшить время поиска и устранения неисправностей. Базу знаний такой системы можно дополнять и изменять, не прибегая к помощи разработчиков системы. В них обычно используются знания из различных справочных пособий и технической документации.
4. Простые экспертные системы индивидуального использования. Часто изготавливаются самостоятельно. Применяются в ситуациях, чтобы облегчить повседневную работу. Пользователь, организовав правила в некоторую базу знаний, создает на ее основе свою экспертную систему. Такие системы находят применение в юриспруденции, коммерческой деятельности, ремонте несложной аппаратуры.

Использование экспертных систем и нейронных сетей приносит значительный экономический эффект. Так, приведем примеры:

- American Express сократила свои потери на 27 млн. долларов в год благодаря экспертной системе, определяющей целесообразность выдачи или отказа в кредите той или иной фирме;
- DEC ежегодно экономит 70 млн. долларов в год благодаря системе XCON/XSEL, которая по заказу покупателя составляет конфигурацию вычислительной системы VAX. Ее использование сократило число ошибок от 30% до 1%;
- Sira сократила затраты на строительство трубопровода в Австралии на 40 млн. долларов за счет управляющей трубопроводом экспертной системы.

#### 6. Технологии обработки изображений (2d и 3d объектов)

Сегодня анализ изображений — одна из наиболее востребованных областей исследований, в том числе в сетевых технологиях. Биометрические данные собирают для отслеживания нарушений в финансовой и правовой сфере. Разрабатывается программное обеспечение, которое обеспечит управление функциями компьютера с помощью привычных движений руки (перелистывание, удаление с экрана и многие другие).

Задачи компьютерного зрения, представляющие методы обработки изображений, как правило, разбиваются на несколько подзадач:

- предобработка,
- выделение характерных свойств изображения объекта,
- классификация и
- заключение (принятие решения).

Этап *предобработки* обычно включает такие операции с изображением, как фильтрация, выравнивание яркости, геометрические корректирующие преобразования для облегчения устойчивого выделения характерных свойств объекта, которые представляются как некоторое множество признаков, приближенно описывающее интересующий объект, его поведение.

*Классификация* строится путем анализа полученного множества признаков, разделения признакового пространства на подобласти, указывающие на соответствующий класс. Принадлежность к некоторому классу зарегистрированного объекта или структуры на изображении используется в последующих этапах принятия решения в интеллектуальной системе.

Методы для исследования, преобразования и обработки изображений можно сгруппировать, выделив следующие операции как наиболее общие для группы:

- классификация различных видов операций над изображениями;
- точечные методы обработки изображений;
- гистограмма, операции с гистограммой: нормализация, эквализация, гамма-коррекция и еще целый ряд операций;
- алгебраические и геометрические преобразования изображения.

Чаще всего требуется произвести каким-либо образом уменьшение данных об изображении, не теряя при этом качества и характеристик этого изображения. Для таких преобразований разработаны различные методы, их можно отнести к направлениям, в которых разнятся принципы исследований. Первую группу представляют следующие методы.

1. *Пространственно-локальные методы обработки изображений (фильтры)*
  - Пространственная частота изображения. Свертка изображения.
  - Построение фильтров. Низкочастотные, полосные и высокочастотные фильтры.
  - Выделение края. Фильтры направленного градиента, Лапласа, Собеля.
2. *Математическая морфология*
  - Нелинейные фильтры, фильтры порядковых статистик, релаксации.
  - Частные случаи фильтров порядковых статистик: эрозия, дилатация, медиана.
  - Морфологические операции на дискретных изображениях, частный случай бинарного изображения
3. *Спектральный анализ изображений*
  - Дискретное преобразование Фурье.
  - Преобразования Адамара, Уолша, Хаара.
  - Быстрые методы расчёта спектральных преобразований.
4. *Вейвлет-анализ*
  - Базис функций Габора
  - Дискретное вейвлет-преобразование
  - Анализ мод вейвлет-преобразования
  - Вейвлет-анализ как метод выделения пространственной и частотной структуры данных
5. *Выделение признаков и сегментация*
  - Текстура и методы её анализа
  - Понятие связности. Метод «лесного пожара».
  - Сегментация методами водораздела и квадрирования.
  - Диаграммы Вороного

Вторая группа использует принципиально иные методы исследования, их можно отнести к ниже перечисленным методам.

1. *Представление и описание областей изображения*
  - Цепные коды, стягивающие ломаные.
  - Методы слияния и разбиения при построении границ.
  - Сигнатуры (одномерные функции описания границ).
  - Остовы областей. Преобразование к главным осям, методы утоньшения.
2. *Поиск объектов заданной формы*
  - Алгебраический и морфологический корреляторы.
  - Обобщённое преобразование Хафа, аналогия со свёрткой
  - Преобразования Хафа для прямых и окружностей
3. *Оптические потоки*
  - Алгоритмы Лукаса-Канаде, Хорна-Шунка.
4. *Проективная геометрия и восстановление трёхмерной структуры*
  - Однородные координаты, эпиполярная геометрия, фундаментальная матрица.
  - Восстановление формы из движения, из стерео, из затенения, из расфокусировки.

## ЗАКЛЮЧЕНИЕ

К 2020 году технологии обработки данных и знаний глубоко вошли в различные сферы нашей жизни. Сотни тысяч привычных вещей не существовали бы без цифровых технологий или были бы гораздо менее удобными в использовании. В значительной мере это относится к высокоскоростной обработке оцифрованных текстовых данных и изображений, графических данных. В большой степени это касается и задач, решаемых Вооруженными силами Российской Федерации.

В работе рассмотрены различные направления, касающиеся обработки самых разных типов знаний, представлена классификация моделей обработки знаний, недостатки и преимущества тех или иных моделей представления и обработки знаний. С точки зрения дальнейшего развития рассматриваемой области важно отметить большой прорыв в направлении компьютерного зрения и разработки манипуляторов, действующих под управлением импульсов нервной системы.

Очень подробно рассмотрены направления развития обработки знаний, модели извлечения знаний, особенностей применения тех или иных моделей к конкретным задачам в реальных условиях, отмечены сильные стороны и недостатки рассматриваемых технологий.

Сегодня прикладные системы работают в разнообразных предметных областях, поэтому использование существующих подходов не является серьезным сдерживающим фактором. Создание нового метода представления знаний станет необходимым при автоматизации работы с более сложной разнородной и распределенной информацией, такой как медицинские карты, данные криминалистических и судебных разбирательств, результаты различных экспертиз, области военно-морских применений и тому подобного.

Эти исследования подтверждают важность подготовки специалистов в области программирования и извлечения знаний в различных родах войск Российской Федерации, в том числе и в ВМФ РФ.



Результаты статьи интересны и весьма важны также с точки зрения знакомства с современными высокоэффективными технологиями обработки знаний, что позволяет более ясно представлять перспективы развития современной вычислительной техники и программного обеспечения, а также использовать для углубленной подготовке специалистов Военно-морского флота в области современных перспективных направлений разработки параллельных архитектур и общего и специального программного обеспечения при обучении курсантов высших военных учебных заведений по инженерным специальностям.

#### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М., Техносфера, 2005 – 1072 с.
2. Методы компьютерной обработки изображений. Под ред. В. А. Сойфера. М., Физматлит, 2003. – 780 с.
3. Шапиро Л, Стокман Дж. Компьютерное зрение. М., БИНОМ. Лаборатория знаний, 2006 – 752 с.
4. Форсайт Д, Понс Ж. Компьютерное зрение. Современный подход. М., Издательский дом «Вильямс», 2004 – 928 с.
5. Яне Б. Цифровая обработка изображений М.: Техносфера, 2007 – 584 с.
6. Duda R. O., Hart P. E., Stork D. G. Pattern classification. New York, Wiley Interscience Publication, 2001 – 654 p.