

Интеллектуальные информационные системы управления

1. **Основные понятия и определения теории интеллектуальных информационных систем управления.**
2. **Структура и назначение экспертных систем.**
3. **Добыча знаний (Data Mining).**

Интеллект. Различают формулировки данного понятия по нескольким направлениям [21; 22]:

- ◆ философскую;
- ◆ биологическую;
- ◆ психологическую.

В *философии* под интеллектом понимают познание, понимание, рассудочную способность к абстрактно-аналитическому расчленению (Г. Гегель), способность к образованию понятий (Э. Кант).

В *психологии* под интеллектом понимают характеристику умственного развития индивидуума, определяющую его способность целенаправленно действовать, рационально мыслить и эффективно взаимодействовать с окружающим миром.

В *биологии* под интеллектом понимают способность адекватно реагировать (принимать решения) в ответ на изменение окружающей обстановки.

Разум. В отличие от интеллекта разум — категория глубоко человеческая, опирающаяся на сознание как высшую форму психологической деятельности. Принципиальным моментом в определении разума, так же как и сознания, является их общественный, социальный характер, поскольку и то и другое понятия сформировались в результате совместной человеческой деятельности.

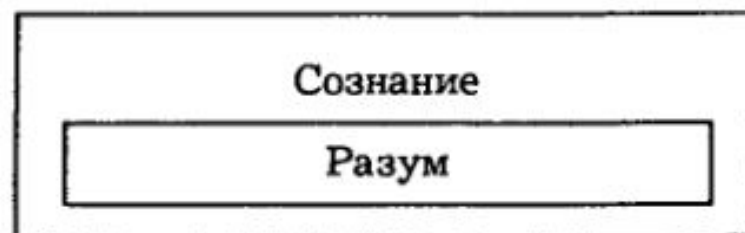
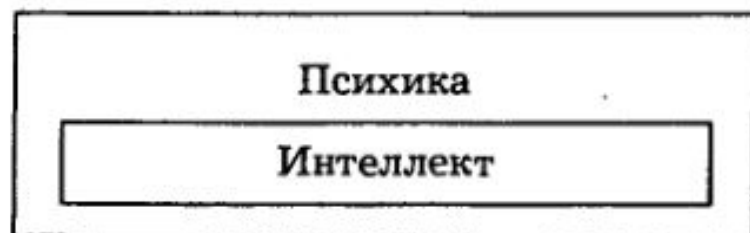
Сознание. Это понятие также трактуется различными науками неоднозначно.

С точки зрения философии сознание — свойство высокоорганизованной материи — мозга, выступающее как осознанное бытие, субъективный образ объективного мира, субъективная реальность.

При социологическом подходе сознание рассматривается прежде всего как отображение в духовной жизни людей интересов и представлений различных социальных групп, классов, наций, общества в целом.

Психика. Психика — это свойство высокоорганизованной материи — мозга, являющееся особой формой отражения действительности и включающее такие понятия, как ощущение, восприятие, память, чувства, воля, мышление и др. Отметим, что мышление и память, которыми обычно характеризуют интеллект, входят в понятие психики составными частями.

Основное отличие второй пары от первой состоит в том, что она образовалась в результате социальной, общественной деятельности людей, и поэтому социальная компонента — неотъемлемая и существенная черта сознания и разума (классическим примером может служить психика Маугли и психика “нормальных” детей).



Отсюда следует очень важный вывод: принципиально невозможно моделировать сознание и разум во всей полноте, так как для этого пришлось бы "моделировать не только человека", но и всю систему его социально-общественных отношений. В то же время моделировать интеллект как одну из компонент психики отдельных индивидуумов вполне возможно, хотя и очень сложно.

Искусственный интеллект

Искусственный интеллект – это область исследований, в рамках которых разрабатываются модели и методы решения задач, традиционно считавшихся интеллектуальными и не поддающимися формализации и автоматизации.

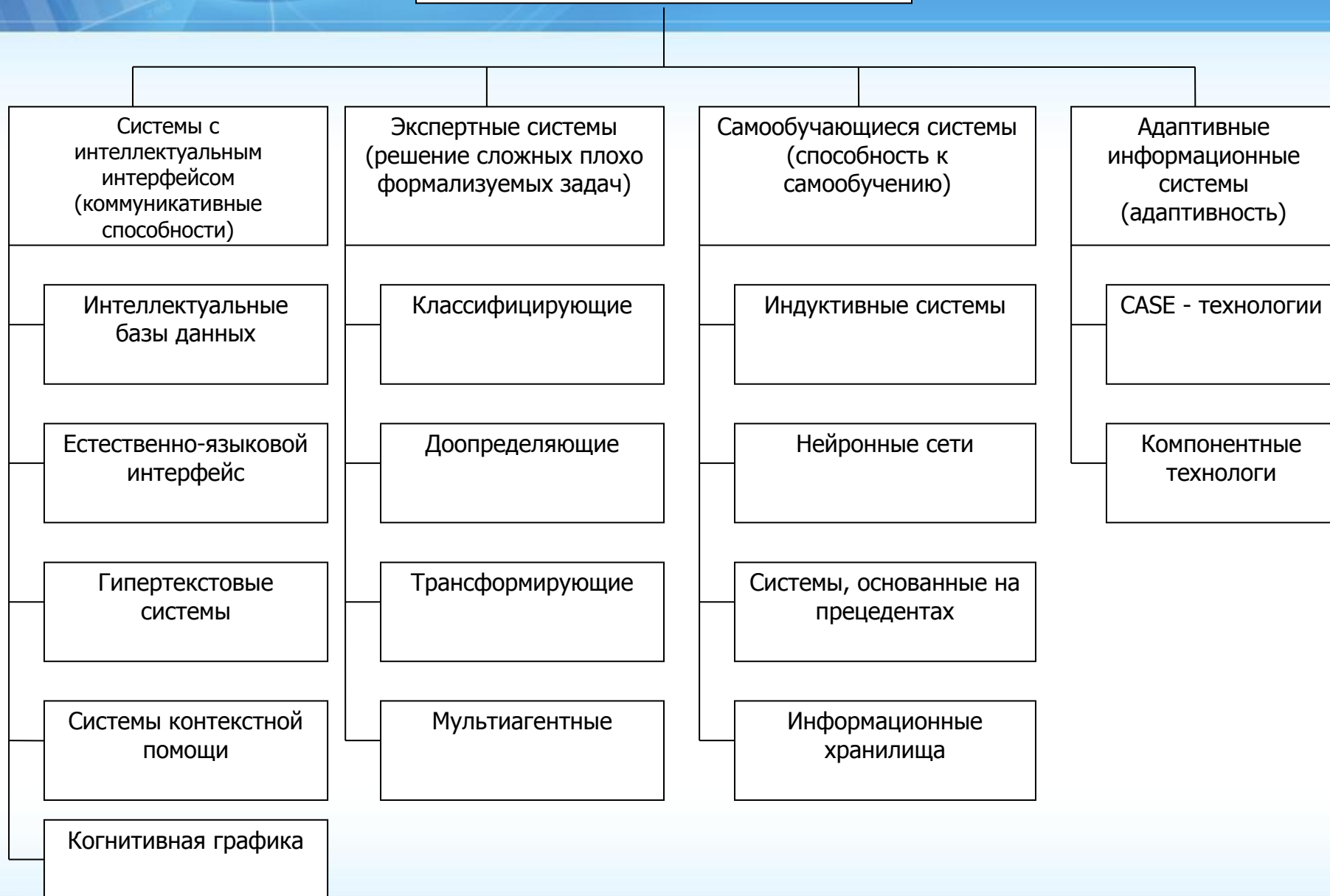
Область применения:

1. Разработка интеллектуальных информационных систем или систем, основанных на знаниях
2. Разработка естественно-языковых интерфейсов и машинный перевод
3. Генерация и распознавание речи
4. Обработка визуальной информации
5. Обучение и самообучение

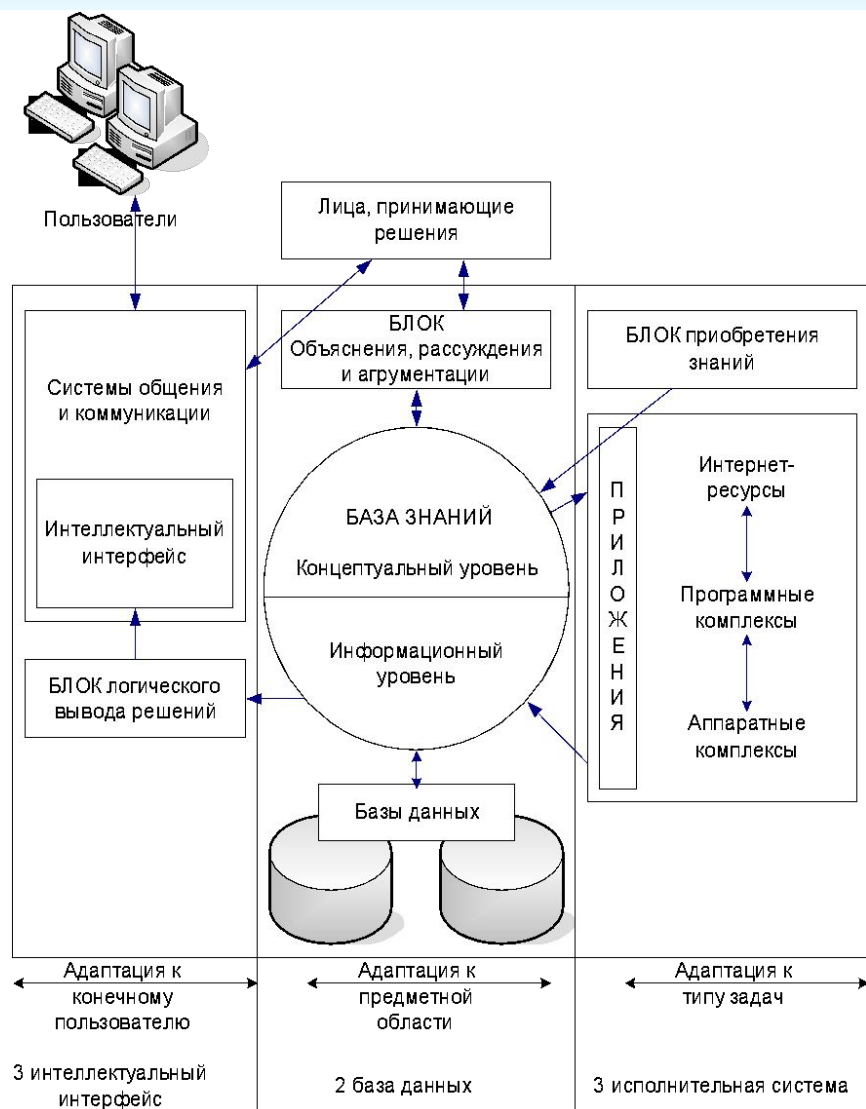
Краткая история искусственного интеллекта

Этап	Начало	Особенности	Задачи
1	конец 50-х годов	Доказательство теорем	Всевозможные игры, головоломки, математические задачи (задачи об обезьяне и бананах, миссионерах и людоедах, Ханойской башне, игра в 15 и другие).
2	начало 70-х годов	Переход к реальным проблемным средам	Создание автономно функционирующих систем, самостоятельно решающих в реальной среде поставленные перед ними задачи.
3	конец 70-х годов	Поведение в динамических средах	Создание человеко-машинных систем, интегрирующих в единое целое интеллект человека и способности вычислительной системы для решения задачи.

Интеллектуальные информационные системы



Структура системы искусственного интеллекта (СИИ)

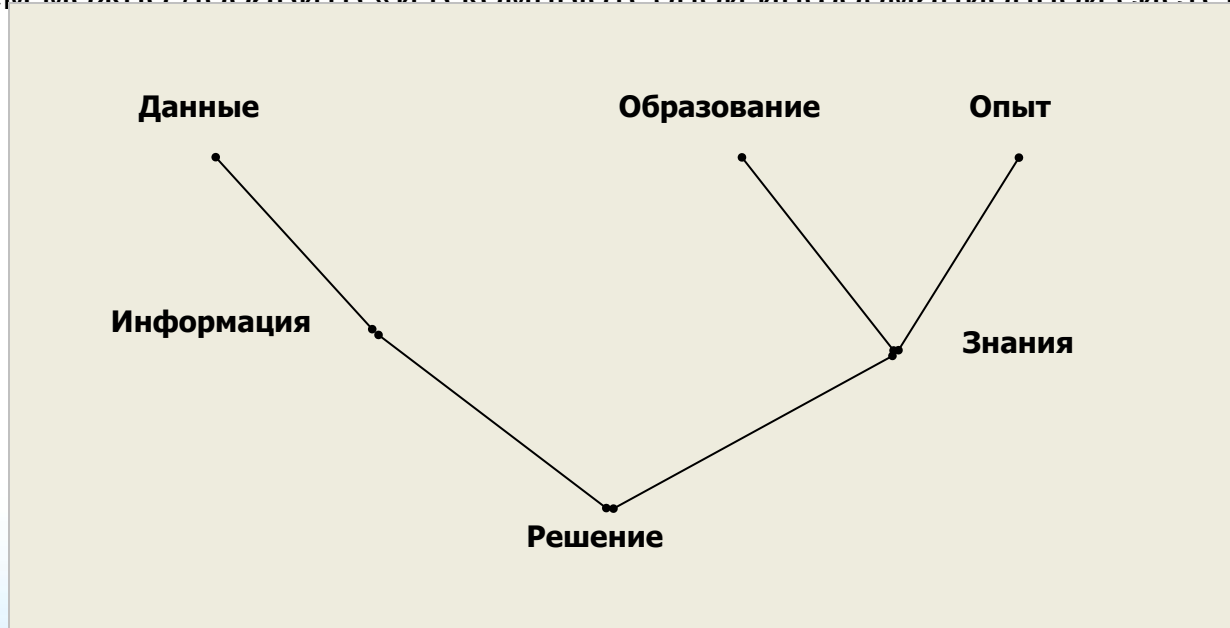


№	Название	Описание
1	Исполнительная система	Исполнительная система объединяет всю совокупность средств, обеспечивающих выполнение программ для эффективного решения задач и часто имеет проблемную ориентацию.
2	База знаний	База знаний занимает центральное положение по отношению к остальным компонентам вычислительной системы в целом, через нее осуществляется интеграция средств, участвующих в решении задач.
3	Интеллектуальный интерфейс	Интеллектуальный интерфейс - система программных и аппаратных средств, обеспечивающих для конечного пользователя использование компьютера для решения задач, которые возникают в среде его профессиональной деятельности либо без посредников либо с незначительной их помощью.

Понятие «знания»

Понятие «знания» рассматривается с различных точек зрения. В соответствии с этим имеется много определений этого понятия.

- Знания - это «закономерности предметной области (принципы, связи, законы), полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области».
- Энциклопедический словарь определяет знания как «проверенный практикой результат познания действительности, верное ее отражение в мышлении человека».
- Применительно к ситуации с использованием компьютерных информационных систем можно добавить «и в компьютерной информационной системе».



Данными называют информацию фактического характера, описывающую объекты, процессы и явления предметной области, а также их свойства.

Знания описывают не только отдельные факты, но и взаимосвязи между ними, поэтому знания иногда называют структурированными данными

Основные отличия знания от данных

Знания по сравнению с **данными** обладают избыточными возможностями, помимо собственно данных в знаниях содержится дополнительная информация, которая позволяет интерпретировать их определенным образом.

№	Особенность	Описание
1	Внутренняя интерпретируемость	Каждая информационная единица имеет уникальное имя, по которому ее можно найти и идентифицировать.
2	Структурированность	Информационные единицы должны обладать гибкой структурой, т.е. рекурсивная вложенность одних информационных единиц в другие.
3	Связность	Между информационными единицами существуют связи.
4	Семантическая метрика	На множестве информационных единиц задается отношение, характеризующее «близость» информационных единиц,
5	Активность	Принципиальное отличие от данных, означающее, что сами знания могут активно влиять на процесс их использования.

Характеристики представления знаний

№	Характеристики	Комментарий
1	Фактические и стратегические знания	Фактические – это такие знания, которые позволяют специалисту предметной области решать конкретные задачи из бизнес-сферы или в каком-либо другом виде деятельности. К ним относятся факты, взаимосвязи, системы понятий, правила. Стратегические – позволяют определить поведение объектов в ближайшем или отдаленном будущем.
2	Факты и эвристики	Факты – это хорошо известные и описанные обстоятельства. К ним относятся также экономические категории, известные и описанные закономерности и так далее. Эвристики – знания, опыт, навыки специалистов в соответствующих предметных областях. Они являются объектом изучения и внедрения в информационные системы различного назначения.
3	Декларативные и процедурные знания	Декларативные знания являются очевидными, например: выручка – сумма, полученная в результате продажи товаров. Товар – изделие, предназначенное для продажи. Процедурные знания – алгоритмы преобразования декларативных знаний, действий над ними.
4	Интенсиональные и экстенсиональные знания	Интенсиональные знания – знания о связях между объектами (их атрибутами) рассматриваемой предметной области. Экстенсиональные знания – свойства объектов, их состояния, значения свойств в пространстве и динамике.
5	Глубинные и поверхностные знания	Глубинные знания содержат подробные сведения о структуре предметной области, законах поведения структуры в целом и отдельных ее элементов, достоверные и полные отражения взаимосвязей элементов структуры и т.д. Например: подробные сведения об устройстве компьютера или мобильного телефона, позволяющие производить проектирование их или ремонт. Поверхностные знания касаются лишь внешних свойств и связей с рассматриваемым объектом(ами).
6	Жесткие и мягкие знания	Жесткие знания отражают системы или объекты с четко выраженными свойствами, связями, поведением, которые легко описываются качественными и количественными признаками, например описываются логико-дедуктивной системой показателей. Мягкие знания отображают соответственно системы и объекты с трудно поддающимися описанию или формализации свойствами и связями. Дают нечеткие, размытые решения и множественность рекомендаций.

Процессы получения знаний

Следует различать два различных процесса получения знаний.

Первый - это «**извлечение**» их из живого источника - **эксперта**, специалиста с целью их идентификации и возможной формализации, помещения в **базу знаний** и построения на этой основе **экспертных систем**, а также в других целях. Такой процесс относят к инженерии знаний.

Второй - это «**добыча**» скрытых от пользователя знаний **из данных**, помещенных в различного рода компьютерные информационные системы, в том числе базы данных различного назначения, информационные хранилища. Процесс второго рода называют **Data mining** – используют русский перевод «**интеллектуальный анализ**».

2. Структура и назначение экспертных систем.

Экспертная система (ЭС) – это программный продукт, позволяющий имитировать творческую деятельность или усиливать интеллектуальные возможности пользователя при поиске решения задачи в конкретной предметной области, используя в основном эвристические знания специалистов-экспертов (т. е. накопленный ранее опыт) и некоторый логический механизм вывода.

Области применения ЭС:

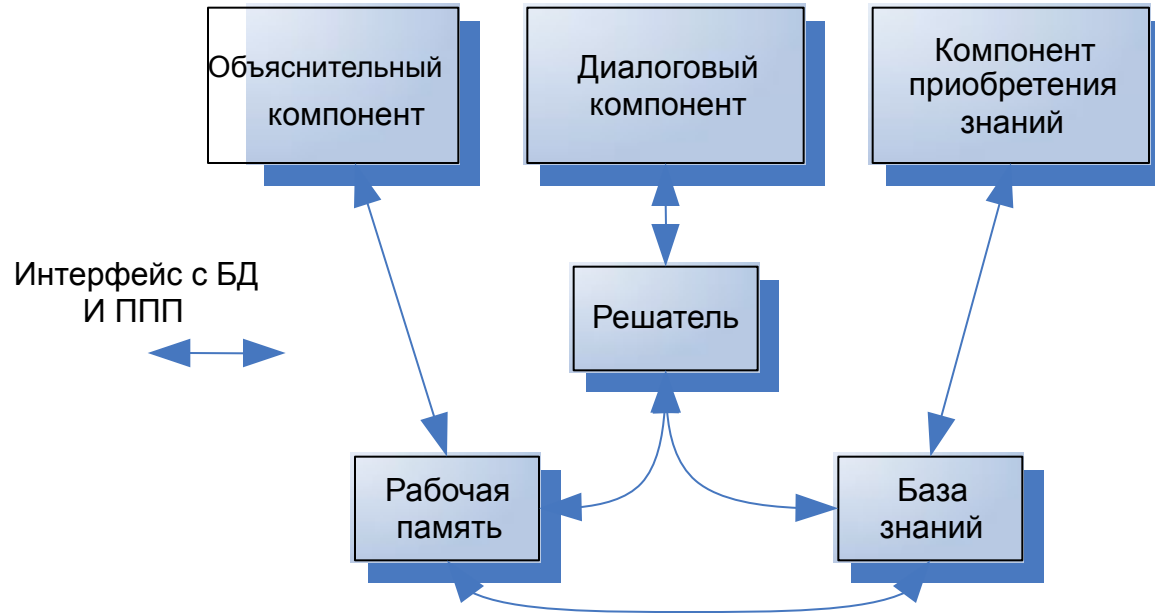
- медицинская диагностика,
- прогнозирование,
- планирование,
- интерпретация,
- контроль и управление,
- диагностика неисправностей,
- обучение,
- финансовое планирование,
- торговля.

Исторически ЭС принято делить на интеллектуальные системы первого и второго поколений.

ЭС первого поколения	ЭС второго поколения
■ знания системы в целом представлены только знаниями эксперта, функция накопления знаний не предусмотрена; ■ методы представления знаний позволяли описывать только статические предметные области; ■ модели представления знаний ориентированы на простые предметные области.	■ накопление и расширение базы знаний; ■ решение задач динамических предметных областей; ■ проведение анализа нечисловых данных; ■ генерация новых и отбраковка ненужных гипотез; ■ оценка достоверности фактов; ■ самостоятельное пополнение базы знаний; ■ контроль над непротиворечивостью базы знаний; ■ выдача собственных заключений на основе прецедентов; ■ решение новых, ранее не рассматривавшихся, задач.

Структура экспертных систем

Статическая экспертная система

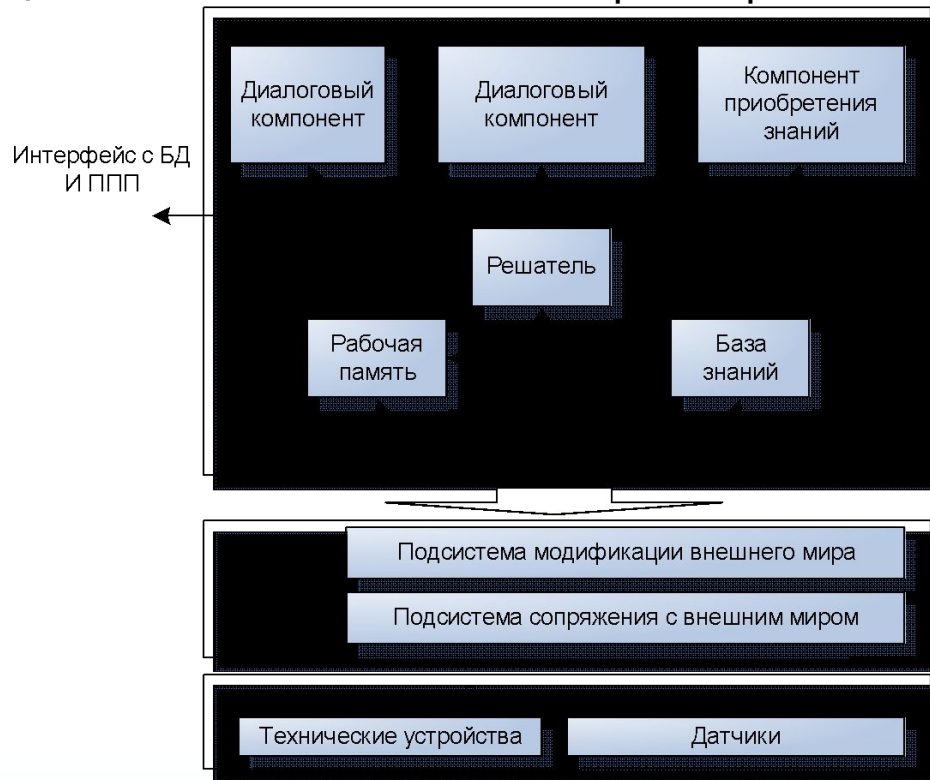


Рабочая память (база данных)	Предназначена для хранения исходных и промежуточных данных решаемой в текущий момент задачи.	Компонент приобретения знаний	Предназначен для автоматизированного наполнения экспертной системы знаниями, осуществляемый пользователем-экспертом.
База знаний	Предназначена для хранения субъективных эвристических знаний экспертов, описывающих рассматриваемую предметную область.	Объяснительный компонент	Предназначен для объяснения, как система получила решение задачи (или почему она не получила решение) и какие знания она при этом использовала, что облегчает эксперту тестирование системы и повышает доверие пользователя к полученному результату.
Решатель	Составляет наиболее важную часть экспертной системы. Решатель, используя исходные данные из рабочей памяти и знания из базы знаний, формирует такую последовательность правил, которые, будучи примененными к исходным данным, приводят к решению задачи.	Диалоговый компонент	Предназначен для организации дружественного общения с пользователем как в ходе решения задач, так и в процессе приобретения знаний и объяснения результатов работы.

Структура экспертных систем

Динамическая экспертная система

В динамической ЭС по сравнению со статической вводятся два компонента: **подсистема моделирования внешнего мира** и **подсистема связи с внешним окружением**. Последняя осуществляет связи с внешним миром через систему датчиков и контроллеров.



Кроме того, традиционные компоненты статической ЭС (база знаний и решатель) претерпевают существенные изменения, чтобы отразить временную логику происходящих в реальном мире событий.

Режимы работы экспертной системы

Режим приобретения знаний	■ В этом режиме эксперт, используя компонент приобретения знаний, наполняет систему знаниями, которые позволяют ЭС в режиме решения самостоятельно (без эксперта) решать задачи из проблемной области. ■ Эксперт описывает проблемную область в виде совокупности данных и правил. Данные определяют объекты, их характеристики и значения, существующие в области экспертизы. Правила определяют способы манипулирования с данными, характерные для рассматриваемой области. ■ Режиму приобретения знаний в традиционном подходе к разработке программ соответствуют этапы алгоритмизации, программирования и отладки, выполняемые программистом. Таким образом, в отличие от традиционного подхода в случае ЭС разработку программ осуществляет не программист, а эксперт (с помощью ЭС), не владеющий программированием.
Режим консультации	■ Общение с ЭС осуществляет конечный пользователь, которого интересует результат и (или) способ его получения. ■ В зависимости от назначения ЭС пользователь может не быть специалистом в данной проблемной области (в этом случае он обращается к ЭС за результатом, не умея получить его сам), или быть специалистом (в этом случае пользователь может сам получить результат, но он обращается к ЭС с целью либо ускорить процесс получения результата, либо возложить на ЭС рутинную работу). ■ Данные о задаче пользователя после обработки их диалоговым компонентом поступают в рабочую память. Решатель на основе входных данных из рабочей памяти, общих данных о проблемной области и правил из БЗ формирует решение задачи. ЭС при решении задачи не только исполняет предписанную последовательность операций, но и предварительно формирует ее. Если реакция системы не понятна пользователю, то он может потребовать объяснения: "Почему система задает тот или иной вопрос?", "как ответ, собираемый системой, получен?".

Модели представления знаний в экспертных системах

К основным моделям представления знаний относятся:

- логические модели;
- продукционные модели;
- сетевые модели;
- фреймовые модели.

Логические модели

В основе моделей такого типа лежит формальная система, задаваемая четверкой вида:

$M = \langle T, P, A, B \rangle$.

Множество T	Множество <i>базовых элементов</i> различной природы, например слов из некоторого ограниченного словаря, деталей детского конструктора, входящих в состав некоторого набора и т.п.
Множество P	Множество <i>синтаксических правил</i> . С их помощью из элементов T образуют <i>синтаксически правильные совокупности</i> . Например, из слов ограниченного словаря строятся синтаксически правильные фразы, из деталей детского конструктора с помощью гаек и болтов собираются новые конструкции.
Множество A	Множество <i>аксиом</i> – подмножество синтаксически правильных совокупностей. Можно считать, что множество A образуют все информационные единицы, которые введены в базу знаний извне.
Множество B	Множество <i>правил вывода</i> . Применяя их к элементам A , можно получать новые синтаксически правильные совокупности, к которым снова можно применять правила из B . С помощью правил вывода B из аксиом A выводятся новые производные знания.

Т.е. данная формальная система представляет собой генератор порождения новых знаний, образующих множество **выводимых** в данной системе **знаний**. Это свойство логических моделей делает их притягательными для использования в базах знаний. Оно позволяет хранить в базе лишь те знания, которые образуют множество **A**, а все остальные знания получать из них по правилам вывода.

Модели представления знаний в экспертных системах

Продукционные модели

Продукции являются наиболее популярными средствами представления знаний.

В общем виде под продукцией понимают выражение вида **A** $\square$ **B**.

Обычное прочтение продукции выглядит так: **ЕСЛИ A, ТО B**.

В состав системы продукций входит база правил (продукций), глобальная база данных и система управления.

- База правил – это область памяти, которая содержит совокупность знаний в форме правил вида ЕСЛИ – ТО.
- Глобальная база данных – область памяти, содержащая фактические данные (факты).
- Система управления формирует заключения, используя базу правил и базу данных.

Существуют два способа формирования заключений – прямые выводы и обратные выводы.

- В **прямых выводах** выбирается один из элементов данных, содержащихся в базе данных, и если при сопоставлении этот элемент согласуется с левой частью правила (посылкой), то из правила выводится соответствующее заключение и помещается в базу данных или выполняется действие, определяемое правилом, и соответствующим образом изменяется содержимое базы данных.
- В **обратных выводах** процесс начинается от поставленной цели. Если эта цель согласуется с правой частью правила (заключением), то посылка правила принимается за подцель или гипотезу. Этот процесс повторяется до тех пор, пока не будет получено совпадение подцели с данными.

При большом числе продукций в продукционной модели усложняется проверка непротиворечивости системы продукций, т.е. множества правил. Поэтому число продукций, с которыми работают современные системы ИИ, как правило, не превышают тысячи.

Модели представления знаний в экспертных системах

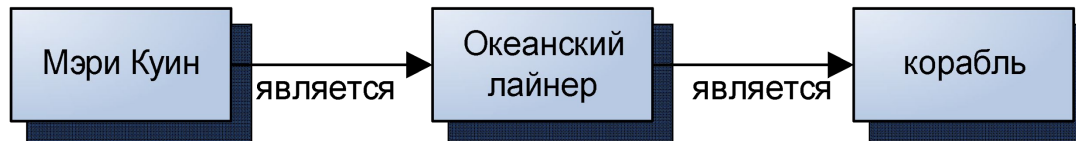
Сетевые модели

В основе моделей этого типа лежит конструкция называемая **семантической сетью**.

В самом общем случае семантическая сеть представляет собой информационную модель предметной области и имеет вид **графа**, вершины которого соответствуют объектам предметной области, а дуги – отношениям между ними.

Дуги могут быть определены разными методами, зависящими от вида представляемых знаний. Обычно дуги, используемые для представления иерархии, включают дуги типа "*множество*", "*подмножество*", "*элемент*". Семантические сети, используемые для описания естественных языков, используют дуги типа "*агент*", "*объект*", "*реципиент*".

В качестве простого примера рассмотрим предложения "Куин Мэри является океанским лайнером" и "Каждый океанский лайнер является кораблем". Они могут быть представлены через семантическую сеть. В этом примере используется важный тип дуг "является".



Семантический подход к построению систем ИИ находит применение в системах понимания естественного языка, в вопросно-ответных системах, в различных предметно – ориентированных системах.

Модели представления знаний в экспертных системах

Фреймовые модели

В отличие от моделей других типов во фреймовых моделях фиксируется жесткая структура информационных единиц, называемых фреймами.

Фрейм является формой представления некоторой ситуации, которую можно (или целесообразно) описывать некоторой совокупностью понятий и сущностей.

Фрейм имеет определенную внутреннюю структуру, состоящую из множества элементов, называемых **слотами**. Каждый слот представляется определенной структурой данных. В значение слота подставляется конкретная информация – **значение слота**, относящаяся к объекту, описываемому этим фреймом.

Значением слота может быть практически что угодно: числа, формулы, тексты на естественном языке или программы, правила вывода или ссылки на другие слоты данного фрейма или других фреймов. В качестве значения слота может выступать набор слотов более низкого уровня, что позволяет реализовывать во фреймовых представлениях "принцип матрешки".

Связи между фреймами задаются значениями специального слота с именем "Связь". Часть специалистов по системам ИИ считают, что нет необходимости выделять фреймовые модели представления знаний, так как в них объединены все основные особенности моделей остальных типов.

Слот	Значение слота
Фрейм:	Имя
Имя слота:	Значение слота
Класс:	Животное
Структурные элементы	Голова, шея, руки,...
Рост	30 – 220 см.
Масса	1 – 200 кг.
Хвост	Нет
Язык	Русский, английский, ...
Фрейм аналогии	Обезьяна

Простой пример фрейма описания человека

3. Добыча знаний (Data Mining)

Определение добычи знаний (Data Mining)

Data Mining переводится как "добыча" или "раскопка данных". Нередко рядом с Data Mining встречаются слова "обнаружение знаний в базах данных" (knowledge discovery in databases) и "интеллектуальный анализ данных".

Data Mining - это процесс обнаружения в сырых данных

- ранее неизвестных
- нетривиальных
- практически полезных
- и доступных интерпретации знаний,
необходимых для принятия решений в различных сферах человеческой деятельности.

Термин Data Mining обозначает не столько конкретную технологию, сколько сам процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов: кластеризации, создания субвыборок, регрессионного и корреляционного анализа.

Цель этого поиска — представить данные в виде, четко отражающем бизнес-процессы, а также построить модель, при помощи которой можно прогнозировать процессы, критичные для планирования бизнеса (например, динамику спроса на те или иные товары или услуги либо зависимость их приобретения от каких-то характеристик потребителя).

Области применения Data mining

Маркетинг	Рыночная сегментация, идентификация целевых групп, построение профиля клиента
Банковское дело	Анализ кредитных рисков, привлечение и удержание клиентов, управление ресурсами
Кредитные компании	Обнаружение подлогов, формирование "типичного поведения" обладателя кредитки, анализ достоверности клиентских счетов
Страховые компании	Привлечение и удержание клиентов, прогнозирование финансовых показателей
Розничная торговля	Анализ деятельности торговых точек, построение профиля покупателя, управление ресурсами
Биржевые трейдеры	Выработка оптимальной торговой стратегии, контроль рисков
Телекоммуникация и энергетика	Привлечение клиентов, ценовая политика, анализ отказов, предсказание пиковых нагрузок, прогнозирование поступления средств
Налоговые службы и аудиторы	Обнаружение подлогов, прогнозирование поступлений в бюджет
Фармацевтические компании	Предсказание результатов будущего тестирования препаратов, программы испытания
Медицина	Диагностика, выбор лечебных воздействий, прогнозирование исхода хирургического вмешательства
Управление производством	Контроль качества, материально-техническое обеспечение, оптимизация технологического процесса
Ученые и инженеры	Построение эмпирических моделей, основанных на анализе данных, решение научно-технических задач

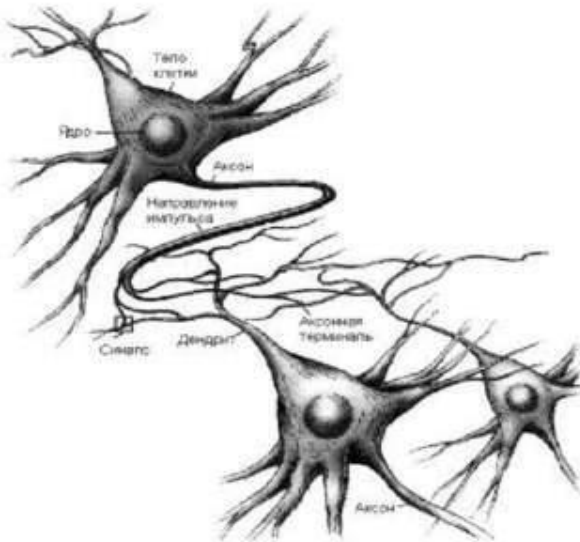
Типы закономерностей, выявляемых методами Data Mining

№	Название	Описание
1	Ассоциация	Высокая вероятность связи событий друг с другом
2	Последовательность	Высокая вероятность цепочки связанных во времени событий
3	Классификация	Имеются признаки, характеризующие группу, к которой принадлежит то или иное событие или объект
4	Кластеризация	Закономерность, сходная с классификацией и отличающаяся от нее тем, что сами группы при этом не заданы - они выявляются автоматически в процессе обработки данных
5	Прогнозирование	Установление закономерности между связанными во времени событиями и их использование для вычисления прогнозного значения

Методы Data Mining

1. Нейронные сети

Нейронные сети представляют собой большой класс систем, условно имитирующих нервную ткань из нейронов.

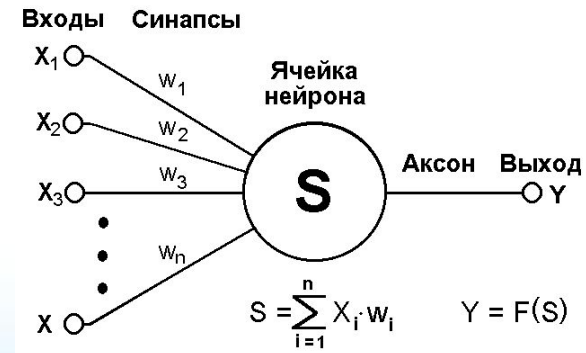


Нейрон (нервная клетка) является особой биологической клеткой, которая обрабатывает информацию. Она состоит из **тела клетки** и двух типов внешних древовидных ветвей: **аксона** и **дендритов**. Тело клетки включает **ядро**, которое содержит информацию о наследственных свойствах, и плазму, обладающую молекулярными средствами для производства необходимых нейрону материалов. Нейрон получает сигналы (импульсы) от других нейронов через дендриты (приемники) и передает сигналы, сгенерированные телом клетки, вдоль аксона (передатчик), который в конце разветвляется на волокна. На окончаниях этих волокон находятся **синапсы**.

Иллюстрация строения нейрона

Математическая модель нейрона представляет собой некоторый универсальный нелинейный элемент с возможностью широкого изменения и настройки его характеристик.

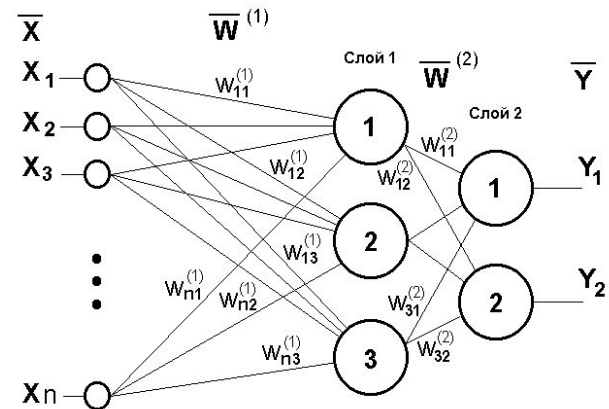
Математическая модель нейрона



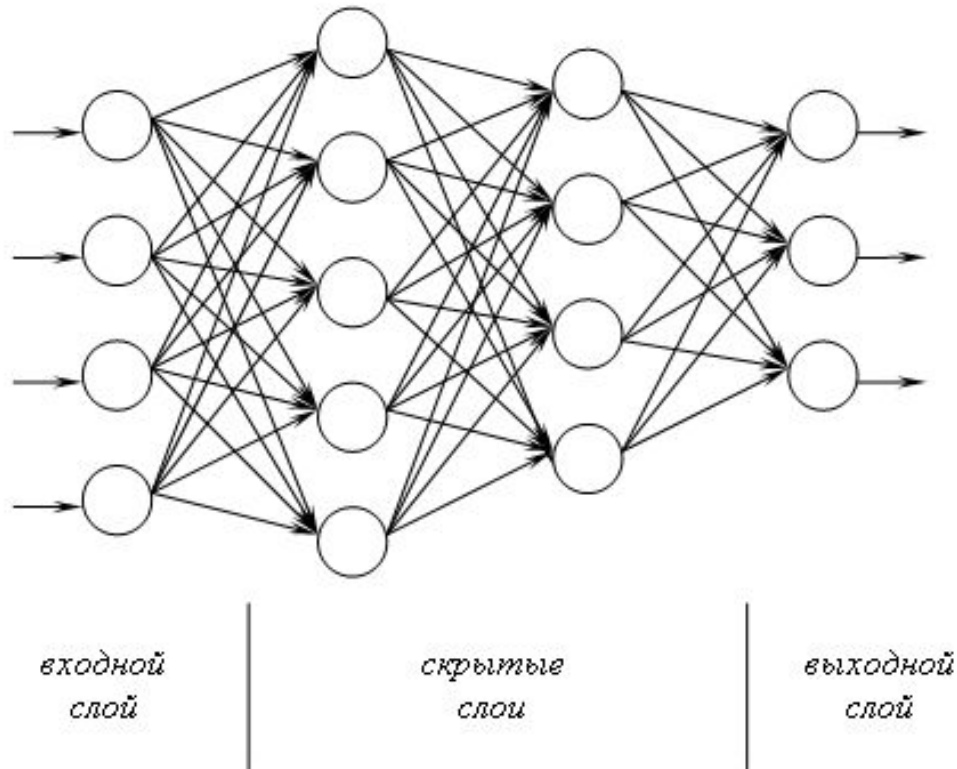
Методы Data Mining

1. Нейронные сети (продолжение)

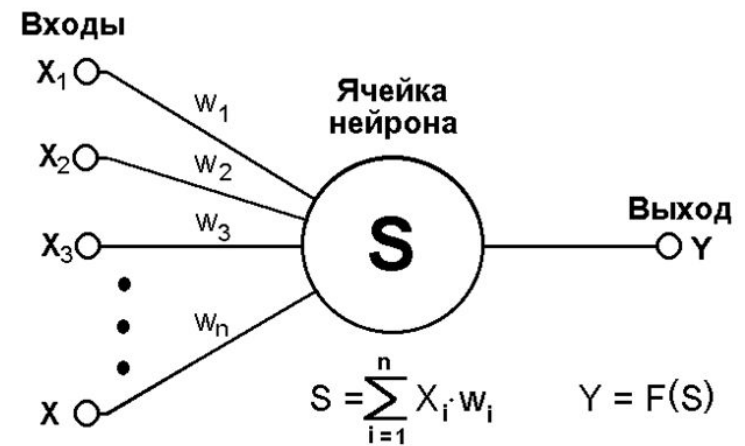
- В одной из распространенных архитектур, **двухслойном персептроне**, имитируется работа нейронов в составе иерархической сети, где каждый нейрон более высокого уровня соединен своими входами с выходами нейронов нижележащего слоя.
- На нейроны самого нижнего слоя подаются значения входных параметров, на основе которых нужно принимать какие-то решения, прогнозировать развитие ситуации и т.д.
- Эти значения рассматриваются как сигналы, передающиеся в следующий слой, ослабляясь или усиливаясь в зависимости от числовых значений (весов), приписываемых межнейронным связям.
- В результате на выходе нейрона самого верхнего слоя вырабатывается некоторое значение, которое рассматривается как ответ - реакция всей сети на введенные значения входных параметров.
- Для того чтобы сеть можно было применять в дальнейшем, ее прежде надо "натренировать" на полученных ранее данных, для которых известны и значения входных параметров, и правильные ответы на них.
- Тренировка состоит в подборе весов межнейронных связей, обеспечивающих наибольшую близость ответов сети к известным правильным ответам.



Нейросетевые технологии



Многослойная нейронная сеть

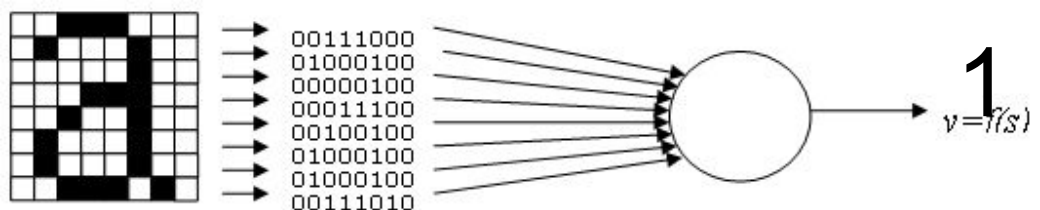


Модель персептрона

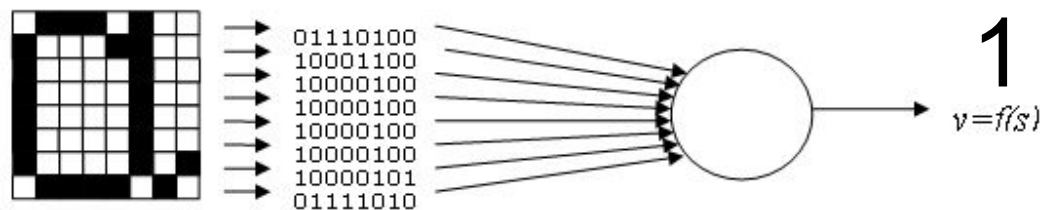
Проблемы практического использования нейросетей

Определение оптимальной архитектуры сети. Выбор активационной функции и алгоритма обучения.

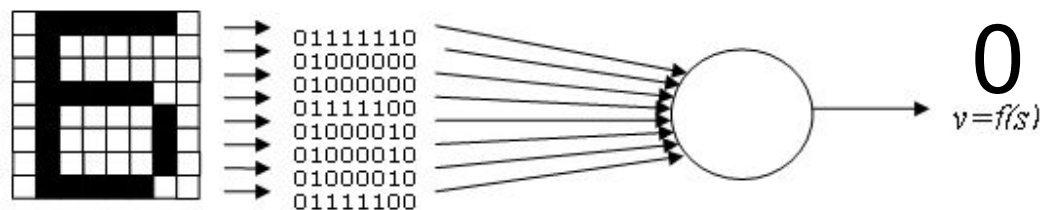
Место нейросетевых технологий как альтернатива и дополнение к алгоритмическим методам обработки информации.



Требования к данным со стороны нейросетевых технологий.



Возможности и методы предобработки данных (понижение размерности, исключение незначущих факторов, восстановление данных, корректировка аномалий).



Состояние и тенденции развития рынка ИИС.

Обучение нейросети с «учителем»

Аналитическая платформа Deductor

Возможности, структура и схема обработки данных.

Методы обработки: извлечение, очистка, манипулирование, моделирование, прогнозирование, кластеризация, поиск закономерностей, обнаружение знаний, добычи данных и др.

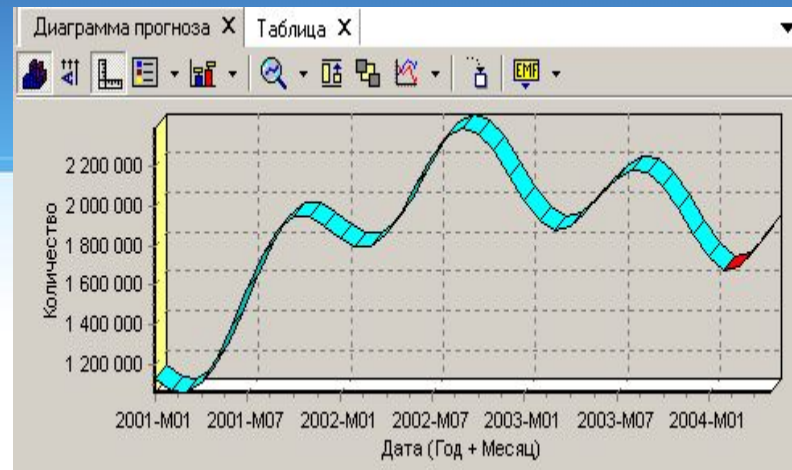
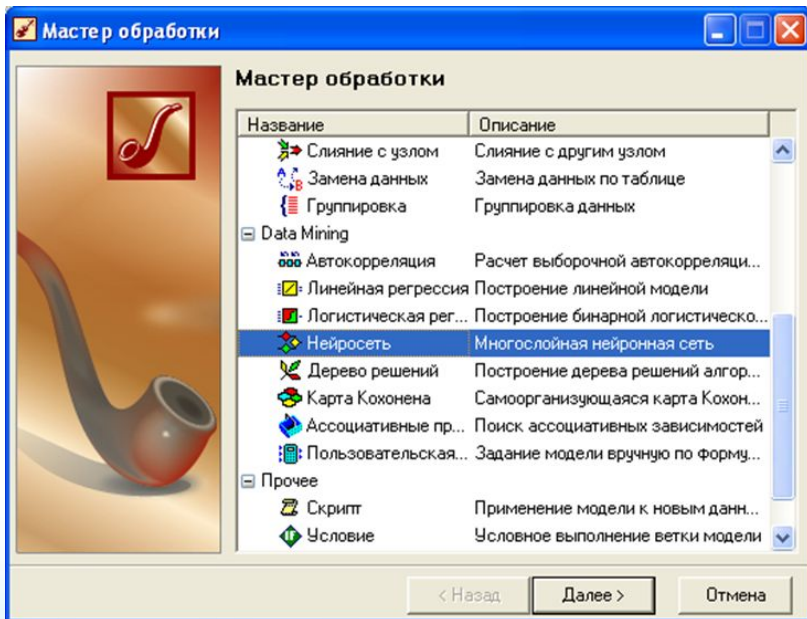
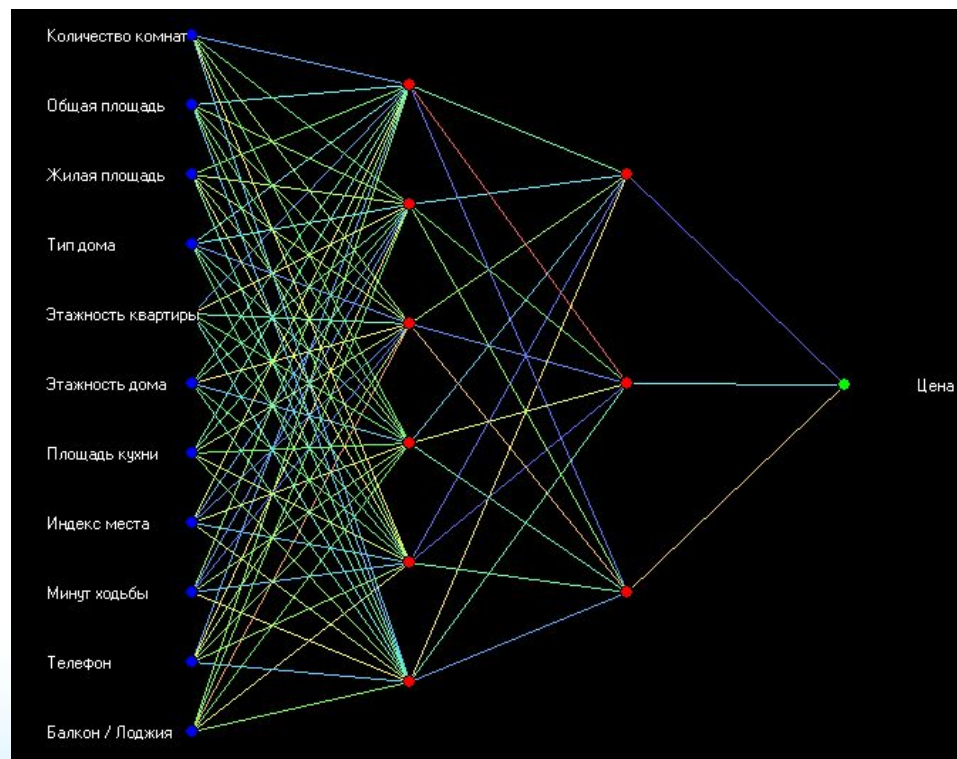


Диаграмма прогноза



Мастер обработки пакета Deductor



Нейросеть

Методы Data Mining

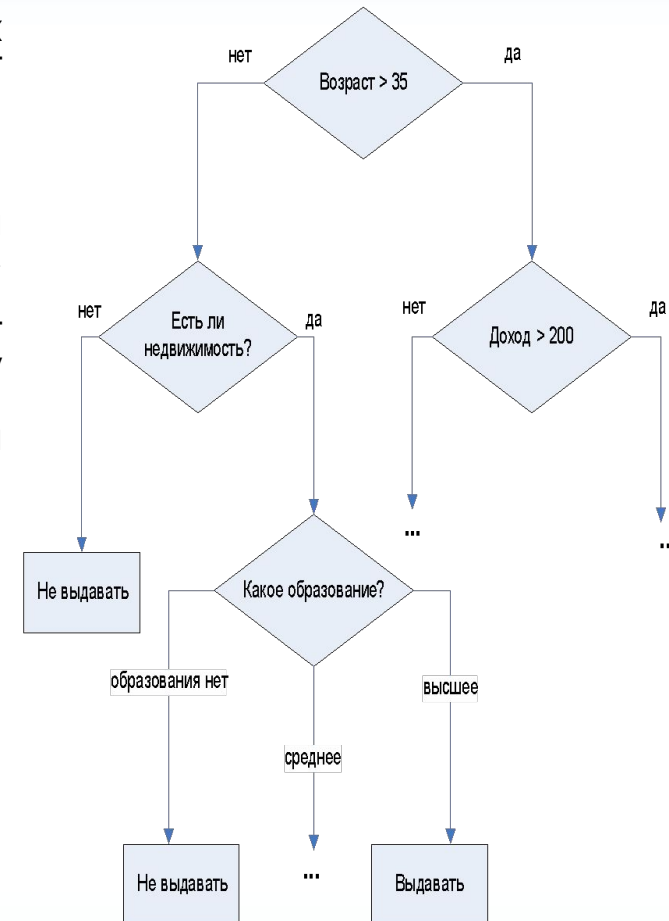
2. Деревья решений

Деревья решений являются одним из наиболее популярных подходов к решению задач добычи данных. Они создают иерархическую структуру классифицирующих правил типа "если...то...", имеющую вид дерева.

Для принятия решения, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Вопросы имеют вид "значение параметра А больше х?". Если ответ положительный, осуществляется переход к правому узлу следующего уровня, если отрицательный - то к левому узлу; затем снова следует вопрос, связанный с соответствующим узлом.

Деревья решений:

- Анализируют решения, принятые человеком (выдать - не выдать кредит, принять - не принять на работу, купить - не купить, закупить - не закупить и т.д.)
- Автоматизируют процесс принятия новых решений на основе исторических данных
- Используются в случае формализованных процедур принятия решений в организации



Методы Data Mining

3. Системы рассуждений на основе аналогичных случаев (case based reasoning - CBR)

Идея систем CBR - на первый взгляд крайне проста. Для того чтобы сделать прогноз на будущее или выбрать правильное решение, эти системы находят в прошлом близкие аналоги наличной ситуации и выбирают тот же ответ, который был для них правильным. Поэтому этот метод еще называют методом "ближайшего соседа" (nearest neighbour). В последнее время распространение получил также термин memory based reasoning, который акцентирует внимание, что решение принимается на основании всей информации, накопленной в памяти.

Например, при рассмотрении нового клиента банка, его атрибуты сравниваются со всеми существующими клиентами данного банка (доход, возраст и т.д.). Множество "ближайших соседей" потенциального клиента банка выбирается на основании ближайшего значения дохода, возраста и т.д.

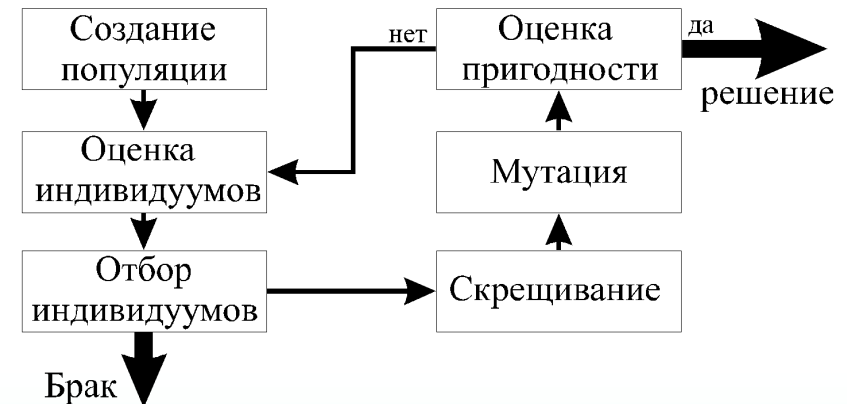
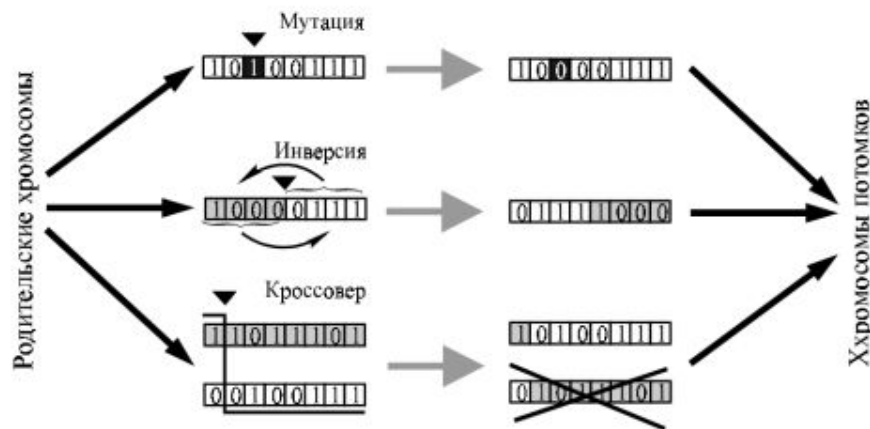
Методы Data Mining

4. Генетические алгоритмы

Методы генетических алгоритмов в какой-то степени имитирует процесс естественного отбора в природе.

Шаги применения генетических алгоритмов:

1. Кодировка исходных логических закономерностей в базе данных, которые именуют **хромосомами**, а весь набор таких закономерностей называют **популяцией** хромосом.
2. Популяция обрабатывается с помощью процедур репродукции, изменчивости (мутаций), генетической композиции. Эти процедуры имитируют биологические процессы. Наиболее важные среди них: **случайные мутации** данных в индивидуальных хромосомах, **переходы** (кроссинговер) и **рекомбинация** генетического материала, содержащегося в индивидуальных родительских хромосомах (аналогично гетеросексуальной репродукции), и **миграции генов**.
3. В ходе работы процедур на каждой стадии эволюции получают популяции со все более совершенными индивидуумами.



Методы Data Mining

5. Эволюционное программирование

Эволюционное программирование – сегодня самая молодая и наиболее перспективная ветвь добычи данных.

Суть метода заключается в том, что гипотезы о виде зависимости целевой переменной от других переменных формулируются системой в виде программ на некотором внутреннем языке программирования. Если это универсальный язык, то теоретически на нем можно выразить зависимость любого вида.

Процесс построения этих программ строится подобно эволюции в мире программ (этим метод похож на генетические алгоритмы). Когда система находит программу, достаточно точно выражающую искомую зависимость, она начинает вносить в нее небольшие модификации и отбирает среди построенных таким образом дочерних программ те, которые повышают точность. Таким образом, система "выращивает" несколько генетических линий программ, которые конкурируют между собой в точности выражения искомой зависимости. Специальный транслирующий модуль переводит найденные зависимости с внутреннего языка системы на понятный пользователю язык (математические формулы, таблицы и пр.), делая их легкодоступными. Для того чтобы сделать полученные результаты еще понятнее для пользователя-нематематика, имеется богатый арсенал разнообразных средств визуализации обнаруживаемых зависимостей.

Методы Data Mining

6. Нечеткая логика (fuzzy logic)

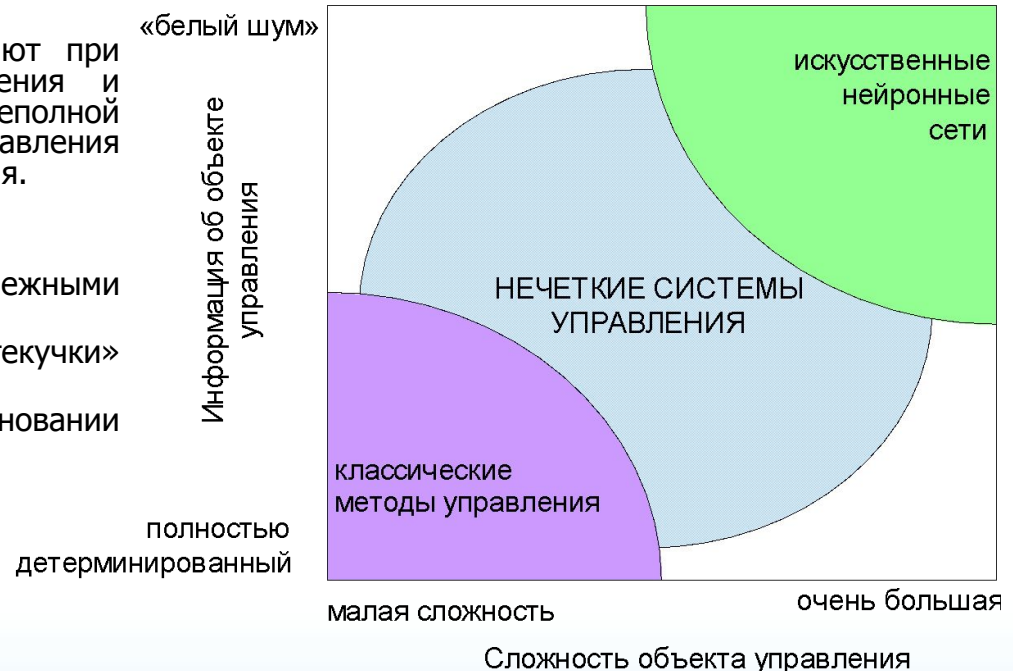
В окружающем нас мире очень редко приходится сталкиваться с задачами, лишенными какого-либо элемента неопределенности. Управленческие решения практически всегда приходится принимать в условиях частичного отсутствия необходимой информации.

Нечеткая логика и алгебра применяются для анализа таких наборов данных, когда невозможно причислить данные к какой-либо группе. Мы можем только отнести данные к какой-либо группе с некоторой вероятностью находящейся в интервале от 0 до 1, но не принимающей крайние значения. Четкая логика манипулирует результатами, которые могут быть либо истиной, либо ложью. Нечеткая логика применяется в тех случаях, когда необходимо манипулировать степенью "может быть" в дополнении к "да" и "нет".

Классические методы управления хорошо работают при полностью детерминированном объекте управления и детерминированной среде, а для систем с неполной информацией и высокой сложностью объекта управления оптимальными являются нечеткие методы управления.

Нечеткая логика:

- Оптимизация закупок товаров, управления денежными потоками и т.п.
- Сведение к минимуму человеческого фактора «текучки» при принятии решений
- Получение оперативных рекомендаций на основании правил, составленных экспертами



Методы Data Mining

7. Статистические пакеты

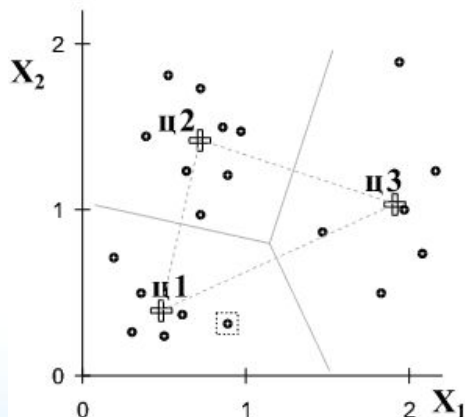
Последние версии почти всех известных статистических пакетов включают наряду с традиционными статистическими методами также элементы Data Mining.

Но основное внимание в них уделяется все же классическим методикам - корреляционному, регрессионному, факторному анализу и другим.

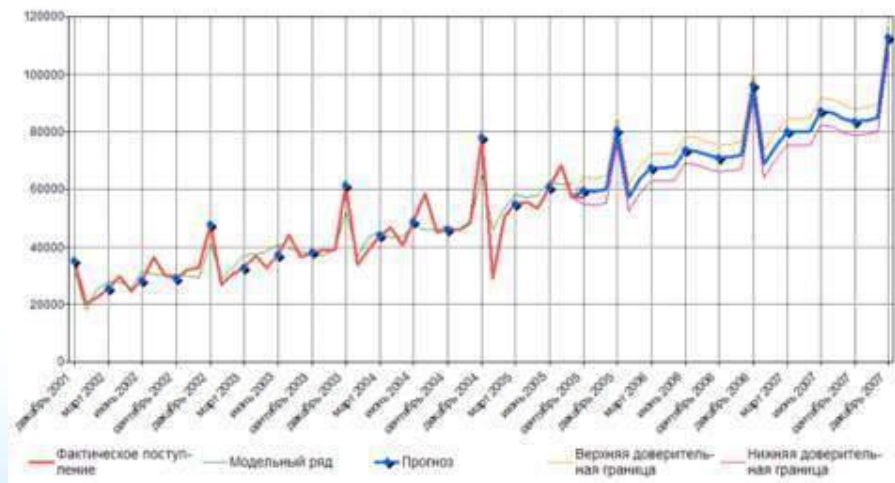
Недостатком систем этого класса считают требование к специальной подготовке пользователя. Также отмечают, что мощные современные статистические пакеты являются слишком "тяжеловесными" для массового применения в финансах и бизнесе.

Большинство методов, входящих в состав пакетов опираются на статистическую парадигму, в которой главными фигурантами служат усредненные характеристики выборки. А эти характеристики, как указывалось выше, при исследовании реальных сложных жизненных феноменов часто являются фиктивными величинами.

Пример кластеризации



Пример прогнозирования



Визуализация инструментов Data Mining

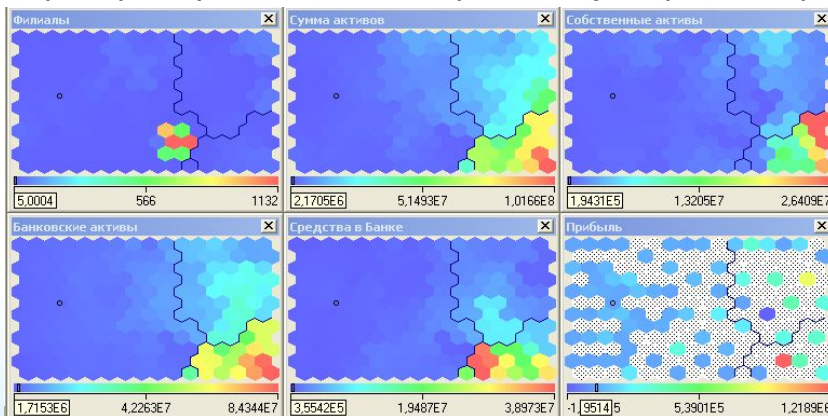
Каждый из алгоритмов Data Mining использует определенный подход к визуализации. В ходе использования каждого из методов Data Mining, а точнее, его программной реализации, получаются визуализаторы, при помощи которых нам удаётся интерпретировать результаты, полученные в результате работы соответствующих методов и алгоритмов.

- Для деревьев решений это визуализатор дерева решений, список правил.
- Для нейронных сетей в зависимости от инструмента это может быть топология сети, график изменения величины ошибки, демонстрирующий процесс обучения.
- Для линейной регрессии в качестве визуализатора выступает линия регрессии.
- Для кластеризации : дендрограммы, диаграммы рассеивания.
- Диаграммы и графики рассеивания часто используются для оценки качества работы того или иного метода.

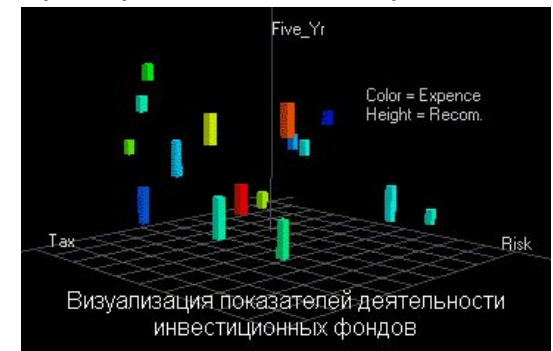
Все эти способы визуального представления или отображения данных могут выполнять одну из функций:

- являются иллюстрацией построения модели (например, представление структуры (графа) нейронной сети);
- помогают интерпретировать полученный результат;
- являются средством оценки качества построенной модели.

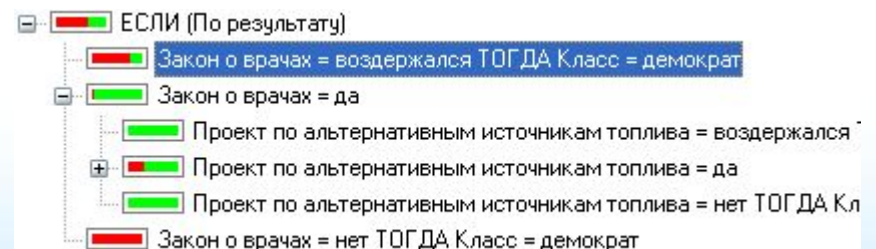
Пример визуализации кластеризации (алгоритм - карты Кохонена)



Пример визуализации классификации



Пример визуализации дерева решений



Литература

- Белов В.С. Информационно-аналитические системы. Основы проектирования и применения: учебное пособие, руководство, практикум/ Московский государственный университет экономики, статистики и информатики – М., 2004. – 116 с.
- Курс лекций по дисциплине "Системы искусственного интеллекта" /// <http://www.mari-el.ru/mmlab/home/AI/index.html>
- С.Л. Сотник. Проектирование систем искусственного интеллекта / Курс лекций // Интернет университет информационных технологий /// <http://www.intuit.ru/department/expert/artintell/9/1.html>
- В.Л. Афонин, В.А. Макушкин. Интеллектуальные робототехнические системы / Курс лекций // Интернет университет информационных технологий /// <http://www.intuit.ru/department/human/isrob/6/1.html>
- Что такое Data Mining // Intersoft Lab - 28 декабря 2003 г. /// <http://citcity.ru/11139/>
- Л. Е. Карпов, В. Н. Юдин. Методы добычи данных при построении локальной метрики в системах вывода по прецедентам // CitForum – 2007 г. /// http://www.citforum.ru/consulting/BI/data_mining/3.shtml
- Вячеслав Дюк. Data Mining - интеллектуальный анализ данных // Портал iTeam - Технологии корпоративного управления /// http://www.iteam.ru/publications/it/section_92/article_1448/
- Наталия Елманова. Введение в Data Mining // Портал iTeam - Технологии корпоративного управления /// http://www.iteam.ru/publications/it/section_92/article_1649/
- Анна Иванова. Аналитическая обработка данных // ВУТЕ Россия - №2 2006г. /// <http://www.bytemag.ru/?ID=608632>
- Сергей Гриняев. Нечеткая логика в системах управления // Компьютерра - №38, 08 октября 2001 года /// <http://www.computerra.ru/offline/2001/415/13052/>