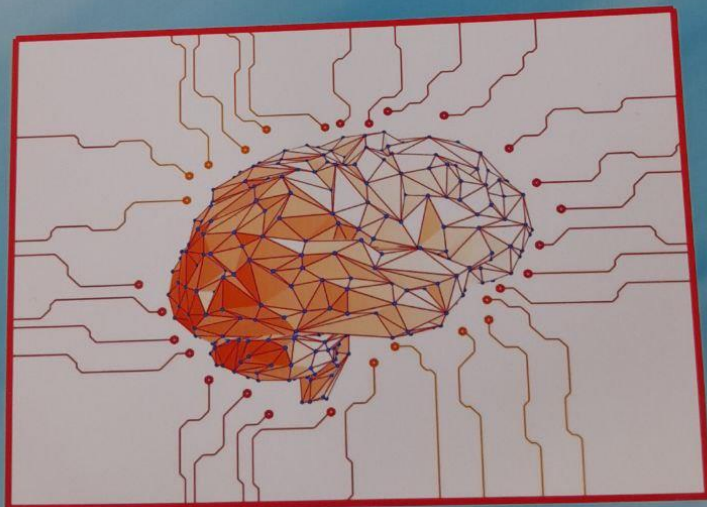


МУСАЕВ М.М., ХУЖАЯРОВ И.Ш.

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НЕЙРОННЫЕ СЕТИ



МИНИСТЕРСТВО ЦИФРОВЫХ ТЕХНОЛОГИЙ  
РЕСПУБЛИКИ УЗБЕКИСТАН

ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ  
ТЕХНОЛОГИЙ ИМЕНИ МУХАММАДА АЛ-ХОРАЗМИЙ

МУСАЕВ М.М., ХУЖАЯРОВ И.Ш.

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НЕЙРОННЫЕ СЕТИ

Рекомендовано в качестве учебного пособия Ташкентским  
университетом информационных технологий имени Мухаммада  
Аль-Хорезми

Ташкент  
“METHODIST NASHRIYOTI”  
2024

УДК: 004.8(075.8)  
ББК: 32.813я7  
М 916

Мусаев М.М., Хужаяров И.Ш.  
Искусственный интеллект и нейронные сети. Учебное пособие.  
– Ташкент: “METHODIST NASHRIYOTI”, 2024. – 268 с.

В книге рассмотрены наиболее широко применяемые методы решения задач искусственного интеллекта на основе баз знаний. Приведены методы эвристического программирования и поиска в пространстве состояний и пространстве задач. Рассмотрены модели представления знаний и даны примеры их построения. Даны основы современной теории нечетких множеств и наиболее применяемые функции принадлежности. Отдельная глава посвящена экспертным системам, их классификации и проектированию. Представлены основные понятия и виды машинного обучения, исследованы задачи классификации и алгоритм регрессии.

Второй раздел учебного пособия посвящен нейронным сетям. Рассмотрены принципы функционирования нейронных сетей с одним и многими слоями. Глава 9 посвящена обучению нейронных сетей, вопросам корректировки весовых коэффициентов, алгоритму обратного распространения ошибок. Детально рассмотрены архитектуры рекуррентных, сверточных и автоэнкодерных сетей на основе персептрона. Наряду с этим представлены структуры релаксационных нейронных сетей, а также глубоких нейронных сетей. Отдельная глава книги посвящена применению нейронных сетей на примере распознавания речи.

#### Рецензенты:

профессор кафедры «Автоматизация производственных процессов» ТГТУ  
д.т.н. Абдукадиров А.А.  
зав кафедрой «Компьютер инжиниринг» ТУИТ, доцент Рахимов М.Ф.

Рекомендовано к публикации на основании приказа  
Ташкентского университета информационных технологий имени  
Мухаммада Аль-Хорезми от 25 апреля 2023 года № 439.

ISBN 978-9910-03-159-5

© Мусаев М.М., Хужаяров И.Ш., 2024.  
© “METHODIST NASHRIYOTI”, 2024.

## ПРЕДИСЛОВИЕ

Только небольшую часть своих знаний человек может сформулировать формальным способом. Большая часть интуитивных возможностей человека, которые необходимы ему для эффективной работы систем обработки данных, остается недоступной из-за отсутствия средств извлечения и представления знаний. Системы, ядром которых является база знаний и модель исследуемой предметной области, называют интеллектуальными. Чаще всего интеллектуальные системы применяются для решения сложных проблем, связанных с использованием слабо формализованных и плохо структурированных задач, где преобладает логическая и эвристическая обработка информации. Системы искусственного интеллекта отличаются от обычных систем обработки данных тем, что в них в основном используются символичный способ представления и эвристический поиск решения, а не исполнение известного алгоритма.

Идея создания искусственного интеллекта связана с постоянным стремлением человека переложить решение сложных задач на компьютерного помощника. Способ реализовать это заключается в моделировании с помощью компьютерных средств интеллектуальных способностей человека. Здесь под интеллектом следует понимать способность мозга решать задачи путем приобретения, запоминания и целенаправленного преобразования знаний в процессе обучения, на опыте и адаптации к разнообразным условиям.

Правильнее воспринимать искусственный интеллект как совокупность методов, реализация которых на компьютере позволяет получать результаты близкие к порождаемым человеческим мышлением.

Первоначально компьютеры могли выполнять лишь простейшие интеллектуальные функции, которые хорошо формализуются с помощью заданного алгоритма, и оперировали с числами. С развитием вычислительной техники появилась возможность обработки символической информации, текстов, изображений. Этому способствовало также и появление языков программирования. Но в середине XX века стали появляться программы, выполняющие творческие функции интеллекта. Первыми такими программами были игровые программы. На начальной стадии развития искусственного интеллекта основой для исследования интеллектуальных, мыслительных процедур человека послужили игровые задачи, которые носили невычислительный характер (игры в

крестики-нолики, шашки, шахматы).

В таких областях, как автоматизация проектирования, автоматическое управление, военные и социально-экономические системы преобладает информация, представленная наборами фактов, гипотез, правил и закономерностей, сформулированных на качественном уровне. Среди таких подходов, применяемых различными исследователями, можно выделить эвристическое программирование, представление и обработка знаний, нечеткое моделирование, искусственные нейронные сети, эволюционное моделирование.

Возникновение и развитие искусственного интеллекта связано с решением следующих классов задач:

- игровые задачи (крестики-нолики, морской бой, шашки, шахматы, нарды);
- задачи доказательств теорем (доказательство теорем в исчислении высказываний в планиметрии, теории групп);
- задачи распознавания образов (анализ изображений, формирование набора признаков, описывающих объекты и процессы, нахождение решающего правила, по которому признаки относятся к тому или иному классу);
- машинный перевод (математическая лингвистика; методы автоматического морфологического и синтаксического анализа и синтеза речи и текста; разделение декларативных и процедурных знаний);
- задачи информационного поиска (методы поиска информации по образцу и типовым структурам).

К сфере решаемых интеллектуальными системами задач относятся задачи, обладающие, как правило, следующими особенностями:

- неизвестен или не может быть реализован алгоритм решения;
- если существует алгоритмическое решение, но его нельзя использовать из-за ограниченности ресурсов (время, память);
- задача не может быть сформулирована в числовой форме;
- цель нельзя выразить в терминах точно определенной целевой функции.

Разработка интеллектуальных систем, как правило, ведется в рамках одного или нескольких направлений искусственного интеллекта, которых в настоящее время существует целое множество.

Существует два основных направления в теории искусственного интеллекта. В первом традиционном направлении используются

методы логических рассуждений и символической обработки информации, второе, связанное с построением сетей, состоящих из нейронных элементов, опирается на биологические основы естественного интеллекта, что позволяет проектировать системы, способные к обучению и самоорганизации.

В данном учебнике рассматриваются оба направления теории искусственного интеллекта.

В первой главе сделана постановка проблемы искусственного интеллекта, введены основные понятия, представлены направления исследований, области применения, рассмотрено понятие знаний как основы искусственного интеллекта.

Во второй главе книги рассматриваются методы эвристического программирования, алгоритмы поиска решения в пространстве состояний и в пространстве задач. На примере игры в пятнадцать рассмотрена игровая модель эвристического программирования.

В третьей главе исследованы традиционные модели представления знаний – логическая, продукционная и фреймовая модели. Отдельно рассмотрены модели представления знаний в виде семантических сетей.

Четвертая глава посвящена теории нечетких множеств. Рассмотрены задачи описания, представления и интерпретации параметров нечетких множеств, даны выполняемые над нечеткими множествами вычислительные операции и стандартные формы функций принадлежности.

Экспертным системам посвящена пятая глава работы. Даны характеристики и базовые функции экспертных систем, структура и свойства экспертных систем. Кроме того, рассмотрены вопросы классификации и проектирования систем

# ВВЕДЕНИЕ В СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

## 1.1. Некоторые понятия искусственного интеллекта

Еще на заре развития вычислительной техники стояла проблема создания систем для различного рода интеллектуальной деятельности – так называемого искусственного интеллекта. Это связано с тем, что многие задачи не могут быть решены точными алгоритмическими методами. Другим аспектом данной проблемы является то, что искусственная система должна не просто функционировать по заданным алгоритмам, а должна быть (по аналогии с человеком) обучаемой, генерировать заложенные знания и создавать алгоритмы решения задач. Под интеллектом следует понимать способность мозга решать задачи путем приобретения, запоминания и целенаправленного преобразования знаний в процессе обучения на опыте и адаптации к разнообразным условиям [1,2].

Человеческий способ познания мира отличается от компьютера наличием двух параллельных систем познания: одна из них – рассудок, интеллект, а вторая система – восприятие и образное мышление. Эти две системы существуют у человека как одно, неразрывное целое. Компьютер же обладает одной системой познания, а именно символично-логической. Компьютеры позволили вынести часть алгоритмов из головы человека в среду, где они могут исполняться в миллионы раз быстрее. Соответственно, все, что удастся формализовать, как правило, переносится из головы человека в эту новую искусственную цифровую среду.

Компьютерные ресурсы, в отличие от остальных ресурсов, постоянно и очень быстро растут и дешевеют, т. е. является сегодня основным потенциальным источником экономического роста. Однако, пока что этот заложенный в компьютерах потенциал практически мало задействован, решаемые на нем задачи используют только часть вычислительных возможностей машин. В то же время есть задачи, где интеллектуальные возможности человека позволяют решать их намного быстрее за

счет эвристики, опирающейся на объем знаний. Для усиления «творческих» возможностей компьютеров их нужно наделить интеллектом, обучить логико-эвристическим приемам решения. Машинный интеллект способен раскрыть истинный потенциал компьютеров, позволяя им решать трудно формализуемые проблемы лучше, быстрее и дешевле человека.

С интеллектуализацией компьютеров связаны последние успехи в задачах машинного перевода, компьютерного зрения, распознавания речи, и других, где узко специализированный машинный интеллект достигает уровня человека и может его частично заменять.

«Искусственный интеллект» (ИИ) – комплекс технологических и программных решений, приводящих к результату, аналогичному интеллектуальной деятельности человека, и используемых для решения сложных, трудно формализуемых прикладных задач.

Интеллектуальной называется система, способная целенаправленно, в зависимости от состояния информационных входов, изменять параметры функционирования и сам способ своего поведения. При этом способ поведения зависит как от текущего состояния информационных входов, так и от предыдущих состояний системы. В упрощенном представлении это система, моделирующая на компьютере мышление человека. Искусственный интеллект – это ключевое направление технологического развития, которое будет определять будущее всего мира, а также создаст новые стандарты в образовании. Уже сегодня механизмы искусственного интеллекта обеспечивают в режиме реального времени быстрое принятие оптимальных решений на основе анализа больших объемов информации, что дает значительные преимущества в качестве и результативности. Такие разработки не имеют аналогов в истории по своему влиянию на экономику и на производительность труда, на эффективность управления, образования, здравоохранения и на повседневную жизнь людей.

Искусственный интеллект (Artificial Intelligence) как научное направление существует с 1956 года, когда британский математик Алан Тьюринг опубликовал свою статью «Can the Machine Think?» («Может ли машина мыслить?»). К первой интеллектуальной

системей относят созданную А.Ньюэллом и А.Тьюрингом программу «Логик-Теоретик», предназначенную для доказательства теорем исчисления высказываний. Слово «intelligence» означает «умение рассуждать разумно». В основе человеческой деятельности лежит мышление, а конечный результат, на который направлены мыслительные процессы человека, называется целью. Мысли, ведущие к конечному результату, не случайны, а строго обоснованы, каждый шаг на пути к главной цели имеет свою локальную цель. Цель заставляет человека думать.

Рассмотрим, каким образом человек решает свои задачи и чем пользуется при этом. Любая программная система, создаваемая в рамках искусственного интеллекта, всегда ориентирована на использование знаний. Знания, выраженные на естественном языке, черпаются из книг, статей и других источников и в том виде, в котором содержатся в этих источниках, не могут быть использованы для обработки на компьютере. Требуется выбрать подходящий способ их формализации (представления) для получения возможности обработки знаний на вычислительных машинах. Сама обработка знаний на компьютере заключается в получении по определенным правилам вывода других знаний на основе имеющихся. Первичными базовыми понятиями искусственного интеллекта являются понятия знание, представление знаний и вывод.

Знаниями принято называть хранимую (в компьютере), формализованную в соответствии с определенными структурными правилами информацию, которую компьютер может автономно использовать при решении проблем по таким алгоритмам, как логические выводы. Знания можно разделить на факты (фактические знания), правила (знания для принятия решений) и метазнания (знания о знаниях).

Факты указывают обычно на хорошо известные в данной предметной области обстоятельства, например, «лев – хищник». Под правилами подразумеваются знания вида «ЕСЛИ ..., ТО ...». Правила позволяют принимать решения, например, сопоставление факта с правилом позволяет принять решение. К метазнаниям относятся знания о способах использования знаний и знания о

свойствах знаний. Метазнания необходимы для управления базой знаний, логическим выводом, обучением. Под выводом подразумевается механизм получения новых знаний на основе имеющихся фактов и правил. Механизм вывода основан на метазнаниях.

**Факты и правила.** Человеческий мозг – это огромное хранилище знаний. Человеку свойственно приобретать новые знания и применять их к возникающим ситуациям, т.е. интеллект можно представить как совокупность фактов и способов их применения для достижения цели. Отчасти цели достигаются с помощью правил использования всех известных фактов.

**Упрощение.** Когда человеческий мозг приступает к решению даже самой простой задачи, для выбора нужных действий в его распоряжении имеется огромный объем информации. Одновременно его мозг обрабатывает впечатления, не имеющие прямого отношения к достигаемой цели. В мозгу существует сложная система, руководящая выбором правильной реакции на конкретную ситуацию. Такой выбор называется упрощением.

**Механизм вывода.** Достигая цели, человек не только приходит к решению поставленной перед ним задачи, но одновременно приобретает новые знания. Часть интеллекта, которая помогает извлекать новые факты, называется «механизмом вывода». Именно механизм вывода позволяет человеку учиться на опыте, так как он дает возможность генерировать новые факты из уже существующих, применяя имеющиеся знания к новой ситуации.

## 1.2. Направления исследований в области искусственного интеллекта

Современные направления исследований в области искусственного интеллекта охватывают широкий круг приложений в различных областях науки, техники, экономики и общественной жизни. Рассмотрим наиболее известные, охватывающие большинство направлений развития. Их можно представить как «распознавание», «осмысление» (системный анализ) и «действие» [4]. В свою очередь эти крупные понятия можно разделить на

отдельные приложения.

1. Компьютерное зрение, IoT-технологии:  
- распознавание жестов, биометрия, распознавание символов.

2. Обработка естественного языка:  
- анализ и синтез речи, диалог на естественном языке, анализ текстов и сообщений, информационный поиск, машинный перевод.

3. Анализ данных:  
- экспертные системы, прогнозные системы, представление знаний, предикативная аналитика.

4. Робототехника:  
- промышленные роботы, беспилотные аппараты, домашние приборы, программные агенты.

Рассмотрим более подробно системы ИИ.

**Разработка систем, основанных на знаниях.** Это одно из главных направлений в искусственном интеллекте. Целью создания таких систем является выявление, исследование и применение знаний специалистов для решения практических задач. Обычно такие знания формализуются в виде некоторой системы правил. В этой области исследований осуществляется разработка моделей извлечения, представления и структуризации знаний в виде «базы знаний». Примеры практических разработок подобных систем обычно ассоциируются с экспертными системами.

**Разработка систем общения на естественном языке и машинного перевода.** Является наиболее важной с точки зрения перехода на новый качественный уровень взаимодействия пользователя с компьютером. Однако для обработки сложного разговорного файла или текста необходимы алгоритмы анализа его смысла, создание которых очень трудоемкая и пока нерешенная полностью задача. Поэтому в настоящее время доступны системы, обеспечивающие диалог между человеком и компьютером на упрощенном естественном языке, программы электронного перевода эффективные преимущественно при работе с односложным текстом, а также функции речевого поиска в электронных словарях и литературных источниках.

**Разработка интеллектуальных систем на основе принципов обучения, самоорганизации и эволюции.** Моделирование этих принципов ориентировано на исследование с помощью законов функционирования биологических систем. Использование принципов эволюции позволяет системе приобретать новые качества и свойства для наиболее оптимального функционирования.

**Распознавание образов.** Является одним из ранних направлений искусственного интеллекта. Оно связано с моделированием особенностей восприятия внешнего мира, узнавания объектов. В основе этого лежит тот факт, что все объекты могут быть проклассифицированы по определенным признакам и, следовательно, умение различать их проявление и позволяет идентифицировать соответствующий объект.

**Программное обеспечение систем искусственного интеллекта.** Инструментальные средства для разработки интеллектуальных систем включают специальные языки программирования, представления знаний, среды создания систем ИИ, а также оболочки экспертных систем.

**Интеллектуальные роботы.** Их создание связано с объединением технологий искусственного интеллекта и методов кибернетики, робототехники. В настоящее время их производство ограничивается манипуляторами с жесткой схемой управления, а также роботами развлекательного и бытового назначения с узкой областью применения и ограниченными функциями. Сдерживающим фактором при разработке более совершенных кибернетических систем являются нерешенные проблемы в области машинного зрения, адаптивного поведения, накопления и обработки трехмерной визуальной информации.

### 1.3. Области прикладных исследований

В приложениях преимущественно создаются такие системы.

1. Системы, имитирующие творческие процессы. Это создание музыкальных произведений, решение игровых задач (шахматы), автоматический перевод, доказательство теорем, распознавание образов, имитация мышления.

2. Информационные системы, основанные на знаниях: поисковые и сервисные системы, системы обучения и

диагностики.

3. Интеллектуальные информационные системы обработки изображений, естественно-языкового диалога, системы библиотечного информационного поиска, военные системы наблюдения и поражения.

4. Первое поколение – роботы-манипуляторы, действующие по заранее утвержденной и неизменной программе (например, подающие заготовки к станку). Второе поколение – адаптивные роботы, оснащенные датчиками для ориентации и точного позиционирования в пространстве. Такие роботы применяются для управления сложными технологическими процессами, сборки автомобилей, поиска взрывчатых веществ, работ на дне морей и океанов.

Решение прикладных задач искусственного интеллекта на компьютерах предполагает создание комплекса аппаратно-программных средств, элементы которого решают задачи хранения, поиска, принятия решения и выдачи результатов (рис. 1.1).

Описание модулей аппаратно-программной системы ИИ.

1. База целей определяет правила поведения, стимулы, действия, критерии.

2. База знаний содержит основные закономерности предметной области и позволяет выводить новые факты, в ней содержатся сведения о структуре и содержании базы данных и средства, обеспечивающие понимание входного языка.

3. База данных содержит структурированные фактографические сведения, которые могут дополняться в процессе работы.

4. Блок представления знаний содержит сведения о среде выполнения вычислений.

5. Блок интерпретации выполняет обработку запросов и формирование результатов в терминах среды решения задачи.

6. Блок выработки решений формирует результат запроса, помогают пользователю сформировать нужную альтернативу среди множества вариантов выбора при принятии ответственных решений. При этом возможен поиск готового решения, полученного на основании опыта или в процессе обучения.

7. Блок усвоения знаний. Имеется в развитых системах ИИ и придает им способность к обучению, накоплению знаний и

коррекции целей.

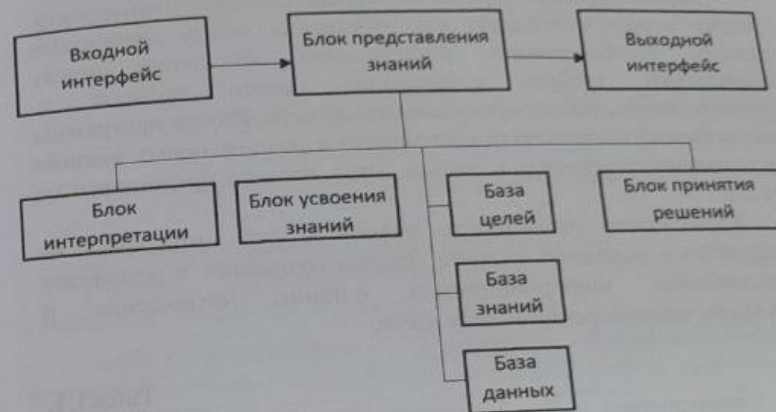


Рис. 1.1. Схема построения системы искусственного интеллекта

Если рассмотреть только компьютерные аппаратно-программные средства, на базе которых реализуются системы ИИ, можно отметить, что такие компьютеры обладают преимуществами, которыми не обладают компьютеры с традиционной архитектурой:

- распределенное представление информации и параллельные вычисления;
- способность к обучению и обобщению;
- адаптивность и самоорганизуемость;
- толерантность к ошибкам;
- низкое энергопотребление и надежность в работе.

Реализация, например, искусственных нейронных сетей, на аппаратном уровне, то есть в виде нейрокомпьютера, обладающего отмеченными выше особенностями, приведет к существенным изменениям большинства характеристик, свойственных традиционным компьютерам (табл. 1.1).

Для того, чтобы компьютерная программа могла решать сложные задачи на уровне человеческого интеллекта она должна обладать знаниями, которые позволяют анализировать, обобщать

начальную информацию и принимать соответствующие решения.  
 Знания – основной термин теории искусственного интеллекта. Знания можно определить как отношения между элементами данных. Необходимость установления отношений между сущностями требует совершенно нового подхода и, соответственно, новых программных средств. Работа программы искусственного интеллекта заключается в «выводе новых знаний» на основании имеющихся данных. Этот процесс представлен на рис. 1.2.

Характерная особенность знаний в том, что они не содержатся в исходной системе. Знания возникают в результате сопоставления информационных единиц, нахождения и разрешения противоречий между ними.

Табл. 1.1.

Характеристика	Традиционный компьютер	Нейрокомпьютер
Процессор	Сложный Высокоскоростной Один или несколько	Простой Низкоскоростной Многопроцессорный
Оперативная память	Адресная Вне процессора Локальная	В процессоре Распределенная Адресация по смыслу
Вычисления	Централизованные Последовательные Алгоритмические	Децентрализованные Параллельные Самообучение
Надежность	Уязвимая	Высокая живучесть
Специализация	Числовые операции	Эвристика
Среда функционирования	Строго ограниченная	Без ограничений
Метод обучения	По правилам	По примерам
Применения	Числовая обработка информации	Распознавание Классификация Управление

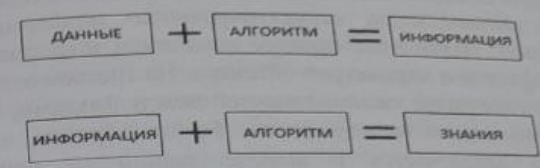


Рис. 1.2. Переход от данных к знаниям

Для знаний характерны следующие свойства:  
 - каждая информационная единица должна иметь уникальное имя и однозначно определяться;  
 - структурированность, между информационными единицами должны быть установлены отношения (например, «часть – целое», «род» – «вид»);  
 - знания образуют некоторое пространство, которое может оказаться как метрическим, так и не метрическим.

Решением проблем создания баз знаний занимается отрасль ИИ, которая называется «инженерия знаний». Создание базы знаний приведено на рис. 1.3.

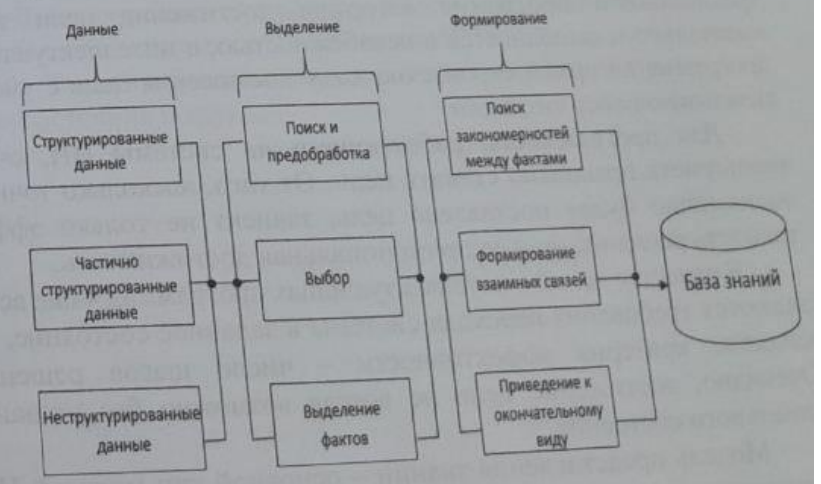


Рис. 1.3. Процесс создания базы знаний



Процесс создания базы знаний (БЗ) начинается с подготовки исходных данных к их структурированию и формальному упорядочению. Далее идет процесс выделения информативных признаков, фактов и параметров объектов. На третьем этапе идет поиск и формирование закономерностей между фактами, а также формирование связей между ними. Таким образом, процесс создания БЗ заключается в выводе знаний на основании взаимосвязанных фактов.

Состояние – второй важный термин, применяемый в системах ИИ. Каждая информационная единица, как и вся система в целом, может находиться в одном из состояний. Переход системы из состояния в состояние может обладать свойством симметричности.

Цель – основная характеристика системы ИИ. Система искусственного интеллекта (интеллектуальная программа) отличается от традиционных программ свойством целенаправленности. Интеллектуальная программа в отличие от обычной не содержит заранее заданный алгоритм, а содержит только цель, которую нужно достичь, и правила движения к этой цели. Можно сказать, что программа ИИ – это не традиционная программа «как нужно делать», а «что нужно сделать?». В традиционной программе алгоритм достижения цели задан изначально и выполняется с неизбежностью, в интеллектуальной программе алгоритм строится по ходу достижения цели с учетом складывающейся ситуации.

Для программиста, работающего на системы ИИ, очень важно уметь правильно ставить цели. От того, насколько точно и экономично будет поставлена цель, зависит не только эффективность решения, но и ее принципиальная достижимость.

В качестве целей в интеллектуальных программах чаще всего задаются требования перехода системы в заданное состояние, а в качестве критерия эффективности – число шагов решения. Очевидно, достижение цели не всегда возможно без указания начального состояния.

Модель представления знаний – основной тип моделей ИИ. Реализация конкретных систем, основанных на знаниях, происходит в рамках языка представления знаний (будут

рассмотрены далее).

Предметная область – это область человеческой деятельности, к решению задач которой применяется теория ИИ. Например, если создается экспертная система, которая по набору результатов анализов ставит диагноз больному, то эта предметная область – медицина. Главное отличие экспертной системы в медицине – это наличие сформированной базы знаний на специальном языке представления знаний.

#### Вопросы для контроля

1. Какая система называется «интеллектуальной»?
2. Что понимается под «механизмом вывода»?
3. Перечислите области исследований по проблемам ИИ
4. Перечислите основные отличия традиционного компьютера от нейрокомпьютера.
5. Дайте определение термину «база знаний».
6. Опишите модули «База целей», «База знаний», «Баз данных».
7. Определите функции блоков «Представление знаний», «Интерпретация» и «Выработка решений».
8. Дайте определение термину «Состояние системы».
9. В чем заключается цель системы ИИ?
10. Какие требования задаются для перехода системы из одного состояния в другое?

## ГЛАВА 2. МЕТОДЫ ЭВРИСТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Эвристическое программирование – разработка стратегий действий на основе заранее заданных эвристик (теоретически необоснованных правил). Эвристический метод – метод решения задачи, основанный на интуиции решающего лица. Этот метод часто не имеет формального доказательства того, что он приводит к цели.

Сам процесс эвристического программирования можно рассматривать как построение некоторого плана действий с применением эвристик. Все задачи, связанные с построением плана действий, можно разбить на два типа:

- планирование в пространстве состояний;
- планирование в пространстве задач.

Для первого типа задач считается заданным некоторое фиксированное пространство состояний, требуется найти путь, ведущий из начального состояний в одно из конечных.

Второй тип задач состоит в поиске декомпозиции исходной задачи на подзадачи посредством введения между ними отношений «часть-целое», «задача-подзадача», «общее-частное» и т.п., что в результате приводит к задачам, решение которых известно.

### 2.1. Поиск в пространстве состояний

Типичным представителем класса задач, для которых подходит представление (формализация) в пространстве состояний, является головоломка, известная как игра в пятнадцать. В ней используется пятнадцать пронумерованных (начиная с 1) подвижных фишек, расположенных в клетках квадрата 4x4. Одна клетка этого квадрата остается всегда пустой. Передвигая соседние с ней фишки необходимо добиться упорядоченного расположения их номеров. Есть также упрощенный вариант - квадрата 3x3 и восьми фишек 8x8. Решением задачи перевода начальной конфигурации в целевую конфигурацию будет подходящая последовательность сдвигов фишек на место пустующей клетки.

Основными особенностями класса задач, к которому принадлежит рассмотренная головоломка, является наличие в каждой задаче точно определенной начальной ситуации и точно определенной цели. Имеется также некоторое множество операций или ходов, переводящих одну ситуацию в другую. Именно из таких ходов состоит искомое решение задачи, которое можно получить (теоретически) методом проб и ошибок. Действительно, отправляясь от начальной ситуации, можно построить все промежуточные конфигурации, возникающие в результате выполнения каждого из возможных ходов, затем построить множество конфигураций после применения следующего хода и так далее – пока не будет достигнута целевая конфигурация.

Для рассматриваемой игры удобнее выделить четыре оператора, соответствующие перемещениям пустой клетки влево, вправо, вверх, вниз. В терминах состояний и операторов решение задачи есть определенная последовательность операторов, преобразующая начальное состояние в целевое. Решение задачи ищется в **пространстве состояний** – множестве состояний, достижимых из начального состояния при помощи операторов. В игре в пятнадцать пространство состояний состоит из всех конфигураций фишек, которые могут быть образованы в результате допустимых перемещений фишек.

Рассмотрим еще один пример представления задачи в пространстве состояний, а именно задачу о коммивояжере (продавце товаров и услуг). В ней коммивояжер, располагая картой дорог между несколькими городами, должен выстроить кратчайший маршрут своей поездки так, чтобы побывать в каждом городе, но не более одного раза. На рис. 2.1 показана карта дорог между семью городами. Для этой задачи фрагмент пространства состояний, представленного деревом, приведен на рис. 2.2.

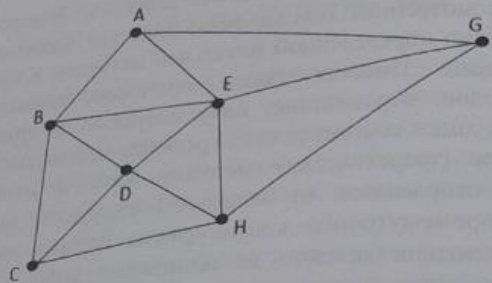


Рис. 2.1. Пример задачи о коммивояжере

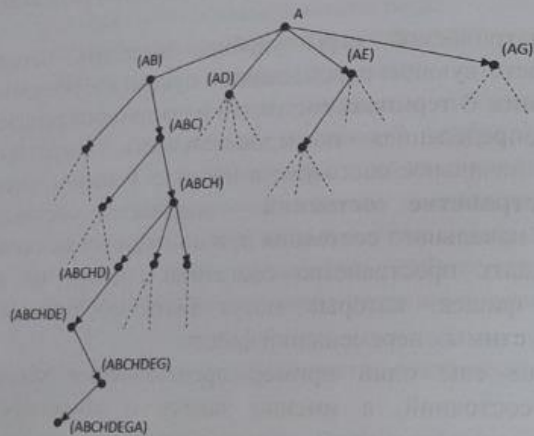


Рис. 2.2. Фрагмент пространства состояний для задачи о коммивояжере

Алгоритм поиска решения задачи заключается в построении дерева пространства состояний посредством перебора его вершин до тех пор, пока не будет обнаружена целевая вершина. При этом эффективность решения зависит от такого параметра, как максимальная глубина дерева. Вершины и указатели, построенные в процессе перебора, образуют поддерево всего пространства состояний (поддерево перебора).

Известные алгоритмы поиска в пространстве

состояний различаются несколькими характеристиками:

- использованием или нет эвристической информации;
- порядком раскрытия (обхода) вершин;
- полнотой просмотра пространства;
- направлением поиска.

На приведенном рисунке 2.2 показано последовательное перемещение из точки А с анализом возможных промежуточных точек движения АВ, АЕ, ... с выходом на конечную точку АВСНДЕГА.

В соответствии с первой характеристикой алгоритмы делятся на два класса – слепые и эвристические. В слепых алгоритмах поиска в пространстве состояний место расположения целевой вершины никак не влияет на порядок, в котором рассматриваются вершины. Такая задача чаще всего решается путем простого перебора.

В представленном эвристическом алгоритме движение идет с учетом наличия ложных, неэффективных путей. Поэтому для уменьшения перебора используют априорную (эвристическую) информацию о том, где в пространстве состояний расположена цель, и для раскрытия обычно выбирается более перспективная вершина [5,6,7].

## 2.2. Эвристический поиск

Идея большинства эвристических алгоритмов в том, чтобы оценивать перспективность нераскрытых вершин пространства состояний и выбирать для продолжения поиска наиболее перспективную вершину. Для этого используется эвристическая оценочная функция. Эта функция определяется на множестве вершин пространства состояний и принимает числовые значения, которые могут интерпретироваться как перспектива раскрытия вершины или вероятность ее расположения на решающем пути. Использование такой функции позволяет сделать поиск упорядоченным [7,8].

Рассмотрим основные шаги алгоритма эвристического поиска.

Пусть имеются: одна начальная вершина  $x_0 \in X_0$  ;

$S$  – множество уже выбранных вершин;  $\bar{S}$  – множество вершин-кандидатов в  $S$ , при  $S \cap \bar{S} = \emptyset$ ;  
 $X_t$  – множество конечных вершин;  
 $x$  – текущая вершина;  
 $x_i$  – дочерняя вершина  $x$ .

Тогда алгоритм поиска в графе пространства состояний заключается в поиске пути, начальная из вершины  $x_0$ , просматривая граф в ширину и представляется следующими шагами.

1. Поместить  $x_0$  в  $\bar{S}$  и вычислить оценочную функцию  $\varphi(x_0)$ .
2. Выбрать такую  $x_0 \in \bar{S}$ , что  $\varphi(x) = \min(\varphi_{y \in \bar{S}}(y))$  и поместить ее в  $S$ , изъяс из  $\bar{S}$ . При равенстве выбрать любую.
3. Если  $x \in X_t$  – путь найден, иначе продолжить.
4. Найти все  $x_i \in F(x)$ , если  $F(x) = \emptyset$ , то перейти к шагу 2, иначе вычислить все  $\varphi(x_i)$ .
5. Для каждого  $x_i$ :
  - а) если  $x_i \notin S \cup \bar{S}$ , то поместить  $x_i$  в  $\bar{S}$ ;
  - б) если  $x_i \in F(x) \cap \bar{S}$ , то сопоставить  $x_i$  наименьшее из старой и вновь полученной оценки  $\varphi(x_i)$ ;
  - в) если  $x_i \in F(x) \cap S$ , то сопоставить  $x_i$  наименьшее из старой и вновь полученной оценки  $\varphi(x_i)$ , поместить  $x_i$  в  $\bar{S}$  (изъяс ее из  $S$ );
  - г) в остальных случаях не изменять  $S$  и  $\bar{S}$ .
6. Перейти к шагу 2.

Пункты 5(б), 5(в) отражают действие алгоритма, когда оператор  $F$  порождает уже рассмотренные вершины, которые к этому моменту находятся в  $S$  или  $\bar{S}$ , поэтому этим вершинам придаются наименьшие из возможных оценочных функций.

На рис. 2.3 показано дерево игры в восемь (уменьшенный вариант игры в пятнадцать), построенного алгоритмом эвристического перебора с указанной оценочной функцией.

Оценка каждой вершины приведена рядом с ней внутри кружка. Отдельно стоящие цифры показывают порядок, в котором раскрывались вершины.

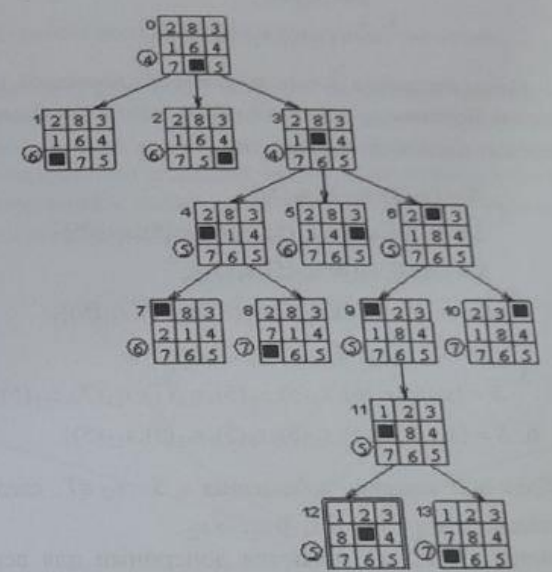


Рис. 2.3. Пример дерева эвристического поиска

Рассмотрим процесс построения дерева поиска пути на основе изложенного ранее алгоритма. Результаты поиска пути представляется в виде двух множеств  $S$  и  $\bar{S}$  содержащих вершины с указанием в скобках значений весов входящих в них дуг.

В качестве оценочной функции используется:  $\varphi(x) = D(x) + K(x)$ , где:

$D(x)$  – глубина вершины  $x$  или число ребер дерева на пути от этой вершины к начальной;

$K(x)$  – число фишек позиции – вершины  $x$ , лежащих не на «своем» месте.

Будем считать, что меньшее значение  $\varphi(x)$  соответствует более перспективной вершине, и вершины раскрываются в

порядке увеличения (возрастания) значения оценочной функции.

$$S = \{x_0(4)\};$$

$$1. \bar{S} = \{x_1(6), x_2(6), x_3(4)\}.$$

Новые вершины в  $\bar{S}$  есть дочерние для последней добавленной вершины. Вершины  $x_4, x_6$  имеют одинаковую минимальную оценку, выберем из них любую и добавим в  $S$ .

$$S = \{x_0(4), x_3(4), x_6(5)\};$$

$$3. \bar{S} = \{x_1(6), x_2(6), x_4(5), x_5(6), x_9(5), x_{10}(7)\}.$$

$$S = \{x_0(4), x_3(4), x_6(5), x_9(5)\};$$

$$4. \bar{S} = \{x_1(6), x_2(6), x_4(5), x_5(6), x_{10}(7), x_{11}(5)\}.$$

$$S = \{x_0(4), x_3(4), x_6(5), x_9(5), x_{11}(5)\};$$

$$5. \bar{S} = \{x_1(6), x_2(6), x_4(5), x_5(6), x_{10}(7), x_{13}(7), x_{12}(5)\}.$$

$$6. S = \{x_0(4), x_3(4), x_6(5), x_9(5), x_{11}(5), x_{12}(5)\}.$$

Последняя вершина, добавленная в  $S$   $x_{12} \in T$ , следовательно, путь найден:  $x_0 \rightarrow x_3 \rightarrow x_6 \rightarrow x_9 \rightarrow x_{11} \rightarrow x_{12}$ .

Вершины  $x_1, x_2, x_3$  являются дочерними для вершины  $x_0$ . Вершина  $x_3$  имеет наименьшее значение оценочной функции, поэтому она исключается из  $\bar{S}$  и добавляется в  $S$ .

В ходе решения задачи может возникнуть ситуация, когда при выборе вершин для множества  $S$  совпадают их оценочные функции, в этом случае необходимо проанализировать всевозможные альтернативные пути, начиная с данного этапа.

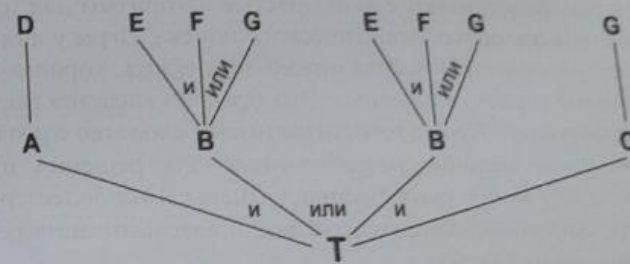
Найденный путь решения задачи длиной в пять ходов может быть получен и другими методами перебора, но использование оценочной функции приводит к существенно меньшему числу раскрытых вершин. Действительно, сравнение трех алгоритмов перебора показывает, что в среднем алгоритм эвристического поиска обнаруживает решение быстрее алгоритмов слепого перебора.

Выбор «хорошей» эвристической функции, существенно

сокращающей поиск, наиболее трудный момент при формализации задачи, особенно это важно в больших пространствах состояний.

### 2.3. Эвристический поиск в пространстве задач

Такой поиск в пространстве задач сводится к нахождению доказательства того, что решение данной задачи выводится из решения совокупности ее подзадач. Решение будет получено тогда, когда множество подзадач будут полностью состоять из заведомо разрешимых задач (аксиом) или будет доказано, что исходная задача не имеет решения.



На рис. 2.4 представлен пример задачи T в виде И/ИЛИ дерева.

Задача T может быть решена, если решены подзадачи A, B или C, D, а задача A решается, если достижима вершина D. Задача B разрешима – если E, F или G, а задача C, если решается подзадача G. Подзадачи D, E, F, G называются разрешимыми.

Существует несколько различных методов эвристического поиска в пространстве задач, например, универсальный решатель задач GPS [8,9].

В основе работы алгоритма GPS лежит метод «анализа целей и средств», согласно которому для достижения каждой подцели выбирается оператор, уменьшающий различие между имеющимся и желаемым состоянием объекта. При этом изначально должны быть заданы операторы, различия, таблица связей, а также цель (суперцель) и объекты, участвующие в задаче.

В GPS используется четыре вида целей:

- преобразование A в объект B;
- уменьшение различия D между объектами A и B;
- применение оператора Q к объекту A;
- выбор из множества S элемента, наилучшим образом удовлетворяющего критерию C.

Работа метода GPS заключается в циклическом поиске способа уменьшить различие между текущим состоянием объекта и заданной целью, посредством рекурсивной генерации из существующих новых подцелей.

#### 2.4. Игровая модель эвристического поиска

Игры представляют собой простые алгоритмы для изучения и испытания процедур эвристического поиска. Игры удобны тем, что известны правила и всегда можно определить, хорошо человек или машины играет, или плохо. Это правило касается всех игр и игровых ситуаций. Кроме того, игры имеют сходство с реальными проблемами, и методы, разработанные для решения простых игровых задач, могут быть распространены и на более трудные, например, поисковое конструирование и автоматизация решения изобретательских задач.

В теории игр для каждой игры характерны следующие особенности.

1. Должны быть как минимум два игрока.
2. Игроки поочередно делают ходы, пытаясь максимизировать свой выигрыш.
3. Варианты ходов, делаемых игроками, могут быть как известны, так и не известны другим игрокам.
4. Существуют критерии окончания игры и определения победителей.
5. Существует мера выигрыша – некая выплата в условных очках, баллах или деньгах.

Игры бывают с полной информацией (шахматы, шашки, крестики – нолики) и с неполной информацией (карточные игры). Критерий полноты информации – знание всех ходов противника, которые произошли ранее или могут быть сделаны в настоящий

момент (в карточных играх игрок не видит карты противника, в то время как в шахматах игрок видит всю доску и все возможные ходы).

#### Структура игровой модели

Игровая модель представляет собой комбинацию следующих элементов:

- дерево игры: граф  $G = (X, E)$ ;
- порождающие процедуры  $F(x)$ ;
- критерий окончания игры ( $x \in X_t$ );
- оценочная функция  $\varphi(x)$ ;
- процедуры оценки и выбора хода.

#### Дерево игры

Процесс игры можно представить в виде дерева, узлы которого служат состояниями, а ребра ходами. Цель построения дерева игры заключается в определении выигрышной стратегии для одного из игроков (в нашем случае это игрок  $X_{max}$ ), отправляясь от некоторой фиксированной конфигурации (позиции) игры (не обязательно начальной) независимо от ответов противника. Решающее дерево заканчивается на позициях, выигрышных для конкретного игрока, и содержит стратегию достижения им выигрыша: для каждого возможного продолжения игры, выбранного противником, в дереве есть ответный ход, приводящий к победе. Схема дерева соответствует состояниям игры, в которых она заканчивается ничьей, проигрышем или выигрышем одного из игроков.

#### Порождающие процедуры

Задача порождающих процедур состоит в формировании дерева игры, то есть генерация приемников для каждой позиции. В общем случае процесс формирования дерева – это упорядоченный перебор возможных конфигураций игры с ограничением глубины перебора (иначе алгоритм никогда не закончит работу). В большинстве игр (шахматы, шашки)

невозможно построить полное дерево. Так, в шашках общее число вершин оценивается как  $10^{40}$ .

### Критерии окончания игры

Имеются указания для прекращения перебора возможных конфигураций игры. Среди них можно выделить следующие:

- задание значения максимальной глубины;
- мертвая позиция, например в шахматах это позиция, в которой невозможны немедленный ход, взятие фигуры, шах.

### Оценочная функция

Назначение оценочной функции заключается в определении достоинства игровой позиции. Вычисление оценочной функции позволяет дать эвристическую оценку шансов на выигрыш одного из игроков. Оценочная функция, как правило, является статической и чаще всего принимает линейную форму, имеющую вид:

$$\varphi(x) = c_1 y_1 + c_2 y_2 + \dots + c_n y_n,$$

где  $y_i$  – признак позиции, то есть параметр, характеризующий развитие игры;

$c_i$  – вес соответствующего признака  $y_i$ . Например, в шашках

$$\varphi(x) = 6k + 4m + u,$$

где  $k$  – перевес в дамах;

$m$  – перевес в шашках;

$u$  – перевес в подвижности.

Значения 6, 4, 1 соответствуют весам признаков.

Выбор оценочной функции и глубина перебора являются одними из главных факторов, определяющих алгоритм игры. Выбор или создание хорошей оценочной функции является необходимым, но не достаточным условием создания эффективной игровой модели. Большинство алгоритмов игр двух игроков с полной информацией и нулевой суммой выигрышей базируется на общих идеях минимаксного поиска, а также совершенствующих его различных модификаций.

### Вопросы для контроля

1. Дайте определение эвристическому программированию.
2. Определите понятие «пространство состояний».
3. В чем особенности поиска в пространстве состояний?
4. Каким образом формируется пространство состояний в задаче окоммивояжере?
5. Постройте дерево поиска пути для игры в «восемь».
6. В чем суть игровой модели эвристического поиска?
7. Какова задача порождающей функции?
8. Какие основные правила обязательны для теории игр?
9. Перечислите критерии окончания игры.
10. Как вычисляется оценочная функция?

## ГЛАВА 3. МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

### 3.1. Основные понятия и определения

Представление знаний является одним из важнейших разделов искусственного интеллекта. Искусственный интеллект как научное направление связан с попыткой формализовать мышление человека – разработать методы, которые позволили бы запрограммировать машину таким образом, чтобы она могла воспроизводить или даже превосходить способности человеческого интеллекта. Исследования в этой области тесно связаны со смежными дисциплинами – информатикой, лингвистикой, психологией и философией.

Реализация систем обработки, основанных на знаниях, происходит в рамках моделей представления знаний или языка представления знаний. Теоретическими и практическими вопросами представления и обработки знаний в компьютерных системах занимается отдельное направление ИИ – инженерия знаний. Программы, использующие достижения в этой области, образуют отдельный класс компьютерных систем, основанный на знаниях. Главными структурными элементами таких систем являются базы знаний и механизм логических выводов.

Для организации знаний информация представляется в виде фактов и правил, которые компьютер может использовать при решении задач по интеллектуальным алгоритмам. Знания можно разделить на декларативные и процедурные. Декларативные знания представляют собой описания фактов и явлений, а также связанных с ними закономерностей. Процедурные знания – это описание действий, которые возможны при манипулировании фактами и явлениями для достижения намеченных целей.

Следовательно, при обработке знаний наиболее фундаментальной проблемой является описание смыслового содержания решаемой проблемы, нахождение такой формы описания знаний, которая гарантирует правильность обработки формальных правил преобразования. Эта проблема называется проблемой представления знаний.

### Этапы формирования знаний в предметной области

1. Этап. Инженерия знаний – с помощью специалиста предметной области формализуются правила решения задачи исходя из опыта и интуиции.

2. Этап. Выделение отношений, связывающих имена и их состояния, в которых они могут находиться. Это построение математических или логико-эвристических моделей поведения системы.

Появление баз данных (БД) позволило организовать работы с декларативной информацией. В базах данных могут одновременно храниться большие объемы информации, а средства, образующие систему управления базами данных (СУБД), позволяют эффективно манипулировать с данными, при необходимости извлекать их из базы данных и записывать их в нужном порядке в базу.

При подготовке к работе данные трансформируются, условно проходя следующие этапы.

1. Данные как результат измерений и наблюдений.
2. Данные на материальных носителях (таблицы, протоколы)
3. Структуры данных в виде диаграмм, графиков, функций.
4. Данные в компьютере на языке описания данных.
5. Базы данных на машинных носителях информации.

По мере развития исследований в области ИИ возникла концепция знаний, которые объединили в себе многие черты процедурной и декларативной информации. В компьютерах знания так же, как и данные, отображаются в знаковой форме – в виде формул, текста, файлов, информационных массивов, знания – это особым образом организованные данные. База знаний, наравне с базой данных, – необходимая составляющая программного комплекса ИИ. База знаний может совершенствоваться и дополняться.

При обработке на компьютере знания трансформируются аналогично данным.

1. Знания в памяти человека как результат мышления.
2. Материальные носители знаний (книги, картины).
3. Поле знаний – условное описание основных объектов предметной области, их атрибутов и закономерностей, их



связывающих.

4. Знания, описанные на языках представления знаний (семантические сети, фреймы).

5. Базы знаний на машинных носителях.

Таким образом знания – это хорошо структурированные металанные (данные о данных).

Инженерия знаний – наука о компьютерном представлении знаний и их обработке. Она рассматривает средства, позволяющие описывать знания с помощью языка представления знаний, организовать хранение знаний в системе ИИ.

### Стратегии получения знаний

1. Извлечение знаний без использования вычислительной техники путем непосредственного контакта инженера по знаниям и источника знания (будь то эксперт, специальная литература или другие источники).

2. Приобретение знаний от эксперта с использованием компьютера при наличии подходящего программного инструментария.

3. Формирование знаний с использованием программ обучения при наличии репрезентативной (т.е. достаточно представительной) выборки примеров принятия решений в предметной области и соответствующих пакетов прикладных программ.

Приобретение знаний подразумевает, что автоматизированные системы действительно непосредственно приобретают уже готовые фрагменты знаний в соответствии со структурами, заложенными разработчиками систем. Большинство этих инструментальных средств специально ориентировано на конкретные базы знаний с жестко обозначенной предметной областью и моделью представления знаний, т.е. не являются универсальными.

На рис.3.1 даны современные типы моделей представления знаний. Представленные виды моделей представления знаний делятся на две условные категории: классические, которые давно и успешно применяются при исследованиях систем ИИ, и

современные, которые являются результатом развития численных методов, архитектур компьютеров и их программного обеспечения.

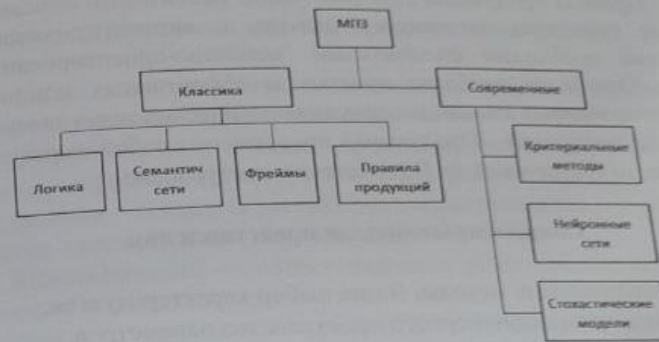


Рис.3.1. Модели представления знаний.

### Классические модели представления знаний

Логическая модель по своей практической результативности истепени внедрения в реальные технические устройства сегодня занимает центральное место. Смысл построения любой формальной теории состоит в том, чтобы выразить мыслительные процессы формально и записать формулами.

Семантические сети. Семантика – наука, устанавливающая отношения между символами и объектами, которые они обозначают, т.е. наука, определяющая смысл знаков. Семантическая сеть – это ориентированный граф, вершины которого – понятия, а дуги – отношения между ними. Узлы в семантической сети обычно соответствуют объектам, концепциям, событиям или понятиям. Дуги могут быть определены разными методами, зависящими от метода представления знаний.

Фреймы – это структура данных для представления типичной, привычной для мозга человека ситуации. Человек смотрит на предмет и подбирает наиболее близкий из известных ему фреймов (аналог). Процессы распознавания, мышления и воображения

можно представить как типовые операции над фреймами: выбор некоторого фрейма из сети подобных, его заполнение и переход к новому фрейму.

Правила продукций. Для того чтобы реализовать описанные выше принципы логического вывода, в автоматизированной системе необходим специальный машинно-ориентированный язык. Одним из наиболее простых и эффективных машинно-ориентированных языков для описания логических задач является правила продукций. «Продукция» представляет собой выражение вида: если А (условие), то В (действие), постусловие.

### Современные модели представления

Критериальные методы. В них выбор характеризуется:

- наличием многомерного пространства параметров;
- описанием желательных областей в этом пространстве;
- критерием выбора (насколько избранный альтернатива близка желательной);
- набором пороговых (предельных) значений, т. е. различий в мерах между двумя альтернативами;
- математическими методами, применяемыми для определения наиболее выгодного направления движения в пространстве состояний.

Нейронные сети. Одним из активно развиваемых сегодня направлений в ИИ являются нейронные сети – модель, представляющая собой в некотором смысле гибриды критериальных, вероятностных и логических методов. Можно также провести аналогию с правилами продукций. Нейронная сеть – это в некотором смысле числовая запись правил «если-то».

Стохастические модели. Широкий класс подходов использует в качестве меры близости альтернатив вероятностные оценки. В настоящее время методы, позволяющие оценить вероятные исходного или иного решения, его правильность, исследуют в рамках так называемой теории рисков. В математическом плане это обыкновенная теория вероятностей, причем в ее самых простых проявлениях. Суть теории рисков составляют не математические, а алгоритмические процедуры,

связанные с особенностями измерения и вероятностной обработки данных в конкретной предметной области, а также интерпретации полученных результатов.

### Основные понятия моделей представления знаний

Имена – выражение языка, означающее отдельные предметы или их совокупность со сходными свойствами.

Объем имени – класс предметов, которые обладают сходными признаками, входящими в содержание имени.

Деление – операция распределения на группы тех предметов, которые входят в объем исходных данных.

Классификация – многоступенчатое разветвленное деление, где делимое является «родом», новые имена – «видами».

Класс – множество предметов с общими признаками, выражаемыми именем данного класса.

Классификационные системы применяются для структурирования и обобщения знаний. Все сущности разбиваются по определенным признакам и группируются вместе. Формируется набор объектов, которые можно описать некоторым множеством признаков. При этом каждый объект принадлежит одному или более классам из некоторого фиксированного множества.

При помощи классификации решают три основные задачи:

- классификация образов – отнесение вновь поступившего объекта к некоторому классу;
- распознавание образов – правило классификации вырабатывается на основе исследования множества объектов с принадлежностью к различным классам (обучающей выборке);
- формирование образов – объекты представляются исследователю без указания их принадлежности к классам, он сам должен сам построить деление на классы.

На рис. 3.2. приведен пример иерархической классификации.

Примеры деления на классы – звуки речи: гласные, согласные, шипящие, телефонный справочник (алфавитная классификация), библиотека – тематическая классификация.

Тезаурус – центральная тема запроса (перечень простых слов в виде класса, координаты документа в базе данных). Пример тезауруса – ключевые слова.

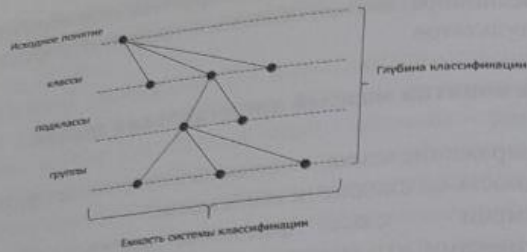


Рис.3.2. Иерархическая классификация

Дескриптор – однозначное ключевое слово, обозначающее класс (не во всех случаях может быть создан).  
 Высказывание (утверждение) – правильное предложение с истинным или ложным смыслом.

### 3.2. Логическая модель представления знаний

Логика – это наука о рассуждениях, которая позволяет определить истинность или ложность того или иного математического утверждения. Основная идея – формальное выражение мыслительных процессов. Для задания формальной логической теории необходимо определить:

- алфавит (символы записи);
- правила синтаксиса (правила записи формул);
- правила вывода (множество отношений для множества формул).

В основе логической модели лежит система исчисления предикатов первого порядка, которая, в свою очередь, основана на исчислении высказываний. Под высказыванием понимают утверждение, о котором в данной ситуации можно сказать, истинно оно или ложно. Эти высказывания являются элементарными, потому что их нельзя разделить на части.

Из простых высказываний могут быть построены сложные с помощью известных логических операций: отрицания ( $\neg$ ), конъюнкции ( $\wedge$ ), дизъюнкции ( $\vee$ ), импликации ( $\rightarrow$ ), эквиваленции

( $\equiv$ ), исключающего ИЛИ ( $\oplus$ ). При этом, любое сложное высказывание может быть упрощено до выражения, содержащего только основные логические операции ( $\neg$ ,  $\wedge$ ,  $\vee$ ).

Примером реализации формальной логики может служить Булева алгебра. Элементами булевой алгебры являются операции:

- конъюнкция (И)
- дизъюнкция (ИЛИ)
- отрицание (НЕ)
- эквивалентность (А тогда, когда В)
- импликация (если А то В).

Для упрощения используются следующие эквивалентные преобразования.

$$A \rightarrow B = \bar{A} \vee B;$$

$$A \equiv B = (A \rightarrow B) \wedge (B \rightarrow A) = (A \wedge B) \vee (\bar{A} \wedge \bar{B});$$

$$A \oplus B = (\bar{A} \wedge B) \vee (A \wedge \bar{B});$$

$$\overline{A \vee B} = \bar{A} \wedge \bar{B};$$

$$\overline{A \wedge B} = \bar{A} \vee \bar{B};$$

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C);$$

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C).$$

В процессе упрощения любая логическая формула может быть приведена к эквивалентной ей конъюнктивной нормальной форме, которая представляет собой конечное число дизъюнктов, объединенных операцией конъюнкции. Каждая логическая операция имеет множество правил интерпретации (табл. 3.1), на основании которых могут быть решены задачи описания знаний из различных прикладных областей.

Табл.3.1.

Операция	Интерпретация
$A$	«Не», «Неверно»
$A \wedge B$	«И», «Одновременно»
$A \vee B$	«Или», «Хотя бы один»

$A \rightarrow B$	«Если А то В», «В необходимо для А», «А только, если В», «А достаточно для В», «В тогда, когда А», «А только тогда, когда В»
$A \equiv B$	«Эквивалентно», «Равносильно»
$A \oplus B$	«Либо А, либо В», «Или А, или В»

Одной из ключевых сторон представления знаний с помощью логических моделей является возможность моделирования рассуждений. В этом случае применяется так называемый дедуктивный вывод от общего к частному. С его помощью могут быть решены такие задачи, как поиск следствия и поиск доказательства. Вывод представляет собой процедуру, которая из заданной группы выражений логических утверждений выводит новое логическое выражение или следствие. Для этого могут быть использованы специальные правила вывода, состоящие из посылок гипотез и утверждения, называемого заключением. Правильным заключением называется такое, которое истинно всякий раз, когда истинны гипотезы.

Рассмотрим наиболее известные из существующих правил вывода.

1. «Если истинна импликация  $A \rightarrow B$  и  $A$  истинно, то  $B$  истинно».
2. «Если истинна импликация  $A \rightarrow B$  и  $B$  ложно, то  $A$  ложно».
3. «Если  $A$  истинно и конъюнкция  $A \wedge B$  ложна, то  $B$  ложно».
4. «Если  $A$  ложно и дизъюнкция  $A \vee B$  ложна, то  $B$  ложно».
5. Цепное заключение: «Если истинна импликация  $A \rightarrow B$  и истинна импликация  $B \rightarrow C$ , то импликация  $A \rightarrow C$  является истинной».

Рассмотрим пример вывода с применением этих правил, при этом заданы следующие посылки.

1.  $P \rightarrow Q$ : если растут мировые цены на топливно-энергетические ресурсы, то увеличиваются поступления в бюджет.

2.  $(R \vee Q) \rightarrow (R \vee S)$ : если наблюдается рост производства или увеличиваются поступления в бюджет, то следует увеличение

производства или укрепление валюты.

3.  $(P \rightarrow Q) \rightarrow ((P \vee Q) \rightarrow (R \vee Q))$ : если растут мировые цены на топливно-энергетические ресурсы, то увеличиваются поступления в бюджет, из чего следует, что при росте цен на сырье или при увеличении поступлений в бюджет происходит рост производства или увеличение бюджета.

Используя правило 1, из посылок 1 и 3 можно вывести следующее заключение.

4.  $(P \vee Q) \rightarrow (R \vee Q)$ : если растут мировые цены на топливно-энергетические ресурсы или увеличиваются поступления в бюджет, то происходит рост производства или увеличение бюджета. Из посылок 4 и 2, используя правило 5, можно получить посылку 5.

5.  $(P \vee Q) \rightarrow (R \vee S)$ : если растут мировые цены на топливно-энергетические ресурсы или увеличиваются поступления в бюджет, то происходит рост производства или укрепление валюты. Таким образом, применяя правила логического вывода, можно получить новые логические формулы на основании исходных без использования таблиц истинности.

Все эти высказывания и заключения составляют основу алгоритмов ИИ, решающих задачи с помощью логических моделей представления знаний.

### 3.3. Продукционная модель представления знаний

Для реализации принципов логического вывода необходим специальный алгоритмический язык [11]. Одним из таких языков являются правила продукций. В продукционной модели описания знаний, знания описываются в форме «ЕСЛИ – ТО». Условная часть правила «ЕСЛИ» называется посылкой или антецедентом, а часть «ТО» – заключением (продукцией). Подобное представление знаний является наиболее простым и легко понимаемым, потому что близко к привычной нам форме рассуждений:

- Если  $A$  (условие), то тогда  $B$  (действие), постусловие  $C$ .
- Условие – это образец, по которому осуществляется поиск в базе знаний.



Проверяется – нет ли среди полученных вершин целевой вершины. Если есть – формируется решение на основе выбранного оператора. Если цель не достигнута – процесс продолжается. Такой перебор гарантирует нахождение целевой вершины.

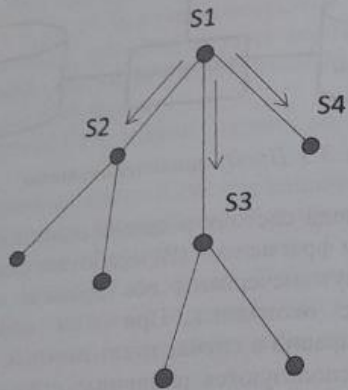


Рис. 3.4. Метод перебора в ширину

Методов полного перебора много, но при сложных построениях графа направлений перебора, такой подход не является оптимальным. Более оптимальным являются эвристические методы. Эвристические методы поиска в пространстве состояний применяют тогда, когда граф поиска очень сложный, простыми программами перебора его не выполнить. Примером является известная задача коммивояжера, описанная ранее (гл.2).

### Механизмы вывода продукционных моделей

Именно продукционная модель представления знаний получила наибольшее распространение. При использовании продукционной модели база знаний состоит из набора правил. Программа, управляющая перебором правил называется механизмом вывода.

Механизм вывода (интерпретатор правил) выполняет две

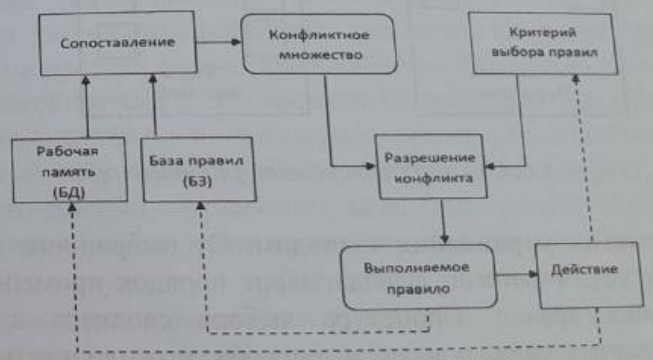
функции:

- просмотр существующих фактов из рабочей памяти (базы данных) и правил из базы знаний;
- определение порядка просмотра и применение правил.

Механизм вывода представляет собой небольшую программу и включает два компонента: один реализует собственно вывод, другой управляет этим процессом. Управляющий компонент определяет порядок применения правил и выполняет четыре функции (рис.3.5).

1. Сопоставление – образец правила сопоставляется с полученными фактами из базы данных.
2. Выбор – если в конкретной ситуации может быть применено сразу несколько правил, то из них выбирается наиболее подходящее по заданному критерию (разрешение конфликта).
3. Срабатывание – если образец правила при сопоставлении совпал с какими-либо фактами из рабочей памяти, правило срабатывает.
4. Действие – в рабочую память записывается заключение (вывод) сработавшего правила. Если в правиле содержится указание на какое-либо действие, то это действие выполняется.

Рис.3.5. Цикл работы интерпретатора



**Интерпретатор** продукционных правил работает следующим образом. В каждом цикле он просматривает все правила, чтобы выявить их совпадение с известными фактами из рабочей памяти (базы данных). Совокупность отобранных правил составляет конфликтное множество. Для разрешения конфликтного множества интерпретатор имеет критерий, с помощью которого он выбирает единственное правило. После выбора правило срабатывает, его заключение заносится в рабочую память, затем цикл повторяется снова.

Работа механизма вывода зависит только от состояния рабочей памяти и от полноты базы знаний. Информация о поведении механизма вывода запоминается в памяти состояний (рис. 3.5). Обычно память состояний содержит протоколы работы системы.

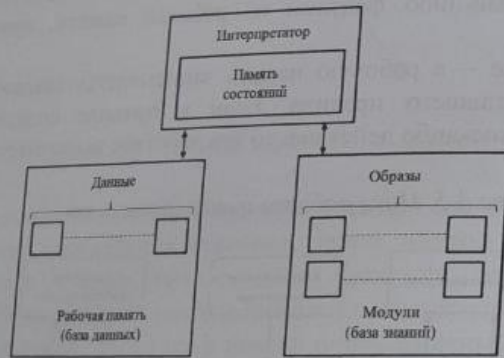


Рис. 3.6. Схема работы интерпретатора

**Стратегия управления выводом.** От выбранного метода поиска, то есть стратегии вывода, зависит порядок применения и срабатывания правил. Процедура выбора сводится к определению направления поиска и способа его осуществления. Процедуры, реализующие поиск обычно «зашиты» в механизм вывода, поэтому инженеры знаний не имеют к ним доступа.

### 3.4. Фреймовая модель представления знаний

Во фреймовой модели за единицу представления принят объект, называемый фреймом. С точки зрения памяти фрейм – единица представления знаний, запомненных в прошлом, детали которой при необходимости могут быть изменены согласно текущей ситуации [13,14].

Во фреймовой модели представления знаний за единицу представления принят объект, называемый фреймом. С точки зрения памяти фрейм – единица представления знаний, запомненных в прошлом, которые при необходимости могут быть изменены.

Алгоритмически фрейм – это структура для описания понятия или ситуации, состоящая из характеристик этой ситуации и их значений. Фрейм имеет имя, служащее для идентификации описываемого им понятия, и содержит ряд слотов, с помощью которых описываются основные элементы этого понятия. Слот может содержать не только конкретное значение, но также имя процедуры, позволяющей вычислить это значение по заданному алгоритму. Процедуры, располагающиеся в слотах, называются присоединенными процедурами, их вызов происходит при обращении к слоту, в котором она помещена.

**Фрейм** (кадр, рамка) – это структура данных для представления типичной, привычной для мозга ситуации. Человек смотрит на предмет и подбирает наиболее близкий из известных ему фреймов (аналог). Это процесс согласования. Согласование может происходить в несколько этапов с остановками и промежуточной оценкой. Если оценка удовлетворяет, тогда терминалы фрейма заполняются заданиями, удовлетворяющими этим ограничениям. Если находится фрейм, точно соответствующий ситуации, происходит сопоставление. Если сопоставление невозможно, то идет процесс оправдания (фрейм подходит, но есть неточности). Далее идет советование (т.е. выработка дальнейших действий). При полной неудаче делается резюме – отказ от фрейма.

Если согласование прошло, идет детализация, элементы фрейма заполняются. Этот экземпляр становится образцом и помещается в память (обучение). Таким же образом происходит

анализ пространственных сцен, понимание смысла предложения (эвристические алгоритмы обработки речи).  
 На рис.3.7 представлены области применения фреймовой модели.



Рис.3.7. Характерные области применения фреймов

Структура данных фрейма содержать следующие атрибуты.

**Имя фрейма.** Служит для идентификации фрейма в системе и должно быть уникальным.

**Имя слота.** Должно быть уникальным в пределах фрейма. Может быть или назначено проектировщиком, или выбрано из системных, зарезервированных имен. Системные слоты служат для управления выводом во фреймовой системе. Примером такого слота может быть слот-указатель дочерних фреймов (IS-A).

**Указатели наследования.** Показывают, какую информацию об атрибутах слотов из фрейма верхнего уровня наследуют слоты с аналогичными именами в данном фрейме.

**Указатель типа данных.** Показывает тип значения слота. Наиболее типичные типы: frame – указатель на фрейм; real – вещественное число; integer – целое число; list – список.

**Значение слота.** Оно должно соответствовать указанному типу данных и условию наследования.

**Демоны.** Это присоединенная процедура, автоматически запускаемая при выполнении некоторого условия, при обращении к некоторому слоту. Типы демонов всегда связаны с условием

запуска процедуры. Демон с условием IF-NEEDED (ЕСЛИ-НУЖНО) запускается, если в момент обращения к слоту его значение не установлено. Демон типа IF-ADDED (ЕСЛИ-ДОБАВЛЕНО) запускается при попытке изменения значения слота. Демон IF-REMOVED (ЕСЛИ-УДАЛЕНО) запускается при попытке удаления значения слота.

**Присоединенная процедура.** Может быть значением слота и запускается по сообщению, переданному от другого фрейма. Демоны и присоединенные процедуры являются процедурными знаниями, объединенными вместе с декларативными в единую систему. Эти процедурные знания являются средствами управления выводом во фреймовых системах.

Совокупность данных предметной области может быть представлена множеством взаимосвязанных фреймов, образующих единую фреймовую систему, в которой объединяются декларативные и процедурные знания. Такая система имеет, как правило сложную иерархическую структуру, в которой фреймы соединены друг с другом с помощью видовых связей. На верхнем уровне такой иерархии находится фрейм, содержащий наиболее общую информацию, истинную для всех остальных фреймов. Для уменьшения информационной избыточности во фреймовых системах часто реализуется принцип наследования информации, позволяющей общую (глобальную для системы) информацию хранить в отдельном родительском фрейме, а во всех остальных фреймах указывать ссылку на место хранения этой информации.

Фундаментальная идея состоит в том, что свойства и процедуры, расположенные выше, являются более или менее фиксированными, поскольку они представляют те вещи или понятия, которые в большинстве случаев являются истинными для интересующей нас сущности.

В то же время, фреймы более нижних уровней имеют слоты, которые должны быть заполнены наиболее динамической информацией, подверженной частым изменениям. Если такая динамическая информация отсутствует из-за неполноты наших знаний о предмете, то слоты фреймов более нижних уровней заполняются данными, унаследованными от фреймов верхних уровней.



Ниже в качестве простого примера показан фрейм, описывающий человека (табл.3.2)

Фрейм: Человек

Табл.3.2.

Имя слота:	Значение слота
Класс:	Млекопитающее
Структурный элемент:	Голова, шея, руки, ...
Рост:	40 % 220 см
Масса:	1 % 200 кг
Хвост:	Нет
Язык:	Русский, английский, китайский ...
Фрейм аналогии:	Обезьяна

Для уменьшения информационной избыточности во фреймовых системах реализуется принцип наследования информации, позволяющей общую (глобальную для системы) информацию хранить в отдельном родительском фрейме, а во всех остальных фреймах указывать ссылку на место хранения этой информации. В качестве примера рассмотрим фреймы, приведенные на рис. 3.8.

Как следует из рис. 3.8 понятие «ученик», описываемое соответствующим фреймом, наследует свойства фреймов «ребенок» и «человек», которые находятся на более высоких уровнях иерархии.

При возникновении вопроса «Любят ли ученики сладкое?», будет получен ответ «да», потому что этим свойством обладают все дети, что указано во фрейме «ребенок». Наследование может быть частным, например, слот «возраст» для учеников не наследуется из фрейма «ребенок», так как явно указан в собственном фрейме.

Над фреймами можно совершать некоторые теоретико-множественные операции, например объединение или

пересечение. В первом случае в результирующем фрейме будут присутствовать все слоты, которые были в исходных. Во втором – результирующий фрейм будет содержать только одинаковые для всех фреймов слоты.

Человек

IS-A Млекопитающее  
Умеет Мыслить

Ребенок

IS-A  
Возраст 9-12 лет  
Рост 50-180 см  
Любит Сладкое

Ученик

IS-A Ребенок  
Учится В школе  
Возраст 7-17 лет  
Носит Форму

Рис. 3.8. Пример сети фреймов

Фреймовые системы подразделяются на статические и динамические, последние допускают изменение фреймов в процессе решения задачи.

### 3.5. Семантические сети для представления знаний

Под семантической сетью (СС) обычно подразумевают систему знаний некоторой предметной области, представленной в виде сети – ориентированного графа, узлы которого выражают понятия, события, а также свойства предметной области, а дуги являются описаниями их отношений. При этом все узлы и дуги могут быть снабжены метками, которые показывают, что именно они описывают [12].

Семантика – наука, устанавливающая отношения между символами и объектами, которые они обозначают. Это наука, определяющая смысл знаков. Семантическая сеть – это ориентированный граф, вершины (узлы) которого – понятия, а

дуги – отношения между ними. Узлы соответствуют объектам, концепциям или событиям, дуги чаще соответствуют понятиям «является», «составляет часть».

Формально семантическую сеть можно представить как  $SN = \langle X, R \rangle$ , где  $X$  – множество объектов сети, а  $R$  – множество отношений между ними.

Понятия представляют собой сведения об абстрактных или физических объектах предметной области.

События представляют собой действия, происходящие в реальном мире, определяются указанием типа действия и ролей, которые играют объекты в этом действии.

Свойства используются для уточнения понятий и событий. Применительно к понятиям они описывают их особенности и характеристики (цвет, размер и др.), а применительно к событиям – продолжительность, время, место.

Дуги графа сети отображают многообразие семантических отношений, которые можно условно разделить на четыре класса: лингвистические, логические, теоретико-множественные и квантифицированные.

Лингвистические отношения выражают смысловую взаимосвязь между событиями, между событиями и понятиями или свойствами. При этом лингвистические отношения бывают:

- глагольные (время, вид, род, залог, наклонение);
- атрибутивные (цвет, размер, форма);
- падежными.

Логические отношения представляют собой операции, используемые в исчислении высказываний: дизъюнкция, конъюнкция, инверсия, импликация.

Теоретико-множественные представляют собой отношение подмножества, отношение части и целого, отношение множества и элемента. Типовыми примерами здесь являются отношения включения или совпадения (IS-A) и отношение «целое-часть» (PART-OF)

Понятиями обычно выступают абстрактные или конкретные объекты, а отношения - это связи типа: «это» («is»), «имеет частью» («has part»), «принадлежит», «любит». Характерной особенностью семантических сетей является обязательное

наличия трех типов отношений:

- класс - элемент класса (цветок - роза);
- свойство - значение (цвет - желтый);
- пример элемента класса (роза - чайная).

В семантических сетях используются следующие отношения:

- связи типа «часть-целое» («класс-подкласс», «элемент-множество»);
- 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).

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

**Пример.** На рисунке изображен пример построения семантической сети. В данной семантической сети в качестве вершин – понятия: Человек, Ибрагимов, Нексия, Автомобиль, Вид транспорта, Двигатель, Цвет, Красный (рис. 3.9).



Рис.3.9. Пример построения семантической сети

Таблица 3.3

Квалификатор отношения	Объекты, с которыми связываются действия
Агент	Предмет, являющийся инициатором действия
Объект	Предмет, подвергающийся действию
Источник	Размещение предмета перед действием
Применик	Размещение предмета после действия
Время	Момент выполнения действия
Место	Место проведения действия
Цель	Действие другого события

При описании событий и действий с помощью семантической сети должны быть определены объекты, которые действуют, и объекты, над которыми эти действия выполняются. Связываются события и объекты с действиями с помощью падежных отношений (табл. 3.3)

На рис. 3.10 показан пример семантической сети, описывающей в виде событий и действий ситуацию, связанную с закупкой и доставкой товаров.

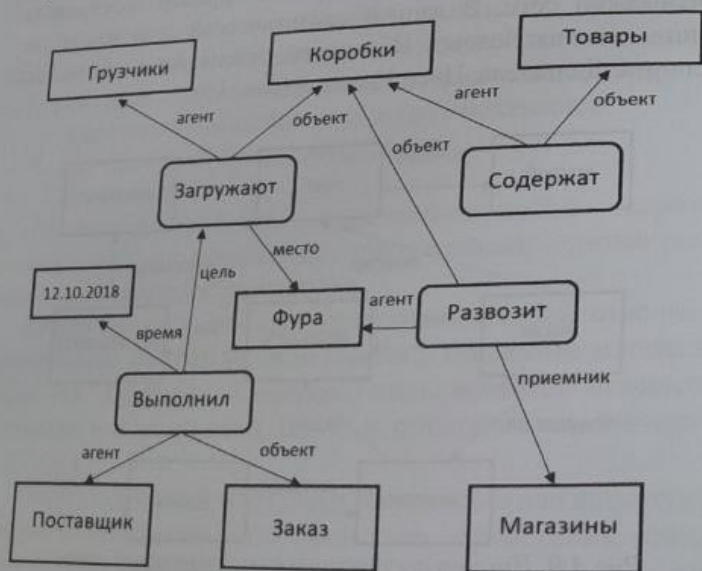


Рис. 3.10. Представление событий в виде семантической сети

Семантические сети обладают набором таких свойств, как хранение сведений об объектах системы и связях между ними, быстрый поиск объекта по заданным характеристикам, возможность обобщения и конкретизации знаний, которые обуславливают широкое использование среди исследователей в качестве средства умозаключений. Кроме этого, семантические сети находят применение при создании экспертных систем и систем распознавания речи

Основное преимущество этой модели представления знаний - в соответствии современным представлениям об организации долговременной памяти человека. Недостаток модели - сложность поиска вывода на семантической сети.

Для реализации семантических сетей существуют специальные сетевые языки, например NET. Широко известны экспертные системы, использующие семантические сети в качестве языка представления знаний PROSPECTOR, CASNET, TORUS.

#### Вопросы для контроля

1. Назовите этапы формирования знаний.
2. Какие существуют формы хранения знаний?
3. Что такое инженерия знаний?
4. Опишите стратегию получения знаний.
5. Перечислите модели представления знаний.
6. Каковы основные понятия моделей представления знаний?
7. В чем суть логической модели представления знаний?
8. Каковы особенности продукционной модели представления знаний?
9. Опишите фреймовую модель представления знаний.
10. Приведите пример семантической сети.

## ГЛАВА 4. ОСНОВЫ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ

### 4.1. Задачи описания нечетких множеств

Одним из ключевых проявлений человеческого мышления является способность обрабатывать нечеткую информацию, характеризующуюся неполнотой и неопределенностью, свойственным многим классам реальных проблем. Важным является создание эффективных методов для отображения таких нечеткостей реального мира и моделирования приближенных рассуждений человека в компьютерных системах [16]. Существует целый класс описаний объектов или процессов, оперирующих качественными характеристиками: много, мало, большой, очень большой. Эти характеристики обычно неоднозначны, хотя и содержат важную информацию, они не могут интерпретироваться полностью истинные или полностью ложные. Такие знания характеризуются некоторой промежуточной цифрой от 0 до 1.

Значительный вклад в эту область был сделан профессором Калифорнийского университета (Беркли, США) Л.Заде, создавшим математический аппарат теории нечетких множеств. С его помощью могут быть эффективно описаны нечеткие понятия и знания, а также выполняться операции над этими знаниями, нечеткие выводы. Широкое практическое применение теория нечетких множеств получила в таких задачах как автоматическое управление, принятие решений, представление и обработка знаний. Далее кратко рассматриваются основы этой теории. Л.Заде предложил для таких параметров ввести понятие «лингвистической переменной» (ЛП) – это переменная, значение которой определяется набором словесных характеристик. Значения лингвистических переменных определяется через так называемые нечеткие множества (НМ).

Нечеткое множество определяется через некоторую базовую шкалу  $V$  и функцию принадлежности нечеткого множества  $\mu(x)$ ,  $x \in V$ , принимающую значения на интервале  $[0,1]$ . Функция принадлежности определяет субъективную степень уверенности эксперта в том, что данное конкретное значение базовой шкалы соответствует определяемому нечеткому множеству.

**Основные понятия нечетких множеств.** В основе теории нечетких множеств лежит расширение классического понятия множества за счет возможности принимать функции принадлежности элемента множеству любых значений в интервале  $[0, 1]$ , а не только фиксированных значений 0 либо 1 [15,16]. Пусть  $U$  полное множество объектов некоторого класса. Нечеткое подмножество  $F$  множества  $U$ , именуемое далее нечетким множеством определяется через функцию принадлежности  $\mu_F(u)$ ,  $u \in U$ . Эта функция отображает элементы  $u$  множества  $U$  на множество вещественных чисел отрезка  $[0,1]$ , которые указывают степень принадлежности каждого элемента нечеткому множеству  $F$ .

Если полное множество  $U$  состоит из конечного числа множеств  $u_1, u_2, \dots, u_n$ , то нечеткое множество  $F$  можно представить в следующем виде:

$$F = \frac{\mu_F(u_1)}{u_1} + \frac{\mu_F(u_2)}{u_2} + \dots + \frac{\mu_F(u_n)}{u_n} = \sum_{i=1}^n \frac{\mu_F(u_i)}{u_i}$$

Важно, что в этом выражении знак «+» означает не сложение, а скорее объединение, символ деления показывает, что значение  $\mu_F(u_i)$  относится к элементу  $u_i$ , а не означает деление на него.

### 4.2. Реализация задачи представления и интерпретации

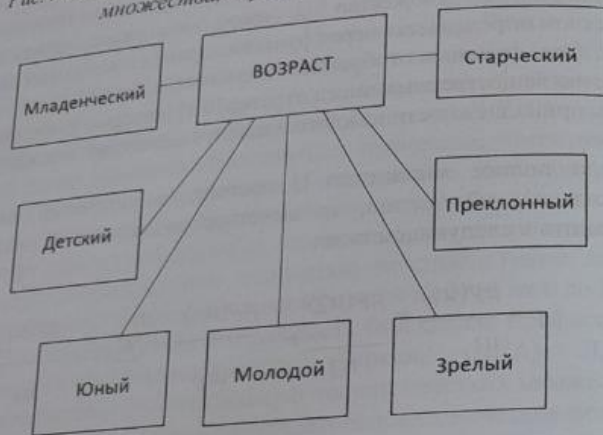
Рассмотрим понятную всем лингвистическую переменную «возраст» и нечеткие множества, которые могут входить в это понятие.

На рис.4.1 представлена интерпретация ЛП «возраст» и ее базовые значения – числовая шкала от 0 до 120 лет. Тогда функция принадлежности определяет, насколько мы уверены в том, что данное количество лет можно отнести к той или иной категории возраста.

На рис. 4.2. показан пример, как одни и те же значения базовой шкалы могут участвовать в определении различных

нечетких множеств.  
Формализуем неточное определение понятия младенческого возраста.

Рис. 4.1. Лингвистическая переменная «возраст» и нечеткие множества, определяющие ее значение



Определить нечеткое множество «младенческий возраст» можно так:

$$\text{«младенческий»} = \frac{1}{0.5} + \frac{0.9}{1} + \frac{0.8}{2} + \frac{0.7}{3} + \frac{0.5}{4} + \frac{0.3}{5} + \frac{0.1}{10}$$

Для правильной и корректной оценки нечеткого множества возьмем усредненное значение нескольких экспертов. При реализации алгоритмов и программ работы с нечеткими множествами такой прием очень часто применяется оценка усредненного эксперта (в числителе дробей – возраст ребенка, в знаменателе – оценка экспертов):

- ребенок до полугода с высокой степенью уверенности относится к категории младенцев ( $\mu=1$ );
- дети до четырех лет с меньшей степенью уверенности тоже причисляются к младенцам ( $0,5 \leq \mu \leq 0,9$ );
- десятилетнего ребенка к младенцам не относят.

На рис.4.3. представлен график функции принадлежности нечеткого множества «младенческий возраст».

Рис. 4.2. Пример формирования нечетких множеств

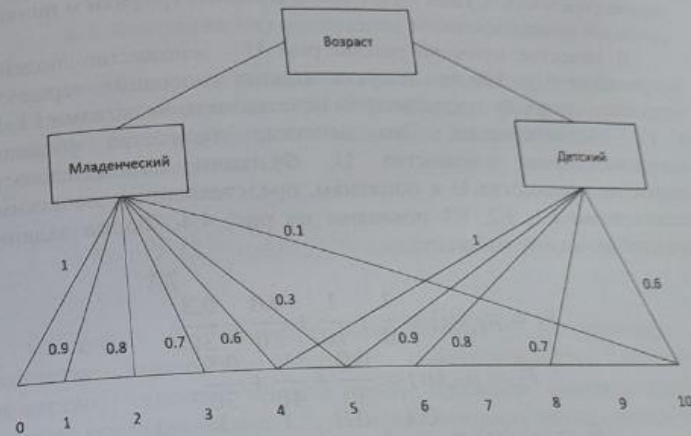
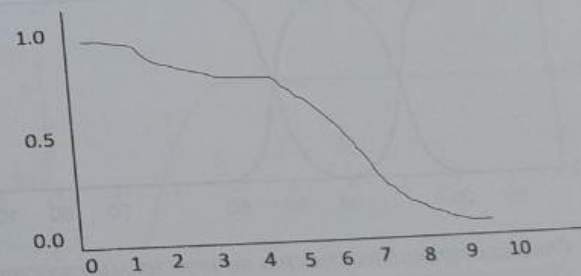


Рис. 4.3. График функции принадлежности НМ «младенческий возраст»



Это функция принадлежности класса L, выражения для которой приведены выше. В этом примере рассмотрен простейший способ формирования функции принадлежности на основе мнений ряда экспертов по построению графика функции принадлежности. Существуют более сложные графики и понятия функций принадлежности.

В качестве примера рассмотрим U – множество людей в возрасте от 0 до 100 лет, и пусть понятия «молодой», «среднего возраста» «старый» представлены нечеткими множествами F1, F2 и F3 соответственно. Эти нечеткие множества являются подмножествами множества U. Функции принадлежности элементов множества U к понятиям, представленным нечеткими множествами F1, F2, F3 показаны на рис. 4.4. Форма задания значений нечетких множеств:

$$F_1 = \mu_{F_1}(u) = \frac{1}{0} + \frac{1}{10} + \frac{0.8}{20} + \frac{0.3}{30}$$

$$F_2 = \mu_{F_2}(u) = \frac{0.5}{30} + \frac{1}{40} + \frac{0.5}{50}$$

$$F_3 = \mu_{F_3}(u) = \frac{0.4}{50} + \frac{0.8}{60} + \frac{1}{70} + \frac{1}{80} + \frac{1}{90}$$

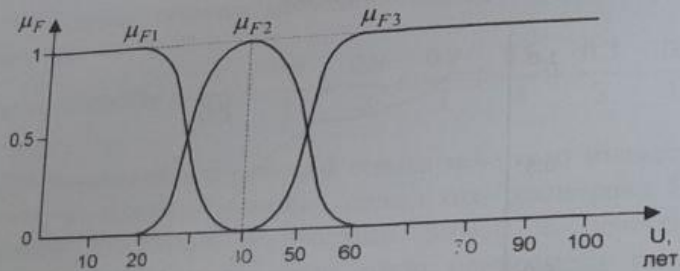


Рис. 4.4. Функции принадлежности нечетких множеств примера

Как следует из рис. 4.4 принадлежность к нечеткому множеству F1 (понятие «молодой»), составляет 1 для детей, 0,8 – для двадцати лет, степень принадлежности тридцатилетнего равна 0,3. При записи функции принадлежности элементы нечеткого множества со значениями  $\mu_F(u_i) = 0$  не включаются. Описаны

функции принадлежности для двух других нечетких множеств. Обычно функции принадлежности для нечетких множеств строятся субъективно по результатам опросов экспертов.

### 4.3. Операции над нечеткими множествами

Над нечеткими множествами можно выполнять теоретико-множественные операции [17,18]. Далее рассматриваются основные из них: дополнение, объединение и пересечение.

1. Операция дополнения:

$$\bar{F} = \sum_{i=1}^n (1 - \mu_F(u_i)) / u_i, \mu_{\bar{F}}(u) = 1 - \mu_F(u)$$

Следуя примеру, дополнением нечеткому множеству «молодой» будет соответствовать нечеткое множество понятия «немолодой», функция принадлежности которого показана на рис. 4.5, а математическая запись имеет вид:

$$\text{молодой} = \mu_{F_1} = \frac{0.2}{20} + \frac{0.7}{30} + \frac{1}{40} + \frac{1}{50} + \frac{1}{60} + \frac{1}{70} + \frac{1}{80} + \frac{1}{90}$$

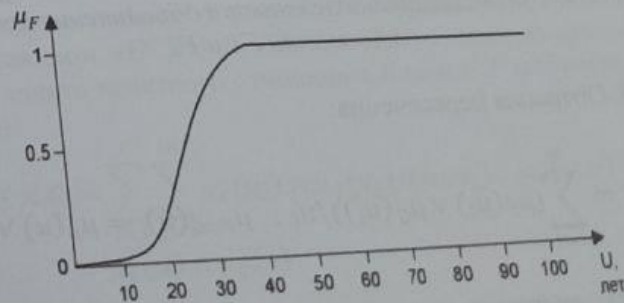


Рис. 4.5. Функция принадлежности дополнения нечеткого множества F1

2. Операция объединения:

$$F \cup G = \sum_{i=0}^n (\mu_F(u_i) \vee \mu_G(u_i)) / u_i, \quad \mu_{F \cup G}(u) = \mu_F(u) \vee \mu_G(u)$$

Здесь операция  $\vee$  соответствует взятию максимума. Для определения понятия, которому будет соответствовать объединение нечетких множеств  $F1$  и  $F2$ , вычислим функцию принадлежности:

$$(\text{молодой} \cup \text{средний}) = \mu_{F1 \cup F2}(u) = 1/0 + 1/10 + 0,8/20 + 0,5/30 + 1/40 + 0,5/50.$$

Ниже на рис. 4.6 приводится график этой функции, а ее лингвистической интерпретацией является понятие «человек молодого или среднего возраста».

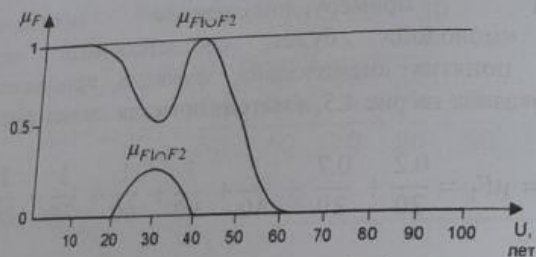


Рис. 4.6. Функция принадлежности объединения нечетких множеств  $F1$  и  $F2$

3. Операция пересечения:

$$F \cap G = \sum_{i=1}^n (\mu_F(u_i) \wedge \mu_G(u_i)) / u_i, \quad \mu_{F \cap G}(u) = \mu_F(u) \wedge \mu_G(u)$$

Здесь операция  $\wedge$  соответствует взятию минимума. Для определения понятия, которому будет соответствовать пересечение множеств  $F1$  и  $F2$ , вычислим функцию принадлежности:

$$(\text{молодой} \cap \text{средний}) = \mu_{F1 \cap F2}(u) = 0,3/30$$

Остальные члены этой функции, соответствующие значениями аргумента, кратным 10, равны нулю. На рис. 4.6 приводится график этой функции, а ее возможными лингвистическими интерпретациями являются понятия «уже не молодой, но еще не средний возраст», «одновременно молодой и средний возраст».

**Нечеткие отношения.** Часто понятия предметной области связаны различными отношениями. Для организации нечетких выводов необходимо формализовать понятие нечеткого отношения. Нечетким отношением  $R$  между полными множествами  $U$  и  $V$  называется нечеткое подмножество прямого декартова произведения  $U \times V$ , определяемое следующим образом:

$$R = \sum_{i=1}^l \sum_{j=1}^m \mu_R(u_i, v_j) / (u_i, v_j),$$

где  $U = \{u_1, u_2, \dots, u_n\}$ ,  $V = \{v_1, v_2, \dots, v_n\}$

Предположим, что между понятиями, представленными нечеткими множествами  $F \subseteq U$  и  $G \subseteq V$ , существует отношение, заданное правилом «ЕСЛИ  $F$ , ТО  $G$ ». Тогда один из способов построения такого нечеткого отношения  $R$  между  $F$  и  $G$  состоит в следующем:

$$R = F \times G = \sum_{i=1}^l \sum_{j=1}^m \mu_F(u_i) \wedge \mu_G(v_j) / (u_i, v_j), \quad \mu_R(u, v) = \mu_F(u) \wedge \mu_G(v)$$

Пусть  $U$  и  $V$  множества натуральных чисел от 1 до 4. Определим понятия «малые числа» и «большие числа» с помощью нечетких множеств  $F$  и  $G$  соответственно.

$$\begin{aligned}
 U = V &= \{1, 2, 3, 4\}; \\
 F &= 1/1 + 0.6/2 + 0.1/3; \\
 G &= 0.1/2 + 0.6/3 + 1/4.
 \end{aligned}$$

Пусть задано правило: «ЕСЛИ  $u$  – малое число, ТО  $v$  – большое» или  $F \rightarrow G$ . Построим для него соответствующее нечеткое отношение  $R = F \times G$ .

$$R = v_j$$

	$u_i$			
	0	0.1	0.6	1
	0	0.1	0.6	1
	0	0.1	0.6	1
	0	0	0	0

Наряду с рассмотренным есть и другие способы задания нечетких отношений.

**Композиция нечетких отношений.** Необходима при наличии цепочки правил, образующих знания в виде нечетких множеств, отношений. Выполнена она может быть с помощью операций свертки, которые включают максиминную, минимаксную, максимультимпликативную и другие операции.

Рассмотрим применение наиболее распространенной – первой из них. Пусть  $R$  – нечеткое отношение из области  $U$  в область  $V$ , а  $S$  – нечеткое отношение из области  $V$  в область множества  $W$ . Тогда нечеткое отношение из  $U$  в область  $W$  определяется как максиминная свертка:

$$R \circ S = \sum_{i=1}^l \sum_{k=1}^n \bigvee_{v_j \in V} \mu_R(u_i, v_j) \wedge \mu_S(v_j, w_k) / (u_i, w_k)$$

где  $\bigvee$  – знак взятия максимума для всех  $v_j$ ;  $\wedge$  – знак взятия минимума.

Пусть для  $W = \{1, 2, 3, 4\}$  определены следующим образом нечеткие множества:

$$\bar{F}(C V) = \text{«не малые числа»} = 0/1 + 0.4/2 + 0.9/3 + 1/4;$$

$$H(C W) = \text{«очень большие числа»} = 0/1 + 0/2 + 0.5/3 + 1/4.$$

Тогда, если есть правило «ЕСЛИ  $v$  не малое число, ТО  $w$  очень большое число», нечеткое отношение  $S$  из  $V$  в  $W$  согласно формуле  $R = F \square G$  определяется как:

$$S = v_j$$

	$w_k$			
	0	0	0	0
	0	0	0.4	0.4
	0	0	0.5	0.9
	0	0	0.5	1

Если далее вычислить по приведенной формуле свертку с нечетким отношением  $R$ , то из двух значений «ЕСЛИ  $u$  – малое число, ТО  $v$  – большое», «ЕСЛИ  $v$  не маленькое число, ТО  $w$  очень большое» получим следующее нечеткое отношение из  $U$  в  $W$ :

$$\begin{aligned}
 R \circ S &= \bigvee_{v \in V} \{ \mu_R(u, v) \wedge \mu_S(v, w) \} = \\
 &= \begin{bmatrix} 0 & 0.1 & 0.6 & 1 \\ 0 & 0.1 & 0.6 & 0.6 \\ 0 & 0.1 & 0.1 & 0.1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & 0.5 & 0.9 \\ 0 & 0 & 0.5 & 1 \end{bmatrix} = u_i
 \end{aligned}$$

	$w_k$			
	0	0	0.5	1
	0	0	0.5	0.6
	0	0	0.1	0.1
	0	0	0	0

Полученное нечеткое отношение показывает взаимосвязь областей  $U$  и  $W$ . При парных сравнениях элементов  $i$ -й строки и  $j$ -го столбца из них выбирается наименьший. Затем из четырех



минимальных элементов выбирается максимум, который является результатом и записывается в ячейку с координатами  $i, j$

**Нечеткие выводы.** Рассмотрим традиционный дедуктивный метод логического вывода Modus Ponendo Ponens для нечетких множеств. Его интерпретация «Если  $F$  – истина И импликация  $F$  тоже истина» то  $G$  – истина» в этом случае остается неизменной. Однако в среде нечетких множеств этот вывод записывается с некоторыми изменениями: из факта  $F'$  и правила  $F$  можно вывести  $G'$ . Здесь  $F', G'$  образуют нечеткие множества рассматриваемого правила вывода на полных множествах  $F, G$  соответственно.

Для организации нечеткого отношения из правила  $F \rightarrow G$  может быть использована формула  $R = F \square G$ , а при наличии цепочки из нескольких правил вывода – максиминная свертка. В свою очередь вывод  $G'$  определяется из максиминной свертки нечеткого множества  $F'$  и отношения  $R$ :

$$G' = F' \circ R = \sum_{i=1}^m v_{u_i \in U} (\mu_{F'}(u_i) \wedge \mu_R(u_i, v_i)) / v_i,$$

где  $F, F' \subset U; G, G' \subset V$ .

Пусть, продолжая предыдущий пример,  $U = V = \{1, 2, 3, 4\}$ ,  $F(\subset U) =$  «малые числа»,  $G(\subset V) =$  «большие числа». Кроме того, пусть в качестве исходной посылки для вывода задан факт « $u$  – число около 2», представленное нечетким множеством  $F' =$  «около 2» =  $0,3/1+1/2+0,3/3+0/4$ .

Пусть также задано правило  $F \rightarrow G$ : «ЕСЛИ  $u$  – малые числа, то  $v$  – большие», формализованное в виде ранее описанного отношения  $R$ . Используя комбинационное правило вывода (3.3), определим вывод  $G'$ , соответствующий ответу на вопрос «Что представляет собой  $v$ , если  $u$  – число около 2?» и если области  $U$  и  $V$  связаны отношением  $R$ :

$$G' = F' \circ R' = [0.3 \quad 1 \quad 0.3 \quad 0] \circ \begin{bmatrix} 0 & 0.1 & 0.6 & 1 \\ 0 & 0.1 & 0.6 & 0.6 \\ 0 & 0.1 & 0.1 & 0.1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = [0 \quad 0.1 \quad 0.6 \quad 0.6]$$

Здесь в результирующем векторе каждый элемент  $j$  представляет значение принадлежности  $v_j$  множества  $G'$ .  
На рис. 4.7 показан график функции принадлежности результата нечеткого вывода.

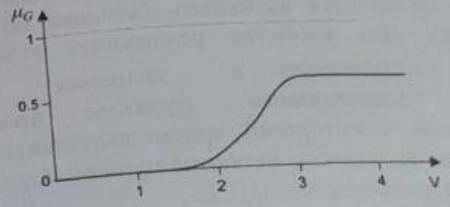


Рис. 4.7. Функция принадлежности результата нечеткого вывода

Исходя из графика можно предложить следующую лингвистическую интерпретацию: « $v$  – не очень большое число» или « $v$  – до некоторой степени большое число». При описании объектов и явлений с помощью нечетких множеств используется понятие нечеткой переменной, которая может принимать нечеткие значения.

**Нечеткая переменная** определяется тройкой  $\langle \alpha, U, F \rangle$ , где  $\alpha$  – наименование переменной;  $U$  – универсальное множество, область определения  $\alpha$ ;  $F$  – нечеткое множество на  $U$ , описывающее ограничения на возможные значения нечеткой переменной  $\alpha$ .

Обычно нечеткие переменные используются для моделирования различных систем, и, следовательно, над этими переменными приходится выполнять большие объемы разных операций. Для удобства реализации этого, а также организации ввода-вывода и хранения информации рекомендуется использовать функции принадлежности стандартного вида, с которыми проще выполнять расчеты. В частности, такими функциями являются трапециевидные (рис. 4.5), где  $\mu F(u)$  характеризуется  $\langle m, n, l, k \rangle$ . Как частный случай при  $m = n$  получается треугольная форма функции принадлежности

#### 4.4. Стандартные формы представления функций принадлежности

Обычно нечеткие переменные используются для моделирования зличных систем, и, следовательно, над этими переменными приходится выполнять большие объемы разных операций [19]. Для удобства реализации этого, а также организации ввода-вывода и хранения информации рекомендуется использовать функции принадлежности стандартного вида, с которыми проще выполнять расчеты. В частности, такими функциями являются трапецевидные (рис.4.8), где  $\mu F(u)$  характеризуется  $\langle m, n, l, k \rangle$ .

Как частный случай при  $m = n$  получается треугольная форма функции принадлежности.

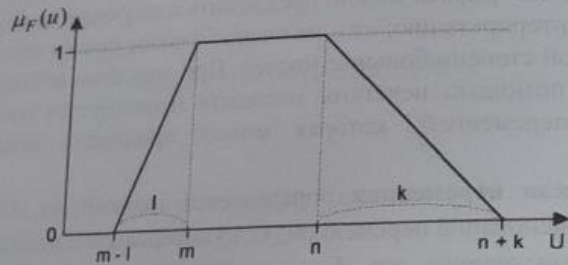


Рис.4.8. Трапецевидная функция принадлежности

В некоторых приложениях используются стандартные формы функций принадлежности. Рассмотрим их графические интерпретации.

1. Функция принадлежности класса S (рис.4.9) определяется как:

$$s(x; a, b, c) = \begin{cases} 0 & \text{для } x \leq a, \\ 2\left(\frac{x-a}{c-a}\right)^2 & \text{для } a \leq x \leq b, \\ 1 - 2\left(\frac{x-c}{c-a}\right)^2 & \text{для } b \leq x \leq c, \\ 1 & \text{для } x \geq c, \end{cases}$$

где  $b = (a + c)/2$ .  
Функция принадлежности, относящаяся к этому классу имеет графическое представление, напоминающее букву «S», причем ее форма зависит от подбора параметров  $a, b, c$ . В точке  $X = b = (a + c)/2$  функция принадлежности класса S принимает значение 0,5.

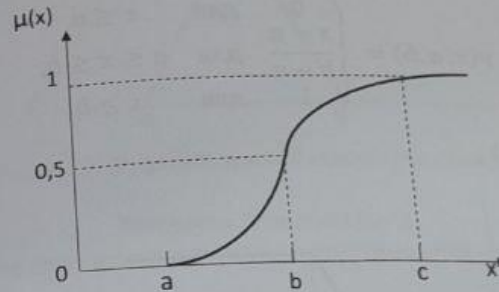


Рис.4.9. Функция принадлежности класса S

2. Функция принадлежности класса π (рис.4.10) определяется через функцию принадлежности класса S:

$$\pi(x; b, c) = \begin{cases} s(x; c - b, c - b/2, c) & \text{для } x \leq c, \\ 1 - s(x; c, c + b/2, c + b) & \text{для } x \geq c. \end{cases}$$

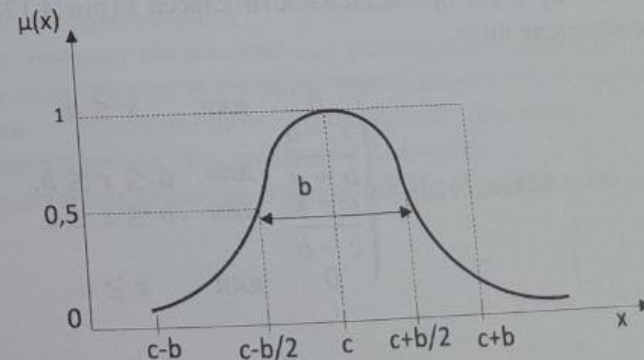


Рис.4.10. Функция принадлежности класса π

Функция принадлежности класса  $\pi$  принимает нулевые значения для  $x \geq c+b$  и  $x \leq c-b$ . В точках  $x = c \pm b/2$  ее значение равно 0,5.

3. Функция принадлежности класса  $\gamma$  (рис.4.11) задается выражением:

$$\gamma(x; a, b) = \begin{cases} 0 & \text{для } x \leq a, \\ \frac{x-a}{b-a} & \text{для } a \leq x \leq b, \\ 1 & \text{для } x \geq b. \end{cases}$$

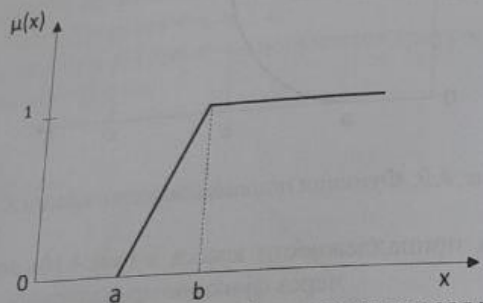


Рис.4.11. Функция принадлежности класса  $\gamma$

Из графиков функций принадлежности можно заметить аналогию между графиками классов  $S$  и  $\gamma$ .

4. Функция принадлежности класса  $t$  (рис.4.12) определяется в виде:

$$t(x; a, b, c) = \begin{cases} 0 & \text{для } x \leq a, \\ \frac{x-a}{b-a} & \text{для } a \leq x \leq b, \\ \frac{c-x}{c-b} & \text{для } b \leq x \leq c, \\ 0 & \text{для } x \geq c. \end{cases}$$

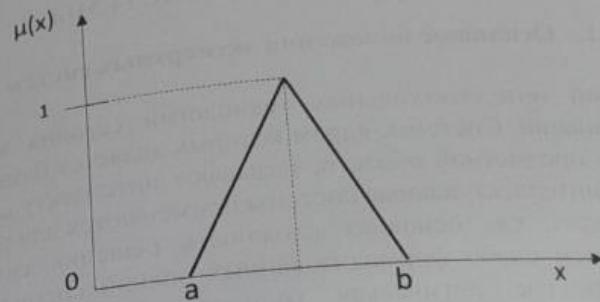


Рис.4.12. Функция принадлежности класса  $t$

### Вопросы для контроля

1. Дайте определение термину «лингвистическая переменная».
2. Опишите пример лингвистической переменной «возраст» и ее нечеткие множества.
3. Дайте пример нечеткого множества «младенческий возраст» и соответствующую функцию принадлежности.
4. Как реализуется операция дополнения над нечеткими множествами?
5. Как реализуется операция объединения над нечеткими множествами?
6. Как реализуется операция пересечения над нечеткими множествами?
7. Когда применяется композиция нечетких отношений?
8. В чем отличие одной функции принадлежности от другой?
9. Какова отличительная особенность функции принадлежности результата нечеткого вывода?
10. Представьте функции принадлежности классов «S» и «П».

## ГЛАВА 5. ЭКСПЕРТНЫЕ СИСТЕМЫ

### 5.1. Основные положения экспертных систем

Основой интеллектуальных технологий сегодня является обработка знаний. Системы, ядром которых является база знаний или модель предметной области, называют интеллектуальными. Чаще всего интеллектуальные системы применяются для решения сложных задач, где основная сложность решения связана с использованием слабо формализованных знаний специалистов-практиков и где логическая (или смысловая) обработка информации превалирует над вычислительной. Например, понимание естественного языка, поддержка принятия решения в сложных ситуациях, постановка диагноза и рекомендации по методам лечения, анализ визуальной информации, управление диспетчерскими пультами.

Технология экспертных систем является одним из направлений исследований в области искусственного интеллекта. Исследования в этой области сконцентрированы на разработке и внедрении компьютерных программ, способных эмулировать (имитировать, воспроизводить) те области деятельности человека, которые требуют мышления, определенного мастерства и накопленного опыта. К ним относятся задачи принятия решений, распознавания образов и понимания человеческого языка.

Перечень типовых задач, решаемых экспертными системами, включает: извлечение информации из первичных данных (радиотехнические сигналы); диагностика неисправностей (особенно в технических системах); анализ сложных объектов и процессов (медицина); распознавание и анализ сложных информационных систем (геоинформационные технологии); прогнозные и расчетные задачи (предсказание погоды).

Цель исследований состоит в разработке программ, которые при решении задач, трудных для эксперта-человека, получают результаты, не уступающие по качеству и эффективности решениям, получаемым экспертом. Исследователи в области ЭС для названия своей дисциплины часто используют также термин "инженерия знаний", понимаемый как "привнесение принципов и

инструментария исследований из области искусственного интеллекта в решение трудных прикладных проблем, требующих знаний экспертов".

Экспертной системой (ЭС) называется комплекс программ, которые решают трудно формализуемую задачу с использованием инженерии знаний в некоторой предметной области [21]. Неформализованные задачи обычно обладают следующими особенностями:

- неполнотой и противоречивостью исходных данных;
- ошибочностью и неполнотой знаний о проблемной области ирешаемой задаче;
- большой размерностью пространства решения, т.е. перебор при поиске решения весьма велик;
- динамически изменяющимися данными и знаниями.

Следует подчеркнуть, что неформализованные задачи представляют большой и очень важный класс задач. Многие специалисты считают, что эти задачи являются наиболее массовым классом задач, решаемых компьютерами. Экспертные системы и системы искусственного интеллекта отличаются от систем обработки данных тем, что в них в основном используются символьный (а не числовой) способ представления и эвристический поиск решения (а не исполнение известного алгоритма).

В настоящее время технология экспертных систем используется для решения различных типов задач (интерпретация, предсказание, диагностика, планирование, конструирование, контроль, отладка, инструктаж, управление) в самых разнообразных проблемных областях, таких, как финансы, нефтяная и газовая промышленность, энергетика, транспорт, фармацевтическое производство, космос, металлургия, горное дело, химия, образование, целлюлозно-бумажная промышленность, телекоммуникации и связь.

Экспертные системы это наиболее распространенный класс систем ИИ, ориентированный на тиражирование опыта высококвалифицированных специалистов в областях, где качество принятия решений традиционно зависит от уровня экспертизы, например, медицина, юриспруденция, геология, экономика, военное дело. Поэтому ЭС эффективны лишь в тех областях, где

важен эмпирический опыт специалистов.

Эти системы интегрируют опыт специалистов компании по ключевым и стратегически важным технологиям и распространяющие этот эмпирический опыт для консультирования менее квалифицированных пользователей.

Экспертная система отличается от прочих прикладных программ наличием следующих признаков [22].

Она моделирует не столько физическую природу определенной проблемной области, сколько механизм мышления человека применительно к решению задач в этой области. Это существенно отличает ЭС от систем математического моделирования или компьютерной анимации. При решении задач основное внимание уделяется воспроизведению компьютерными средствами методики решения проблем, которая применяется экспертом.

Система, помимо выполнения вычислительных операций, формирует определенные соображения и выводы, основываясь на тех знаниях, которыми она располагает. Знания в ЭС представлены, как правило, на некотором специальном, функционально ориентированном языке и хранятся отдельно от собственно программного кода, который и формирует выводы и соображения. Этот компонент программы принято называть базой знаний.

При решении задач основными являются эвристические и приближенные методы, которые, в отличие от алгоритмических, не всегда гарантируют успех. Эвристика, по существу, является правилом влияния (rule of thumb), которое в машинном виде представляет знание, приобретенное человеком по мере накопления практического опыта решения аналогичных проблем. Такие методы являются приближенными, они не требуют полноты исходной информации, существует определенная степень уверенности в том, что предлагаемое решение является верным.

Отличие ЭС от других видов программ из области искусственного интеллекта заключаются в следующем.

Экспертные системы имеют дело с предметами реального мира, операции с которыми обычно требуют наличия значительного опыта, накопленного человеком. Множество программ из области искусственного интеллекта являются сугубо

исследовательскими и основное внимание в них уделяется абстрактным математическим проблемам. Экспертные системы имеют ярко выраженную практическую направленность в научной или коммерческой области.

Одной из основных характеристик ЭС систем является оперативность в получении результата и его достоверность. Исследовательские программы искусственного интеллекта могут и не быть очень быстрыми, можно примириться и с существованием в них отказов в отдельных ситуациях, поскольку, так как это инструмент познания. ЭС система должна за нужное время найти решение, которое было бы не хуже, чем то, которое может предложить специалист в этой предметной области.

Экспертная система должна обладать способностью объяснить, почему предложено именно такое решение, и доказать его обоснованность. При этом пользователь должен получить всю информацию, необходимую ему для того, чтобы быть уверенным, что решение принято обоснованно. Экспертная система проектируется в расчете на взаимодействие с разными пользователями, для которых ее работа должна быть, по возможности, прозрачной.

Суммируя все сказанное, отметим – экспертная система содержит знания в определенной предметной области, накопленные в результате практической деятельности человека, и использует их для решения проблемно ориентированных задач. Процесс создания экспертной системы часто называют «инженерией знаний» (knowledge engineering) и он рассматривается в качестве "применения методов искусственного интеллекта".

## 5.2. Характеристики экспертных систем

В настоящее время в мире насчитывается несколько тысяч промышленных ЭС, которые дают советы:

- при управлении сложными диспетчерскими пультами, например сети распределения электроэнергии;
- при постановке медицинских диагнозов;
- при поиске неисправностей в электронных приборах, диагностике отказов контрольно-измерительного

- по проектированию интегральных микросхем;
  - по управлению перевозками;
  - по прогнозу военных действий;
  - по формированию портфеля инвестиций, оценке финансовых рисков, налогообложению.
- Причины широкого применения ЭС заключаются их основных характеристиках.

### Интегрированность

Разработанные инструментальные средства, легко интегрируются с другими информационными технологиями и средствами (с CASE, СУБД, контроллерами, концентраторами данных).

Открытость и переносимость. ЭС разрабатываются с соблюдением стандартов, обеспечивающих открытость и переносимость в рамках тех областей знаний, на которые они рассчитаны.

Использование языков традиционного программирования. Переход к языкам традиционного программирования (C, C++), упростил обеспечение интегрированности, снизил требования приложений ЭС к быстродействию компьютеров и объемам оперативной памяти.

Архитектура клиент-сервер. Разработаны ЭС, поддерживающие распределенные вычисления по архитектуре «клиент-сервер», что позволило снизить стоимость оборудования, используемого в приложениях, децентрализовать приложения, повысить надежность и общую производительность.

Проблемная ориентация. Переход от разработок ЭС общего назначения к проблемно ориентированным системам обеспечивает сокращение сроков разработки приложений, увеличение эффективности использования ЭС, упрощение и ускорение работы эксперта, повторную используемость информационного и программного обеспечения (объекты, классы, правила, процедуры).

### 5.3. Базовые функции экспертных систем

Поскольку теория ЭС выросла из более общей науки ИИ, то нет ничего удивительного в том, что проблематика этих областей имеет много общего.

Приобретение знаний. Существует высказывание, что приобретение знаний – это передача потенциального опыта решения проблемы от некоторого источника знаний и преобразование его в вид, который позволяет использовать эти знания в процессе функционирования ЭС. Передача знаний выполняется в процессе достаточно длительных собеседований между специалистом по проектированию экспертной системы (он называется инженером по знаниям) и экспертом в определенной предметной области, способным достаточно четко сформулировать имеющийся у него опыт. Исследователи рассматривают функцию приобретения знаний в качестве одной из главных проблем создания технологии экспертных систем [23]. Перечислим некоторые причины этого.

Специалисты в узкой области, как правило, пользуются собственным жаргоном, который трудно перевести на обычный язык пользователя. Но смысл жаргонного слова не очевиден, а потому требуется достаточно много дополнительных вопросов для уточнения его логического или математического значения (это особенно проявляется в медицине). Факты и принципы, лежащие в основе многих специфических областей знания эксперта, не могут быть четко сформулированы в терминах математической теории или модели, свойства которой хорошо понятны. Так, эксперту в финансовой области может быть известно, что определенные события могут стать причиной изменений на фондовой бирже, но он ничего не сможет сказать пользователю точно о механизмах, которые приводят к такому эффекту.

Для того чтобы решить проблему в определенной области, эксперту недостаточно просто обладать суммой знаний о фактах и принципах в этой области. Опытный специалист знает, какого рода информацией нужно располагать для формулировки того или иного суждения и как можно расчленить сложную проблему на более простые. Выявить в процессе собеседования такого рода

знания, основанные на личном опыте и плохо поддающиеся формализации, часто очень сложно. Экспертный анализ даже в очень узкой области, выполняемый человеком, очень часто нужно поместить в довольно обширный контекст, который включает и многие вещи, кажущиеся эксперту само собой разумеющимися, но не для пользователя (например, в юриспруденции). Очень трудно очертить количество и природу знаний общего рода, которые оказываются вовлечены в расследование того или иного дела. Представление знаний - еще одна функция экспертной системы. Предмет исследования в этой области — методы ассоциативного хранения информации как в мозгу человека. Основное внимание, естественно, уделяется логической, а не биологической стороне процесса.

#### 5.4. Структура и свойства экспертных систем

Основными участниками процесса интеллектуальной обработки являются: высококвалифицированный эксперт в данной предметной области, инженер формирования знаний, компьютерная система (рис. 5.1).

Наиболее полезным свойством ЭС является применение опыта профессионального квалифицированного эксперта. Система имеет возможность наращивать базу знаний за счет мнений эксперта при решении новых задач [11, 14].

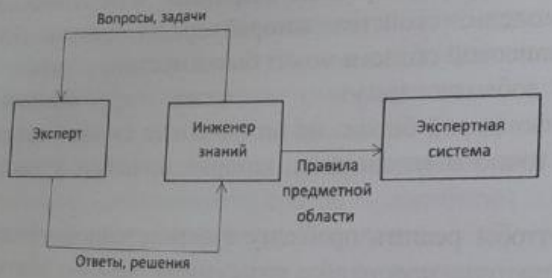


Рис. 5.1. Основные участники процесса обработки. Обычно ЭС используются как инструмент для

автоматизации работы эксперта. ЭС может также выступать в роли:

- консультанта для неопытных или непрофессиональных пользователей;
- ассистента эксперта-человека в процессах анализа вариантов решений;
- партнера эксперта в процессе решения задач, требующих привлечения знаний из разных предметных областей.

Также ЭС имеет возможность самообучаться на решаемых задачах, автоматически дополняя базу знаний результатами полученных выводов и решений.

В общем случае ЭС состоит из следующих основных компонент: базы знаний, рабочей памяти, решателя, системы объяснений, модуля приобретения знаний, интерфейса с пользователем (рис. 5.2).

Экспертная система — это программный комплекс, позволяющий на основе диалога с пользователем помогать в выборе решения путем оценки вариантов, предлагаемых пользователем, и их коррекции.

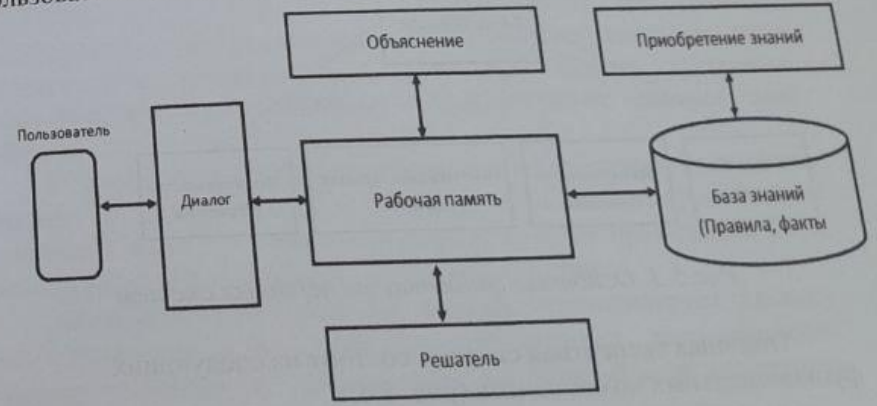


Рис. 5.2. Структура экспертной системы

Эксперт — человек, квалифицированный специалист, умеющий находить правильные решения в конкретной области знаний и исследований.

Пользователь — человек, который использует уже

построенную ЭС, и для которого эта система создана.  
Компьютерная часть ЭС, содержит набор прикладных программ интеллектуального анализа, а также программы поддержки: тестирующие, отладочные, редактирующие, средства графического ввода/вывода.

Многие правила ЭС являются эвристическими, т.к. задачи, которые решает ЭС, трудны для понимания, не до конца понятны и имеют несколько вариантов ответа. Общие знания о нахождении решений называются механизмом вывода. Структура механизма вывода (метода организации вывода) зависит от специфики предметной области и от правильной структуры знаний. В системе всегда присутствует интеллектуальный редактор базы знаний – программа, позволяющая добавлять, удалять, модифицировать факты и правила, содержащиеся в БЗ в конкретной ситуации.

Таким образом, основные свойства экспертных систем основаны на опыте эксперта, базе знаний и возможностях компьютерной обработки (рис. 5.3).

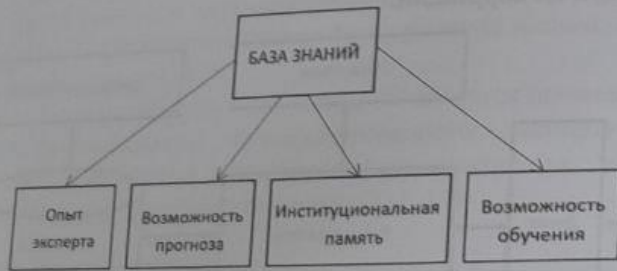


Рис. 5.3. Основные свойства экспертных систем

Типичная экспертная система состоит из следующих функциональных компонентов (рис. 5.4.):

- решателя (интерпретатора);
- рабочей памяти, называемой также базой данных (БД);
- базы знаний (БЗ);
- компонентов приобретения знаний;
- объяснительного компонента;
- диалогового компонента с экспертом и пользователем.

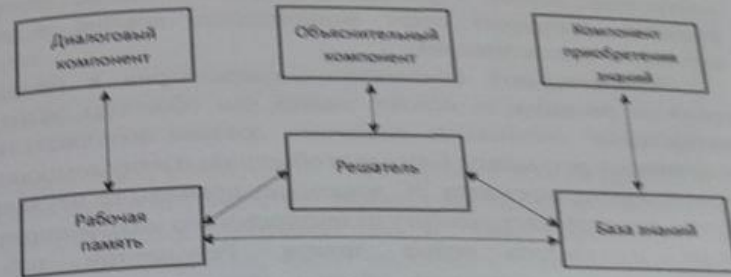


Рис. 5.4. Функциональные компоненты экспертной системы

База данных (рабочая память) предназначена для хранения исходных и промежуточных данных решаемой в текущий момент задачи. Этот термин совпадает по названию, но не по смыслу с традиционным термином СУБД, используемым в информационно-поисковых системах для обозначения всех данных, хранимых в системе.

База знаний (БЗ) в ЭС предназначена для хранения долгосрочных, накапливаемых при решении текущих задач данных, описывающих рассматриваемую область, и правил, описывающих целесообразные преобразования данных этой области.

Решатель – программа, читающая и использующая исходные данные из рабочей памяти и знания из БЗ, формирует такую последовательность правил, которые, будучи примененными к исходным данным, приводят к решению задачи.

Компонент приобретения знаний автоматизирует процесс наполнения ЭС знаниями, осуществляемый пользователем-экспертом. Любая ЭС должна иметь по крайней мере два режима работы.

В режиме «приобретение знаний» эксперт наполняет систему знаниями, обычно в виде совокупности данных и правил. В дальнейшем эти знания будут использоваться для самостоятельного решения ЭС задач из конкретной предметной области.

В режиме консультаций пользователь ЭС сообщает системе



конкретные данные о решаемой задаче, на основе которых решатель формирует ответ пользователю, обычно в виде некоторого предположения.

Объяснительный компонент — комментарии к процессу принятия решения на основе знаний что облегчает эксперту тестирование системы и повышает доверие пользователя к полученному результату. Благодаря объяснительному компоненту эксперт при тестировании ЭС локализует причины ее неудачной работы, что позволяет эксперту целенаправленно модифицировать старые и вводить новые знания. Результатом работы объяснительного компонента обычно является статистика по использованию из базы знаний правил и данных в процессе выполнения ЭС процедуры логического вывода.

Диалоговый компонент ориентирован на организацию дружественного общения с пользователем как в ходе решения задач, так и в процессе приобретения знаний и объяснения результатов работы. В настоящее время диалоговому компоненту экспертных систем уделяется большое внимание, так как от его возможностей зависит эффективность использования накопленных знаний и формирование новых знаний и данных. Диалоговый компонент позволяет преодолеть разницу в уровне понимания решаемой проблемы пользователем и эксперта.

Главное отличие ЭС от традиционных систем обработки состоит в том, что в них преобладает символьный, а не числовой способ представления данных, а в качестве методов обработки информации применяются процедуры логического вывода и эвристического поиска решений (табл. 5.1).

Таблица 5.1.

Характеристика	Разработка ЭС	Традиционные программы обработки
Тип обработки	Символьный	Числовой
Метод	Эвристический поиск	Точный алгоритм
Задание шагов решения	Неявное	Явное

Искомое решение	Удовлетворительное	Оптимальное
Управление и данные	Смешаны	Разделены
Знания	Неточные	Точные
Модификации	Частые	Редкие

Кроме этого, в отличие от традиционной программы ЭС при решении задачи выполняет не только предписанную алгоритмом последовательность действий, но и сама предварительно формирует их. Также ЭС имеет возможность самообучаться на решаемых задачах, автоматически дополняя базу знаний результатами полученных выводов и решений.

Главным структурным отличием ЭС от других типов программ является наличие базы знаний и способность к обучению и самообучению. Ее конкретный вид сильно зависит от избранной модели представления, но в наиболее общем виде она всегда будет содержать информационную (факты и данные) и алгоритмическую (машина вывода) части [6]. На рис.5.5 представлена схема взаимодействия подсистем типовой экспертной системы.

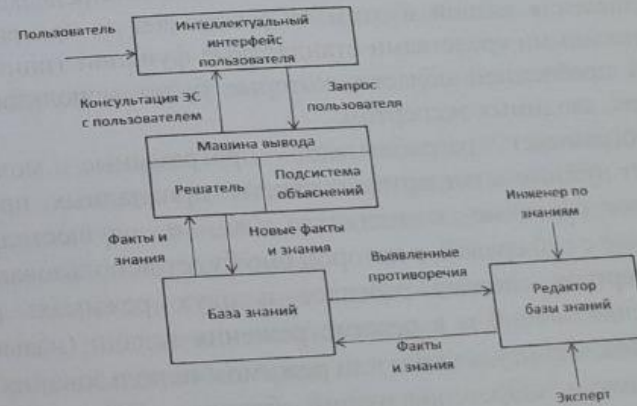


Рис.5.5. Технология работы типовой экспертной системы

Обязательным условием успешной работы ЭС является присутствие и участие в системе как эксперта, так и инженера по

формированию и редактированию базы знаний исходя из вновь выявляемых фактов, знаний и возникающих противоречий. Эффективное выполнение функций со стороны ЭС может быть обеспечено во многом за счет успешной работы интеллектуального интерфейса системы с пользователем.

В разработке и эксплуатации ЭС участвуют представители следующих специальностей [22]:

- эксперт в проблемной области, задачи которой будет решать ЭС;
- инженер по знаниям – специалист по разработке ЭС (используемые им технологии, методы называют технологией (методами) инженерии знаний);
- программист по созданию инструментальных средств, предназначенных для ускорения разработки ЭС.

Эксперт определяет знания (данные и правила), характеризующие проблемную область, обеспечивает полноту и правильность введенных в ЭС знаний.

Инженер по знаниям помогает эксперту выявить и структурировать знания, необходимые для работы ЭС; осуществляет выбор того инструментального средства, которое наиболее подходит для данной проблемной области, и определяет способ представления знаний в этом ИС; выделяет и программирует традиционными средствами стандартные функции типичные для данной проблемной области, которые будут использоваться в правилах, вводимых экспертом.

Программист разрабатывает программные модули и выбирает нужные стандартные пакеты прикладных программ, определяет основные компоненты ЭС, и осуществляет их сопряжение с той средой, в которой оно будет использовано.

Экспертная система работает в двух режимах: режиме приобретения знаний и в режиме решения задачи (называемом также режимом консультации или режимом использования ЭС).

В режиме приобретения знаний общение с ЭС осуществляет (через посредничество инженера по знаниям) эксперт. 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.

В режиме консультации общение с ЭС осуществляет конечный пользователь, которого интересует результат и способ его получения. В зависимости от назначения ЭС пользователь может не быть специалистом в данной проблемной области (в том случае он обращается к ЭС за результатом), или быть специалистом - в этом случае пользователь может сам получить результат. В режиме консультации 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 операции, но и предварительно формирует ее. Если реакция системы не понятна пользователю, то он может потребовать объяснения.

### 5.5. Представление знаний в экспертных системах

В соответствии с общей схемой ЭС (рис. 5.4) для ее функционирования требуются следующие знания [23]:

- знания о процессе решения задачи (управляющие знания), используемые решателем;
- знания о языке общения и способах организации диалога, используемые диалоговым компонентом;
- знания о способах представления и модификации знаний, используемые компонентом приобретения знаний;
- поддерживающие структурные и управляющие знания, используемые объяснительным компонентом.

Вопросы организации знаний необходимо рассматривать в любом представлении, и их решение в значительной степени не зависит от выбранного способа (модели) представления. Выделим

следующие аспекты проблемы организации знаний :

- организация знаний как по уровням представления, так и по уровням детальности;
- организация знаний в рабочей памяти;
- организация знаний в базе знаний.

Уровни представления и уровни детальности. Для того чтобы экспертная система могла управлять процессом поиска решения, была способна приобретать новые знания и объяснять свои действия, она должна уметь не только использовать свои знания, но и иметь знания о том, как представлены ее знания о проблемной среде. Если знания о проблемной среде назвать знаниями нулевого уровня представления, то первый уровень представления содержит метазнания, т.е. знания о том, как представлены во внутреннем мире системы знания нулевого уровня. Первый уровень содержит знания о том, какие средства используются для представления знаний нулевого уровня. Знания первого уровня играют существенную роль при управлении процессом решения, при приобретении и объяснении действий системы. В связи с тем, что знания первого уровня не содержат ссылок на знания нулевого уровня, знания первого уровня независимы от проблемной среды.

Выделение уровней детальности позволяет рассматривать знания с различной степенью подробности. Количество уровней детальности во многом определяется спецификой решаемых задач, объемом знаний и способом их представления. Как правило, выделяется не менее трех уровней детальности, отражающих соответственно общую, логическую и физическую организацию знаний. Введение нескольких уровней детальности обеспечивает дополнительную степень гибкости системы, так как позволяет производить изменения на одном уровне, не затрагивая другие.

**Организация знаний в рабочей памяти.** Рабочая память экспертных систем (рис.5.4) предназначена для хранения данных. Данные в рабочей памяти могут быть однородны или разделяются на уровни по типам данных. В последнем случае на каждом уровне рабочей памяти хранятся данные соответствующего типа. В современных экспертных системах данные в рабочей памяти хранятся как изолированные или как связанные. В первом случае

рабочая память состоит из множества простых элементов, а во втором - из одного или нескольких сложных элементов (объектов). При этом сложный элемент соответствует множеству простых, объединенных в единую сущность. Теоретически оба подхода обеспечивают полноту, но использование изолированных элементов в сложных предметных областях приводит к потере эффективности.

Данные в рабочей памяти являются константами или переменными. При этом переменные могут трактоваться как характеристики некоторого объекта, а константы - как значения соответствующих характеристик. Если требуется анализировать одновременно несколько различных объектов рабочей памяти, то необходимо указывать, к каким объектам относятся рассматриваемые характеристики. Если рабочая память состоит из сложных элементов, то связь между отдельными объектами указывается явно, например заданием семантических отношений. При этом каждый объект может иметь свою внутреннюю структуру.

**Организация знаний в базе данных.** Показателем интеллектуальности системы с точки зрения представления знаний считается способность системы использовать в нужный момент необходимые (релевантные) знания. Эта проблема является одной из основных причин, ограничивающих сферу применения экспертных систем. В проблеме доступа к знаниям можно выделить три аспекта:

- связность знаний и данных;
- механизм доступа к знаниям;
- способ сопоставления.

Связность или агрегация знаний является основным способом, обеспечивающим ускорение поиска релевантных знаний. Все знания, характеризующие некоторую сущность, связываются и представляются в виде отдельного объекта. При подобной организации знаний, если системе потребовалась информация о некоторой сущности, то она ищет объект, описывающий эту сущность, а затем уже внутри объекта отыскивает информацию о данной сущности. В объектах целесообразно выделять два типа связей между элементами: внешние и внутренние. Внутренние связки объединяют элементы в единый объект и предназначены для выражения структуры

объекта. Внешние связи отражают взаимозависимости, существующие между объектами в области экспертизы.

Основной проблемой при работе с большой базой знаний является проблема поиска знаний (механизма доступа), релевантных решаемой задаче. В связи с тем, что в обрабатываемых данных может не содержаться явных указаний на значения, требуемые для их обработки, необходим более общий механизм доступа, чем метод прямого доступа (метод явных ссылок). Задача этого механизма состоит в том, чтобы по некоторому описанию сущности, имеющемуся в рабочей памяти, найти в базе знаний объекты, удовлетворяющие этому описанию. Очевидно, что упорядочение и структурирование знаний могут значительно ускорить процесс поиска.

Нахождение желаемых объектов в общем случае уместно рассматривать как двухэтапный процесс. На первом этапе, соответствующем процессу выбора по ассоциативным связкам, совершается предварительный выбор в базе знаний потенциальных кандидатов на роль желаемых объектов. На втором этапе путем выполнения операции сопоставления потенциальных кандидатов с описаниями кандидатов осуществляется окончательный выбор искомых объектов.

Операции сопоставления весьма разнообразны, выделяются следующие формы: синтаксическое, параметрическое, семантическое и принуждаемое сопоставления.

В случае синтаксического сопоставления соотносят формы (образцы), а не содержание объектов. Успешным является сопоставление, в результате которого образцы оказываются идентичными. Обычно считается, что переменная одного образца может быть идентична любой константе (или выражению) другого образца. Иногда на переменные, входящие в образец, накладывают требования, определяющие тип констант, с которыми они могут сопоставляться. Результат синтаксического сопоставления является бинарным: образцы сопоставляются или не сопоставляются.

В параметрическом сопоставлении вводится параметр, определяющий степень сопоставления.

В случае семантического сопоставления соотносятся не сами

образцы объектов, а их функции. В случае принуждаемого сопоставления один сопоставляемый образец рассматривается с точки зрения другого. В отличие от других типов сопоставления здесь всегда может быть получен положительный результат. Вопрос состоит в силе принуждения. Принуждение могут выполнять специальные процедуры, связываемые с объектами. Если эти процедуры не в состоянии осуществить сопоставление, то система сообщает, что успех может быть достигнут только в том случае, если определенные части рассматриваемых сущностей можно считать сопоставляющимися.

#### Поиск решений в экспертных системах (задача решателя).

Методы решения задач, основанные на сведениях их к поиску, зависят от особенностей предметной области, в которой решается задача, и от требований, предъявляемых пользователем к решению. Особенности предметной области с точки зрения методов решения можно характеризовать следующими параметрами:

- размер, определяющий объем пространства, в котором предстоит искать решение;
- изменяемость области, характеризует степень изменчивости области во времени и пространстве;
- полнота модели, описывающей область, характеризует адекватность модели, используемой для описания данной области;
- определенность данных о решаемой задаче, характеризует степень точности и полноты данных.

Требования пользователя к результату задачи, решаемой с помощью поиска, можно характеризовать количеством решений, свойствами результата и способом его получения. Показатель количества решений может характеризовать одно или несколько решений. Свойства задают ограничения, которым должен удовлетворять полученный результат или способ его получения. Свойства могут определять и такие особенности, как время решения, объем памяти, используемой для получения результата, указание об обязательности использования каких-либо знаний и данных.

Сложность задачи, определяемая вышеприведенным набором параметров, варьируется от простых задач малой размерности с

неизменяемыми определенными данными и отсутствием ограничений на результат и способ его получения до сложных задач большой размерности с изменяемыми, ошибочными и неполными данными и произвольными ограничениями на результат и способ его получения.

Существующие методы решения задач, используемые в экспертных системах, можно классифицировать следующим образом:

- методы поиска в одном пространстве, предназначенные для использования при следующих условиях: области небольшой размерности, полнота модели, точные и полные данные;
- методы поиска в иерархических пространствах, предназначенные для работы в областях большой размерности;
- методы поиска при неточных и неполных данных;
- методы поиска, использующие несколько моделей, предназначенные для работы с областями, для адекватного описания которых одной модели недостаточно.

Предполагается, что перечисленные методы при необходимости должны объединяться для того, чтобы позволить решать задачи сложность которых возрастает одновременно по нескольким параметрам.

### 5.6. Классификация экспертных систем

В зависимости от своих функциональных свойств экспертные системы принято классифицировать по следующим признакам [2]:

- способу формирования решения;
- способу учета временного признака;
- виду используемых данных и знаний;
- числу используемых источников знаний.

По способу формирования решения ЭС можно разделить на анализирующие и синтезирующие. В системах первого типа выбирается решение из множества известных решений на основе анализа знаний, в системах второго типа решение синтезируется из отдельных фрагментов знаний.

В зависимости от способа учета временного признака ЭС

делят на статические и динамические. Большинство ЭС являются статическими, в них не учитываются изменения в окружающем мире за время решения задачи. Однако существует класс задач, где выполнения этого требования является обязательным. Такие ЭС относятся к динамическим, их связь с реальным миром осуществляется через систему контроллеров и датчиков. Следствием этого является существенное изменение в механизмах логического вывода для адекватного отражения временной логики происходящих в реальном мире событий.

По видам используемых данных и знаний различают ЭС с детерминированными и неопределенными знаниями. ЭС могут создаваться с использованием одного или нескольких источников знаний.

Рассмотрим классификацию по основным типам решаемых экспертной системой задач [2, 14].

Интерпретация данных – одна из традиционных задач для ЭС. Под интерпретацией данных понимается процесс определения смысла данных, результаты которого должны быть согласованными и корректными. Обычно подразумевается многовариантный анализ данных (идентификация силуэтов транспортных средств, психологическое тестирование).

Диагностика – процесс соотнесения объекта с некоторым классом объектов, обнаружение неисправностей (отклонений от нормы). Позволяет с единых теоретических позиций рассматривать неисправность оборудования в технических системах, болезни человека и животных, природные аномалии, искать ошибки в компьютерных программах.

Мониторинг – непрерывная интерпретация данных в реальном масштабе времени о выходе тех или иных параметров за допустимые пределы. Главная задача – контроль тревожной ситуации и ложного срабатывания (задачи диспетчерских служб, контроль аварийных ситуаций).

Проектирование – подготовка спецификаций на создание объектов с заранее определенными свойствами. Под спецификацией понимается весь набор необходимых документов и справочной информации – чертежей, стандартов, расчетных формул. Основная задача – получение четкого структурного описания знаний об создаваемом объекте. Для эффективного

проектирования и перепроектирования необходимо формировать не только сами проектные решения, но и мотивы их принятия. В задачах проектирования связываются два основных процесса ЭС: процесс решения и процесс объяснения (проектирование БИС система CADHELP).

Прогнозирование – предсказание последствий некоторых событий или явлений на основании имеющихся данных. Прогнозирующие ЭС выводят вероятные следствия из заданных ситуаций. Для этого обычно используется параметрическая динамическая модель, в которой значения параметров подгоняются под заданную ситуацию. Выводимые из этой модели следствия составляют основу для прогноза с вероятностными оценками предсказания погоды – система WILLARD, оценка будущего урожая – система PLANT.

Управление. Под управлением понимается функция организованной системы, поддерживающая определенный режим деятельности. Такого рода ЭС осуществляют управление поведением сложных систем в соответствии с заданными спецификациями. Примеры: помощь в управлении газовой котельной-GAS; управление системой календарного планирования Project Assistant и другие технологические системы.

Обучение – использование компьютера для обучения какой-либо дисциплине или предмету. Системы обучения диагностируют ошибки при изучении дисциплины и с помощью компьютера подсказывают правильные решения. Вначале формируются знания об ученике и его характерных ошибках, затем в работе эти знания помогают оценивать ошибки обучаемых и найти средства для их ликвидации. ЭС такого типа имеют развитый интерфейс общения с учеником (обучение языку программирования – учитель ЛИСП).

Поддержка принятия решения – это совокупность процедур, обеспечивающая лицо, принимающее решение, необходимой информацией и рекомендациями, облегчающими процесс принятия решения. ЭС помогают специалистам сформировать нужную альтернативу среди множества вариантов выбора при принятии ответственных решений (выход фирмы из кризисной ситуации – CRYISIS).

В общем случае ЭС и все системы, основанные на знаниях, можно разделить на системы, решающие задачи анализа, и на системы, решающие задачи синтеза.

Задачи анализа – интерпретация ситуаций и процессов, диагностика состояния и функционирования, поддержка принятия решения.

Задачи синтеза – разработка и проектирование объектов, изделий, планирование структур и процессов развития.

Комбинированные задачи – процессы обучения, мониторинг действия объектов и их промежуточных состояний, прогнозирование производства и развития процессов.

### 5.7. Проектирование экспертных систем

Разработка, проектирование и введение в пользование экспертных систем выполняется группой людей, включающей экспертов, инженеров по знаниям и программистов [21]. Инженер по знаниям помогает эксперту выявить и структурировать знания, необходимые для работы ЭС, определяет способ представления знаний в ЭС. Применяются известные способы экспертного оценивания: методы ранжирования, парных сравнений, непосредственной оценки.

Разработка ЭС имеет существенные отличия от разработки обычного программного продукта. Использовать ЭС следует только тогда, когда разработка ЭС возможна, а инженерии знаний соответствуют решаемой задаче. Чтобы разработка ЭС была возможной для данного приложения, необходимо выполнение следующих основных требований:

- имеются эксперты в данной области, которые решают задачу значительно лучше, чем начинающие специалисты;
- эксперты сходятся в оценке предлагаемого решения;
- эксперты способны выразить на естественном языке и объяснить используемые ими методы;
- задача не должна быть слишком трудной, ее решение должно занимать у эксперта установленное время;
- задача должна относиться к достаточно понятной и структурированной области, должны быть выделены основные понятия, отношения и известные способы получения решения

задачи.

Перед тем как приступить к созданию экспертной системы, инженер по знаниям должен рассмотреть вопрос о целесообразности разработки системы для данной области. Положительное решение принимается, если отсутствуют традиционные вычислительные средства решения задачи, а также имеется возможность извлечь и эффективно формализовать экспертные знания.

Среди исполнителей наиболее принципиальными являются такие специалисты, как эксперт и инженер по знаниям.

Эксперт – человек, способный ясно выражать свои мысли и пользующийся репутацией специалиста, умеющего находить правильные решения проблем в конкретной предметной области. Эксперт обладает особыми приемами, чтобы сделать поиск решения более эффективным. Инженер по знаниям – человек, как правило, имеющий познания в информатике и искусственном интеллекте и знающий как строить ЭС. Инженер по знаниям опрашивает экспертов, систематизирует знания, решает, каким образом они должны быть представлены в ЭС.

На рис. 5.6 показано взаимодействие основных участников создания и эксплуатации ЭС.

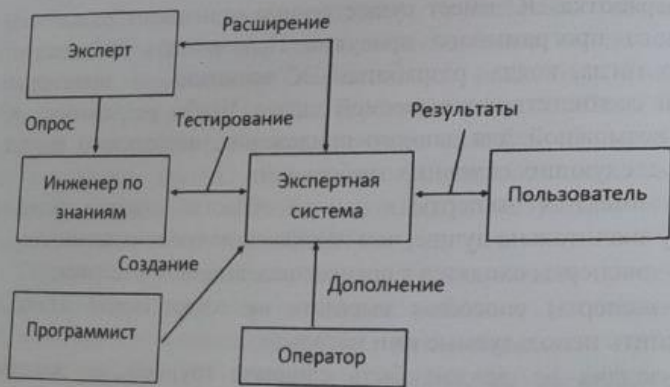


Рис. 5.6. Функции создателей и пользователей ЭС  
В основе технологии разработки ЭС лежат шесть этапов (рис.

5.7), включающих идентификацию, концептуализацию, формализацию, выполнение, тестирование и опытную эксплуатацию.



Рис. 5.7. Этапы технологии разработки ЭС

На этапе идентификации определяются задачи, подлежащие решению, цели разработки, эксперты и типы пользователей. На этапе концептуализации проводится содержательный анализ предметной области, выявляются используемые понятия и их взаимосвязи, определяются методы решения задач. Для определения основных понятий предметной области могут быть использованы методы, основанные на различных психологических эффектах.

Эффективность выполнения указанных этапов во многом определяется успешным формированием авторитетной группы экспертов и получением от них качественных знаний.

На этапе формализации выбираются инструментальные средства и способы представления видов знаний, формируются основные понятия, определяются способы интерпретации знаний.

На этапе выполнения осуществляется заполнение базы знаний. Так как основу ЭС составляют именно знания, то данный этап является наиболее важным и трудоемким. На нем происходит занесение в ЭС предварительно формализованных в понятной

системе знаний экспертов, а также возможно создание одного или нескольких опытных прототипов экспертной системы.

На этапе тестирования эксперт и инженер по знаниям в интерактивном режиме проверяют компетентность ЭС. Процесс тестирования продолжается до тех пор, пока эксперт не решит, что система достигла требуемого уровня компетентности.

На этапе опытной эксплуатации проверяется пригодность ЭС для передачи конечным пользователям. Полученные результаты могут показать необходимость в существенной модификации ЭС.

Как и разработка любого программного обеспечения, процесс создания ЭС не сводится к строгой последовательности перечисленных этапов. В ходе разработки возникает потребность неоднократно возвращаться на более ранние этапы и пересматривать принятые там решения.

#### Вопросы для контроля

1. Что называется экспертной системой?
2. Назовите области применения ЭС.
3. Представьте основные компоненты ЭС и участников процесса обработки данных.
4. Перечислите основные характеристики экспертных систем.
5. Каковы функциональные компоненты ЭС и их функции?
6. Рассмотрите технологию работы типовой ЭС.
7. Перечислите основные типы решаемых ЭС задач.
8. Какого типа знания бывают представлены в ЭС?
9. Расскажите об этапах технологии разработки ЭС.
10. В чем функции опытной эксплуатации?

## ГЛАВА 6. ОСНОВНЫЕ ПОНЯТИЯ И АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

### 6.1. Общая постановка задачи

Машинное обучение, раздел искусственного интеллекта – это процесс, в ходе которого вычислительная система обрабатывает большое число примеров, выявляет закономерности и использует их, чтобы прогнозировать характеристики новых данных. Машинное обучение – дисциплина, которая зародилась как часть теории статистики. Но сегодня она актуальна как никогда, так как позволяет извлекать знания из данных.

Машинное обучение (machine learning – ML) – передовая технологическая дисциплина, класс методов искусственного интеллекта, особенностью которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме. Основная задача машинного обучения – восстановление заранее неизвестной зависимости по выборке, составленной из пар «вход-выход» [31,32].

Цель машинного обучения – предсказать результат по входным данным. Чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат. На основе данных требуется построить неявную зависимость, т.е. построить алгоритм, способный для любого сочетания входных данных выдать квалифицирующий ответ. Реализуется принцип не аналитического, а эмпирического формирования решения. Для измерения точности ответов вводится оценочный функционал качества. На рис. 6.1. представлена общая схема процесса машинного обучения.

На вход системы поступают параметры объекта исследуемой предметной области. На первом этапе из набора параметров выделяются признаки, которые затем в качестве исходных данных используются моделью обработки. Модель работает на базе алгоритма обучения, который построен на обучающих данных.



Так как цель машинного обучения – предсказать результат по входным данным, то чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат. При обучении компьютера нужны три вещи: данные, признаки и алгоритм обработки.

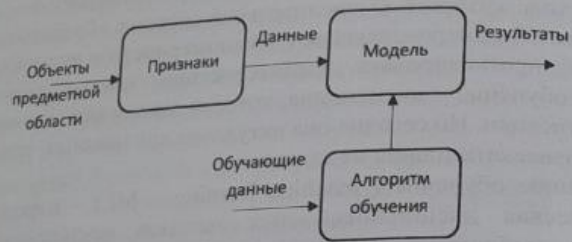


Рис.6.1. Схема машинного обучения

Входные данные являются набором параметров, характеризующих исследуемый объект. Если это изображение – то параметром будет набор пикселей, если это сигнал, то параметрами являются значения дискретных отсчетов во времени. Это могут быть статистические показатели в виде временных рядов. Данные могут быть представлены в виде цифровых показателей возраста, геометрических размеров тел, цвета, спектра.

В результате предварительного анализа из измеренных или считанных входных данных формируются признаки или наборы признаков исследуемого объекта. Признак – это некоторое количественное измерение, характеристика объекта произвольной природы. Совокупность признаков, относящихся к одному образу, называется вектором признаков. Считается, что каждому образу ставится в соответствие единственное значение вектора признаков и наоборот: каждому значению вектора признаков соответствует единственный образ объекта.

На этапе предварительного анализа реализуется генерация и селекция признаков: выбор тех признаков, которые с достаточной полнотой описывают объект исследования (генерация), отбор

наиболее информативных признаков (селекция). Наиболее информативные признаки объекта играют роль входных данных для математической модели обработки.

Модель – это численный метод, в результате работы которого происходит отображение признаков на результат. Математическая модель содержит переменные в виде входных данных и весовых регулируемых коэффициентов (обучающие данные). Алгоритмы машинного обучения можно описать как обучение целевой функции  $f$ , которая наилучшим образом соотносит входные переменные  $X$  и выходную переменную  $Y$ , т.е. реализует функцию  $Y = f(X)$ . Мы не знаем, что из себя представляет функция  $f$ , если бы знали, то использовали бы её напрямую, а не пытались обучить с помощью различных алгоритмов. В этом и есть суть эмпирически формируемого решения с использованием обучающих данных. Как было показано ранее, для оценки точности ответов вводится установленный функционал качества.

В реальных прикладных задачах входные данные об объектах могут быть неполными, неточными, нечисловыми, разнородными. Эти особенности приводят к большому разнообразию моделей и методов машинного обучения.

## 6.2. Основные методы машинного обучения

К настоящему времени разработано большое количество методов и алгоритмов машинного обучения. Разнообразие объектов моделирования и управления, вероятностные характеристики и количественное разнообразие входных данных способствовали разработке самых разных по сложности алгоритмов обучения. Такое же разнообразие наблюдается и при классификации методов. Мы рассмотрим наиболее широко применимые, относительно простые для понимания и успешно реализуемые с помощью нейронных сетей классические алгоритмы машинного обучения (рис.6.2). Из них наиболее широко применяемыми являются: классификация, регрессия и кластеризация [31,33].

Если выход представляет вещественную переменную, то задача называется задачей восстановления – регрессией.

Если выход принимает конечное число значений – это задача

классификации.  
Если входы и выходы – значения величин в некоторые моменты времени – это задача прогнозирования.

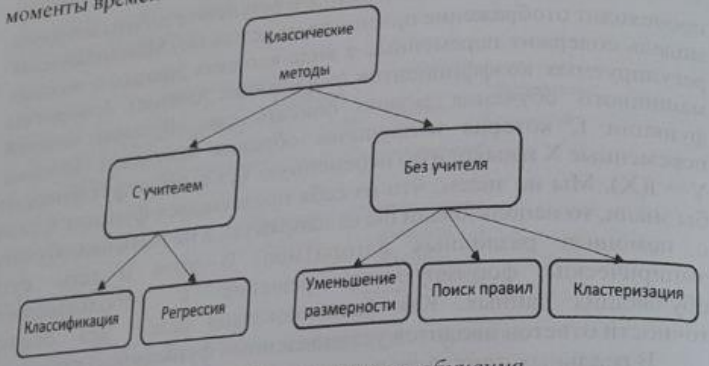


Рис. 6.2. Методы машинного обучения

Приложения задач машинного обучения:

- распознавание образов;
- интеллектуальный анализ данных (Data Mining);
- обработка естественных языков;
- компьютерное зрение, робототехника;
- медицинская диагностика;
- биоинформатика.

### 6.3. Задачи классификации объектов, обучение с учителем

При исследовании широкого типа задач считается, что все объекты или явления разбиты на конечное число классов [24,26,30]. Для каждого класса известно и изучено конечное число объектов – прецедентов. Прецедент – это образ ранее классифицированного объекта, принимаемый как образец при решении задач классификации. Идея принятия решений на основе прецедентности – основополагающая в естественно-научном мировоззрении. При такой постановке задачи каждому образу объекта ставится в соответствие единственное значение вектора

признаков и наоборот: каждому значению вектора признаков соответствует единственный образ исследуемого объекта.

Классификатором или решающим правилом называется правило отнесения образа к одному из классов на основании его вектора признаков.

В зависимости от наличия или отсутствия прецедентной информации различают задачи классификации с обучением и без обучения.

Задача классификации на основе имеющегося множества прецедентов называется классификацией с обучением (или с учителем). В том случае, если имеется множество векторов признаков, полученных для некоторого набора образов, но правильная классификация этих образов неизвестна, возникает задача разделения этих образов на классы по сходству соответствующих векторов признаков. Эта задача называется кластеризацией или распознаванием без обучения (без учителя).

Очевидно, что с учителем машина обучится быстрее и точнее, потому в реальных задачах его используют намного чаще. Эти задачи делятся на два типа: классификация – предсказание категории объекта, и регрессия – предсказание места на числовой прямой.

При классификации алгоритм разделяет объекты по заранее известному признаку. Примером может быть распознавание рукописных цифр или букв по клеткам изображения.

Для реализации процедур классификации в зависимости от характера решаемой задачи использовали известный из теории вероятностей «алгоритм Байеса» или алгоритм «дерево решений», где компьютер автоматически разделяет все данные по вопросам, ответы на которые «да» или «нет».

Сегодня для классификации всё чаще используют нейронные сети, первые разработки которых были ориентированы на решение задач классификации. Преимущество нейронных сетей заключается в том, что они предполагают наличие правил, с помощью которых сеть может программироваться автоматически.

Качество работы нейронной сети сильно зависит от вводимого в процессе обучения набора учебных данных, поэтому в начале обучения весовые коэффициенты устанавливаются равными случайным малым значениям. Расхождение между

входными и выходными, требуемыми данными, есть величина ошибки, которая может использоваться для корректировки весовых коэффициентов.

Среди задач обучения математические методы классификации включают несколько самостоятельных моделей. Рассмотрим наиболее широко применяемые методы  $k$ -ближайших соседей и метод бинарной классификации.

### Метод $k$ -ближайших соседей

Сознание человека при изучении объектов окружающего мира сопоставляет их с уже известными объектами и оценивает степень их сходства. На основе этой оценки объект относится к определённой группе, классу объектов.

Классификация является наиболее естественным для человеческого интеллекта способом получения знаний о процессах и явлениях.

Представителем методов классификации является метод « $k$ -ближайших соседей» ( $k$ -nearest neighbors algorithm - KNN).

При реализации данного метода предполагается, что уже имеется какое-то количество объектов, для которых известно, какому классу они принадлежат. Нужно выработать правило, позволяющее отнести новый объект к одному из возможных классов, классы известны.

В основе  $k$ -NN лежит правило: объект считается принадлежащим тому классу, к которому относится большинство его ближайших соседей. Под «соседями» здесь понимаются объекты, близкие к исследуемому в установленном смысле.

При решении данной задачи необходимо уметь определять, насколько объекты близки друг к другу, т.е. уметь измерять «расстояние» между объектами. Это не обязательно евклидово расстояние. Это может быть мера близости объектов, например, по цвету, форме, вкусу, запаху, весу, то есть по тем параметрам, которые характеризуют эти объекты.

Следовательно, для применения метода  $k$ -NN в пространстве признаков объектов должна быть введена некоторая метрика, т.е. функция расстояния. Предполагается, что объекты с близкими

значениями признаков относятся к одному классу.

Предсказание для новой точки делается путём поиска  $k$  ближайших соседей в наборе данных и суммирования выходной переменной для этих  $k$  экземпляров.

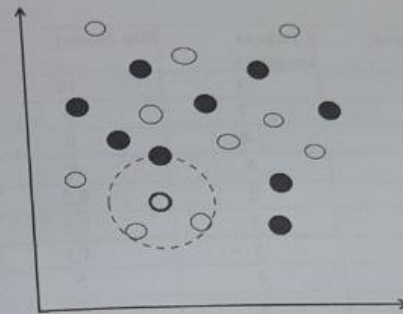


Рис. 6.3. Иллюстрация задачи классификации  $k$ -NN

На рисунке изображено пространство признаков двух классов (белые и черные кружки), неизвестный объект для распознавания изображен в виде круга с утолщенной окружностью.

В данном варианте решения задачи ближайших соседей выбрано  $k=3$ . Попавшие внутрь окружности три объекта (два белых и один черный) дают возможность сделать заключение, что неизвестный объект относится к категории белых кружочков. Другими словами, в данном примере предсказание сделано на основе трех ближайших соседей. Вопрос заключается в том, как определить сходство между экземплярами данных.

Если все признаки имеют один и тот же масштаб (например, сантиметры), то самый простой способ заключается в использовании евклидова расстояния – числа, которое можно рассчитать на основе различий (расстояний) с каждой входной переменной.

Рассмотрим работу метода  $k$ -NN на простом примере распознавания типов автотранспорта (табл.6.1). Метод  $k$ -NN относится к категории алгоритмов искусственного интеллекта,

широко применяемых при решении сложных, слабо структурированных задач обработки данных большого объема. Это алгоритм обучения с учителем, реализующий алгоритм классификации.

Табл.6.1.

Тип транспорта	Средняя скорость	Вес (тонн)	Класс автомобиля
1. Самосвал	4	10	Грузовик
2. Пожарная	7	9	Грузовик
3. Фура	4	9	Грузовик
4. Автобус	6	8	Грузовик
5. Эвакуатор	2	6,5	Спец-авто
6. Реанимобиль	3	5,5	Спец-авто
7. Доставка	4	5	Спец-авто
8. Фургон	4	6	Спец-авто
9. Маршрутка	10	5	Легковое авто
10. Малибу	9	5,5	Легковое авто
11. Джип	7	5	Легковое авто
12. Кобальт	6	3	Легковое авто
13. Спарк	6	3	Легковое авто

Здесь технические параметры автомобилей (средняя скорость и вес) оцениваются по 10-балльной шкале. Эти значения можно рассматривать как координаты точек (автомобилей) в 2-мерном пространстве. По горизонтальной оси откладывается один параметр, по оси ординат – другой параметр. В результате получается график, изображённый на рис.6.4.

Для каждого из названных автомобилей точно определен тип (класс) на последнем столбце таблицы. Можно заметить, что точки (марки автомобилей) на графике можно разбить на классы:

- в левом верхнем углу группируются легковые авто (марки Малибу, Джип, Кобальт, Спарк) – они легкие по весу и скоростные;
- в левом нижнем углу группируются автомобили специального применения (эвакуаторы, реанимобили экстренной помощи, машины доставки на дом, туристические автофургоны,

городские маршрутные такси);  
– справа сгруппированы тяжелые машины (самосвалы, пожарные, фуры-контейнеры, автобусы) – они тяжелые по сравнению с другими классами, но средние по скорости.

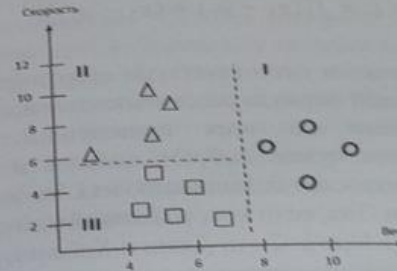


Рис.6.4. График размещения автомобилей по классам I, II, III.

Теперь предположим, что необходимо классифицировать автомобиль неизвестной модели, который размещен в центре рисунка в соответствии со своими параметрами веса и средней скорости (рис.6.5).

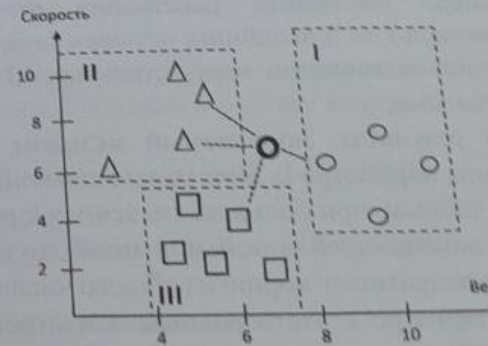


Рис.6.5. Размещение автомобиля неизвестной модели

Согласно методу k-NN неизвестная модель будет отнесена к тому классу, к которому принадлежит большинство из k -

ближайших соседей. Расстояние между объектами будет измеряться по Евклидовой норме, т.е. расстояние между объектами с координатами  $(x_1, y_1)$  и  $(x_2, y_2)$  равно:

$$L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Здесь значения  $x$  и  $y$  соответствуют координатам скорости и веса. По приведенной формуле рассчитываются все расстояния от неизвестной модели до всех приведенных на рисунке автомобилей. Теперь нужно выбрать число  $k$  и определить, к какому классу принадлежит большинство из  $k$  ближайших соседей этой модели. Так, если  $k=1$ , то ближайший сосед – один, и это Джип (легковая). При  $k=2$  это Джип и Кобальт, оба легковых автомобиля. При  $k=3$  мы имеем 2 легковых и один грузовой автомобиль. При  $k=4$  метод  $k$ -NN опять даст ответ: «легковушка». При  $k=5$  к двум легковым и одному грузовому прибавится одна спецмашина, но ответом будет «легковушка». При  $k=6$  в круг попадут две легковые, две спецмашины и один грузовик. Ответом метода  $k$ -NN будет «легковушка и спецмашина с равной вероятностью». Этот процесс можно продолжать и далее, увеличивая значение  $k$ .

Если теперь вычислить расстояния от новой точки (неизвестная модель) до ближайших четырех моделей, то из всех расстояний от новой точки до всех остальных 13 точек графика будут минимальными.

Очевидно результат, получаемый методом  $k$ -NN, сильно зависит от выбора параметра  $k$ . При малых значениях  $k$  результат может быть случайным, при больших значениях  $k$  результат будет зависеть числа автомобилей в той или иной группе. Обычно  $k$  берется равным квадратному корню из общего числа автомобилей. В приведенном примере  $k$  взято равным 4 и ответом будет «легковой автомобиль».

Решение задачи классификации упрощается, если классы признаков имеют четко разделенные области представления. Задача заключается в выработке правил деления областей, т.е. вывести функцию классификации путем построения линии,

разделяющей эти два класса.

### Бинарная классификация методом опорных векторов

В данном случае решается задача бинарной (когда класса всего два) классификации. Сначала алгоритм тренируется на объектах из обучающей выборки, для которых заранее известны метки (признаки) классов. Далее уже обученный алгоритм предсказывает признак класса для каждого объекта из тестовой выборки. Метки классов могут принимать значения  $Y = (+1, -1)$ . Объектом является вектор с  $N$  признаками  $x = (x_1, x_2, x_3, \dots, x_n)$ .

При обучении алгоритм должен построить функцию  $F(x) = y$ , которая принимает аргумент  $x$  (объект из пространства) и выдает метку класса  $y$ .

Данная задача относится к обучению с учителем, применяется алгоритм метода опорных векторов (SVM – Support Vector Machines). Данный алгоритм может применяться и для задач регрессии, но в данном случае он выполняет функции классификатора [32].

Главная цель SVM-классификатора – найти уравнение разделяющей гиперплоскости:

$$w_1x_1 + w_2x_2 + \dots + w_nx_n + w_0 = 0,$$

которое бы разделило два класса оптимальным образом. Общий вид преобразования  $F$  объекта  $x$  в метку класса  $Y$ :

$$F(x) = \text{sign}(w^T x - b) \quad [26].$$

Ранее были введены обозначения переменных:  $w = (w_1, w_2, \dots, w_n)$ ,  $b = -w_0$ .

После настройки весов алгоритма  $w$  и  $b$  в процессе обучения все объекты, попадающие по одну сторону от построенной гиперплоскости, будут предсказаны как первый класс, а объекты, попадающие по другую сторону – как второй класс.

Пример реализации алгоритма.

Задача заключается в выработке правил классификации самолетов (бомбардировщиков и истребителей) в зависимости от

их максимальной скорости и максимального взлетного веса. Формально такие правила записываются в виде:

ЕСЛИ вес  $> 0.80$  И скорость  $< 0.55$ , ТО это бомбардировщик; ЕСЛИ вес  $< 0.90$  И скорость  $> 0.25$  ТО это истребитель.

Эти правила используют дискретные граничные значения, разделяющие пространство всех значений на прямоугольные области. Разделение, порожденное этими правилами, успешно классифицирует самолеты, представленные на нашей диаграмме, но оказывается не слишком гибким, если по этим правилам придется классифицировать новый самолет.

Необходимо вывести функцию классификации путем построения прямой, разделяющей эти два класса (рис. 6.6).

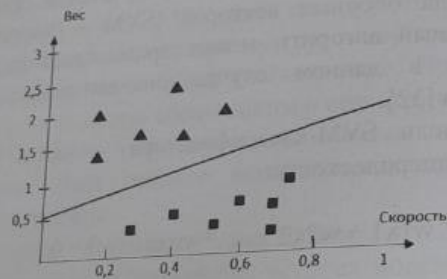


Рис. 6.6. Разделение абстрактных данных на два класса

Для нового самолета нужно указать точку на плоскости, соответствующую известным значениям максимальной скорости и максимального взлетного веса и посмотреть, по какую сторону от прямой будет расположена эта точка. В данном случае рассмотрены всего два признака (скорость и взлетный вес), поэтому данные можно представить в виде изображения на плоскости. Уравнение прямой, полученной показанным выше методом и разделяющей два типа самолетов, записывается в виде:

$$x_2 = 1,5 x_1 + 0,5 x_2,$$

где  $x_1$  представляет скорость, а  $x_2$  – вес. Это уравнение можно

использовать для создания функция выбора решения:

$$f(x_1, x_2) = -x_2 + 1.5x_1 + 0.5$$

$$d = \begin{cases} \text{истребитель,} & \text{если } f(x_1, x_2) \geq 0 \\ \text{бомбардировщик,} & \text{если } f(x_1, x_2) < 0 \end{cases}$$

Например,  $d(0.4, 0.5)$ , даст

$$f(0.4, 0.5) = -0.5 + 1.5 * 0.4 + 0.5 = 0.6$$

истребитель, представленный точкой (0.4,

и функция выбора решения правильно классифицирует эту точка, как истребитель.

Реализация данного принципа классификации с применением нейронной сети и корректировкой ее весовых коэффициентов приведена подробно в главе 9 «Обучение нейронных сетей» [31]. Более сложный пример дан в главе 12.

### Теорема Байеса (метод *Maive Bayes*)

Теорема Байеса – один из законов теории вероятности [28]. Она помогает нам пересматривать и понимать вероятности, когда возникают новые свидетельства. То есть, теорема дает количественную оценку того, насколько должны измениться наши убеждения в связи со вновь обнаруженными доказательствами.

Прежде чем рассмотреть реализацию алгоритма по формуле Байеса, исследуем простой способ представить саму идею метода без формул, т.к. применение теоремы не интуитивно, по крайней мере, для большинства людей. Визуализация использования теоремы намного упрощает понимание сути идеи Байеса и способствует применению ее в жизни.

Для упрощения графического представления задачи и ее решения опытные объекты лучше представить в виде квадратных фигур с неким содержимым.

Пусть имеется несколько возможностей, которые одинаково вероятны, их удобно представить в виде прямоугольника

разделенного на равные части (рис. 6.7). Возможные события и их вероятности представлены в виде фигур.

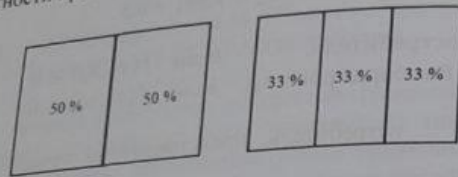


Рис. 6.7. Возможности событий и их вероятности

В следующем примере на рис. 6.8 приведены два объекта с равновероятным исходом: А и В.

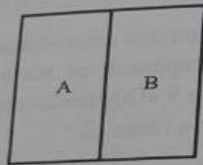


Рис. 6.8. Два объекта с равновероятным исходом

Представим, что каждая возможность – это коробка с 10 конфетами. В коробке А – 10 шоколадных конфет. Чтобы продемонстрировать это, мы заштрихуем ее. В коробке В – 5 шоколадных конфет и 5 ирисок. Мы разделим коробку пополам, и заштрихуем только часть с шоколадными конфетами.

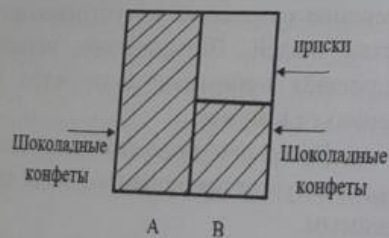


Рис. 6.9. Две коробки с разными типами конфет

Из графического представления видно, что заштрихованные области представляют все шоколадные конфеты в обеих коробках, в то время как белая область представляет все ириски.

Представим себе, что содержимое обеих коробок перемешано, наугад вытащена одна конфета и она оказалась шоколадной. Если бы была необходимость угадать, из какой коробки вытащена шоколадная конфета, большинство выбирало бы коробку А, так как в коробке А их в два раза больше! Это результат оценки мозга человека.

Это пример очень простого, естественного использования теоремы Байеса. В результате выборки шоколадной конфеты исчезла вероятность выбрать ириску. Чтобы визуализировать это, удалим часть коробки, которая относилась к ирискам (рис. 6.10).

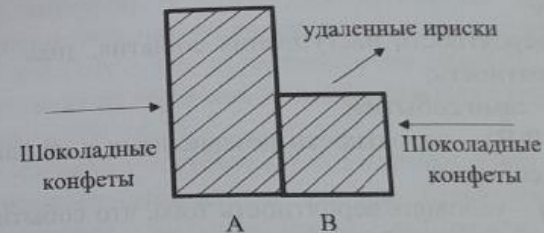


Рис. 6.10. Части объемов коробок с шоколадными конфетами

Теперь в коробке А вдвое больше конфет, чем в коробке В. Если разбить коробки на равные секции, образуется 3 равные области: 2 секции в коробке А и 1 секция в коробке В (рис. 6.11).

При произвольном извлечении шоколадная конфета из коробки В может быть выбрана с вероятностью  $1/3$ , или 33%. Конфета из коробки А может быть выбрана с вероятностью  $2/3$ , или 66%. Эта разница в вероятности – то, что человеческий мозг примерно рассчитал раньше. Именно это – причина, по которой была выбрана коробка А. Только что мы продемонстрировали концепцию теоремы Байеса и решена проблема без использования формулы. Теперь, прежде чем решать эту же проблему с помощью формулы, введем определение теоремы Байеса.

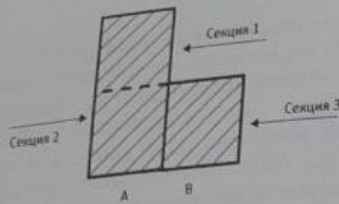


Рис. 6.11. Три равные секции в двух коробках

Формула содержит три компонента, комбинирование которыми позволяет решать задачи:

$$P(A|B) = P(B/A) * P(A) / P(B)$$

Обозначения:

- P – вероятность наступления события, знак условная вероятность;
- A, B – сами события;
- P(A), P(B) – вероятности независимых и не влияющих друг на друга событий;
- P(A/B) – условная вероятность того, что событие A истинно, если истинно событие B;
- P(B/A) – условная вероятность того, что событие B истинно, если истинно событие A.

Использование формулы Байеса предполагает, что нужно определить все три компонента формулы, включить их в формулу и получить обновленную вероятность, основанную на новой информации. Искомый ответ будет P(A|B), который называется апостериорной вероятностью и является нормализованным взвешенным значением.

Во вступлении задача с конфетами решалась без использования формулы. Сейчас мы используем Байеса и помня, что всего в обеих коробках 20 конфет, из них 15 шоколадных (рис. 6.9).

Используем четыре шага, чтобы найти ответ с помощью формулы Байеса. Шаг 1. Для начала мы должны определить, что мы хотим найти. Мы хотим знать вероятность того, что выбрали коробку A с учетом того, что мы выбрали шоколадную конфету. Шаг 2. Запишем поиск решения в виде формулы для данной задачи:

$$P(\text{кор. A шоко. конф}) = P(\text{шоко. конф кор. A}) / P(\text{шоко. конф})$$

Шаг 3. Найдем каждый компонент. P(коробка A) = 0,5. Чтобы ответить на этот вопрос, необходимо определить какова вероятность выбрать из коробки A. Эта вероятность не зависит от других событий. Поскольку существует только две коробки – A и B (рис. 4), и вероятности их выбора равны, ответ 0,5.

P(шоко. конфета) = 0,75. Для ответа на этот вопрос необходимо определить какова вероятность того, что мы выберем шоколадную конфету. Эта вероятность не зависит от других событий. Всего в обеих коробках 20 конфет, из них 15 шоколадных (рис. 6.9). Итак,  $15/20 = 0,75$ .

P(шоко. конфета | коробка A) = 1. Чтобы ответить на этот вопрос необходимо определить какова вероятность выбора шоколадной конфеты, учитывая, что мы выбрали из коробки A. Поскольку в коробке A есть только шоколадные конфеты, вероятность равна 1. Подставим все компоненты в формулу:

$$P(\text{кор. A шоко. конф}) = 1 * 0,5 / 0,75 = 0,66 = 0,66\%$$

Итак, существует вероятность 66%, что мы выбрали из коробки A, если выбранная конфета была шоколадной.

Автор знаменитой формулы Томас Байес родился в Лондоне в 1702 году. Он изучал логику и теологию в Эдинбургском университете, любил математику и увлекался теорией вероятности. Именно в 1740-х годах сформулировал решение, которое можно понимать следующим образом: первоначальное мнение + новое доказательство = новое суждение. В 1774 году французский математик Пьер-Симон Лаплас повторил решение Байеса. В то время он ничего не знал об открытии Байеса и пришел к нему независимо.

Во время Второй Мировой войны формула была



использована Аланом Тьюрингом, чтобы взломать немецкий шифрокод, но эта информация была засекречена в течение длительного времени. Только в 1980-х годах с появлением персональных компьютеров теорема Байеса получила широкое признание. Сейчас она широко используется во многих отраслях и касается нашей повседневной жизни.

### Логистическая регрессия (Logistic Regression)

В математической статистике логистическая регрессия является широко применяемой статистической моделью. Регрессия в общем виде применяется, когда входные и выходные переменные непрерывные. А логистическая регрессия лучшим образом подходит, когда выходная переменная принимает только два значения.

В отличие от множественной регрессии, которая игнорирует ограничения на диапазон значений выходной переменной, задача логистической регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной мы предсказываем непрерывную переменную со значениями на отрезке  $[0,1]$  при любых значениях независимых переменных. Это достигается применением следующего регрессионного уравнения:

$$P = 1 / (1 + e^{-y})$$

где  $p$  – вероятность того, что произойдет интересующее событие;  $e$  – основание натуральных логарифмов 2,71...;  $y$  – стандартное уравнение регрессии:

$$y = W_0 + W_1x_1 + W_2x_2 + \dots + W_nx_n$$

Это уравнение позволяет комплектовать все  $n$  признаков на одну ось абсцисс и вычислять значение признака.

При построении графической зависимости регрессионного уравнения учитываются особенности сигмоиды: при  $y=0$  значение  $e^{-y}$  обращается в 1, величина  $P$  становится равной 0,5. Кроме того, результаты подсчета коэффициентов влияют на характер кривизны сигмоиды. Массив  $X$  – параметры объекта наблюдения,

$W$  – это массив коэффициентов. Путем подбора коэффициентов выбирается характер кривой, наилучшим образом описывающей конкретные данные машинного обучения.

Зависимость, связывающая вероятность события и величину  $X$ , показана на следующем графике:

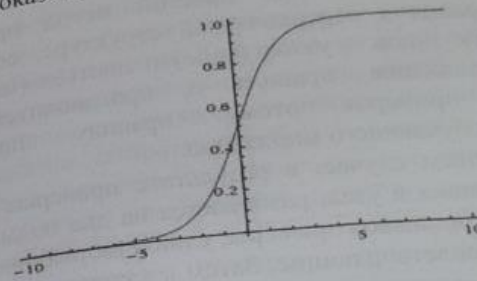


Рис. 6.12. График зависимости между вероятностью и переменной

Сигмоида, представленная на рис. 6.12, берет какое-то число от  $+\infty$  до  $-\infty$  и переводит его в число от 0 до 1.

Данная модель преобразования решает задачу классификации. В теории машинного обучения (Machine Learning – ML) если выход алгоритма преобразования представляет вещественную переменную, то задача называется задачей восстановления – регрессией. Если же выход принимает конечное число значений – это задача классификации. В задачах такого вида множество ответов является конечным, каждый ответ выражению  $y = f(WX)$  соответствует некоторому классу объектов, и задача ML заключается в вычислении по каждому объекту соответствующего ему класса.

### Алгоритм «дерево решений»

Деревья решений являются одним из наиболее эффективных инструментов интеллектуального анализа данных, которые позволяют решать задачи классификации и регрессии. Поскольку правила в деревьях решений получаются путём обобщения множества отдельных наблюдений (обучающих примеров),

описывающих предметную область, то по аналогии с соответствующим методом логического вывода их называют индуктивными правилами, а сам процесс обучения – индукцией деревьев решений. Метод относится к категории обучения с учителем [31].

Дерево решений – это простой метод представления решающих правил в иерархической структуре, состоящей из элементов двух типов – узлов (node) и листьев (leaf). В узлах находятся решающие правила и производится проверка соответствия примеров этому правилу по какому-либо атрибуту обучающего множества.

В простейшем случае, в результате проверки, множество примеров, попавших в узел, разбивается на два подмножества, в одно из которых попадают примеры, удовлетворяющие правилу, а в другое – не удовлетворяющие. Затем к каждому подмножеству вновь применяется правило и процедура рекурсивно повторяется пока не будет достигнуто некоторое условие остановки алгоритма. В результате в последнем узле проверка и разбиение не производится и он объявляется листом. Лист определяет решение для каждого попавшего в него примера. Для дерева классификации это класс, ассоциируемый с узлом, а для дерева регрессии – соответствующий листу модальный интервал целевой переменной. Таким образом, в отличие от узла, в листе содержится не правило, а подмножество объектов, удовлетворяющих всем правилам ветви, которая заканчивается данным листом.

Основная сфера применения деревьев решений – поддержка процессов принятия управленческих решений, используемая в статистике, анализе данных и машинном обучении. Задачами, решаемыми с помощью данного аппарата, являются:

- классификация – отнесение объектов к одному из заранее известных классов, при этом целевая переменная должна иметь дискретные значения;
- регрессия – предсказание числового значения независимой переменной для заданного входного вектора.
- описание объектов – набор правил в дереве решений позволяет компактно описывать объекты.

Процесс построения деревьев решений заключается в

Искусственный интеллект

последовательном, разбиении обучающего множества на подмножества с применением решающих правил в узлах. Процесс разбиения продолжается до тех пор, пока все узлы в конце всех ветвей не будут объявлены листьями. Объявление узла листом может произойти естественным образом (когда он будет содержать единственный объект, или объекты только одного класса), или по достижении некоторого допустимого числа задаваемого пользователем (минимально глубина дерева).

При формировании правила для разбиения в очередном узле дерева необходимо выбрать атрибут, по которому это будет сделано. Общее правило для разбиения в очередном узле следующего поколения: выбранный атрибут должен разбить подмножество наблюдений в узле так, чтобы результирующие классы, или были максимально приближены к этому, т.е. количество объектов из других классов («примесей») в каждом из этих множеств было как можно меньше.

Как было отмечено выше, если «рост» дерева не ограничить, то в результате будет построено сложное дерево с большим числом узлов и листьев. Как следствие оно будет трудно интерпретируемым. В то же время решающие правила в таких деревьях, создающие узлы, в которые попадают два-три примера, оказываются малозначимыми с практической точки зрения.

Гораздо предпочтительнее иметь дерево, состоящее из малого количества узлов, которым бы соответствовало большое число примеров из обучающей выборки. Поэтому представляет интерес подход, альтернативный ранней остановке – построить все возможные деревья и выбрать то из них, которое при разумной глубине обеспечивает приемлемый уровень ошибки распознавания, т.е. найти наиболее выгодный баланс между сложностью и точностью дерева.

Рассмотрим основные достоинства алгоритма:

- относительно быстрый процесс обучения;
- извлечение правил на естественном языке;
- высокая точность предсказания;
- построение непараметрических моделей.

В силу этих и многих других причин, деревья решений

являются важным инструментом в работе каждого специалиста, занимающегося анализом данных. Основные области применения: банковское дело (оценка кредитоспособности клиентов), оценка качества продукции и выявление дефектов, диагностика заболеваний.

Рассмотрим, как законы энтропии используются при построении дерева решений. Энтропия для системы с  $N$  возможными состояниями определяется следующим образом:

$$S = - \sum_{i=1}^N p_i \log_2 p_i$$

где  $p_i$  – вероятности нахождения системы в  $i$ -ом состоянии. Это очень важное понятие. Опуская предпосылки введения этого понятия, отметим, что, интуитивно, энтропия соответствует степени хаоса в системе. Чем выше энтропия, тем менее упорядочена система и наоборот.

Для иллюстрации того, как энтропия поможет определить хорошие признаки для построения дерева, приведем пример с черными и белыми шариками. Будем предсказывать цвет шарика по его координате, это позволит показать, как энтропия используется для построения дерева решений (рис.6.13).

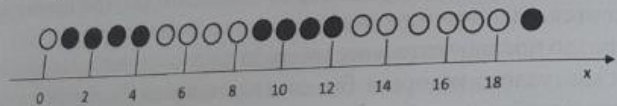


Рис.6.13. Произвольное расположение черных и белых шариков

На рисунке 9 черных шариков и 11 белых шариков. Если наудачу вытащить шарик, то он с вероятностью  $p_1=9/20$  будет черным и с вероятностью  $p_2=11/20$  будет белым. Значит, энтропия состояния:

$$S = - 9/20 \log_2 9/20 - 11/20 \log_2 11/20 \approx 1.$$

Само значение «1» ни о чем не говорит. Посмотрим, как

изменится энтропия если разбить шарики на две группы – с координатой равной или меньше 12 (рис.6.14).

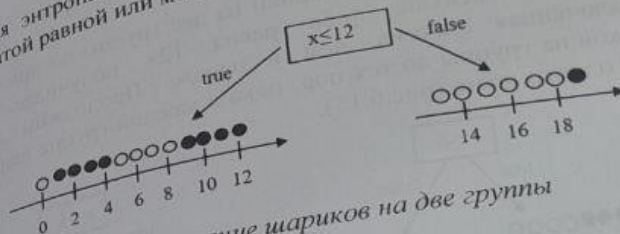


Рис.6.14. Разделение шариков на две группы

В левой группе оказалась 13 шаров, из которых 8 черных и 5 белых. Энтропия этой группы равна:

$$S_1 = - 5/13 \log_2 5/13 - 8/13 \log_2 8/13 \approx 0,96.$$

В правой группе оказалось 7 шаров, из которых 1 черный и 6 белых. Энтропия правой группы равна:

$$S_2 = - 1/7 \log_2 1/7 - 6/7 \log_2 6/7 \approx 0,6.$$

Как видно из формул, энтропия уменьшилась в обеих группах, особенно во второй. Поскольку энтропия – это степень хаоса (неопределенности) в системе, уменьшение энтропии называют приростом информации. Формально прирост информации (information gain) при разбиении выборки по признаку  $Q$  (в нашем примере это признак «  $x \leq 12$  ») определяется как:

$$IG(Q) = S_0 - \sum_{i=1}^q N_i/N * S_i$$

где  $q$  – число групп разбиения,  $N_i$  – число элементов выборки, у которых признак  $Q$  имеет  $i$ -ое значение.

В данном случае после разделения получилось две группы ( $q=2$ ) – одна из 13 элементов ( $N_1=13$ ), вторая группа – из 7 элементов ( $N_2=7$ ). Прирост информации получился:



### Наземный сегмент и организация связи

Система спутниковой связи Odyssey предназначена для организации радиотелефонной связи, передачи данных и коротких сообщений о местоположении подвижных объектов. Наземный сегмент Odyssey включает в себя узловые (базовые) станции и терминалы. Двухрежимный радиотелефонный терминал обеспечивает работу в сетях стандарта GSM, TDMA, CDMA, PHS. Он позволяет работать не только в системе Odyssey, но и в наземных сотовых сетях, причем доступ к наземной сотовой сети является приоритетным.

Связь регламентирована так, что после определения свободных частот вызов всегда направляется в адрес базовой станции сотовой сети. В случае невозможности соединения с базовой станцией (вызов заблокирован или все частоты заняты) терминал автоматически передает запрос на спутник системы Odyssey.

Передача речи осуществляется со скоростью 4,2 кбит/с; вероятность ошибки в речевом канале — не более 10<sup>-3</sup>. Кроме речевой связи терминал Odyssey предоставляет возможность приема сообщений персонального радиовызова (пейджинг) с буквенно-цифровой индикацией, обеспечивает режим электронной почты, а также определение местоположения абонента. Скорость передачи данных составляет 2,4—64 кбит/с; вероятность ошибки на бит — не более 10<sup>-5</sup>. Для коррекции ошибок применяется сверточное кодирование ( $R = 1/2$ ,  $K = 7$ ).

Определение координат производится по собственным сигналам системы Odyssey. В связи с относительно большим (для средневысотной орбитальной группировки) числом спутников в любой точке обслуживаемой территории можно наблюдать «созвездие» из двух или трех спутников, находящихся под большими углами видимости. Это делает возможным установление местоположения объекта только по сигналам КА Odyssey. Погрешность определения местоположения — не более 15 км.

В системе не предусмотрены межспутниковые связи. Весь график данного региона передается через узловые станции (таб. 5.3), связанные между собой многоканальными линиями связи.

В задачи узловой связи входят не только прием/передача персонального графика, но и обеспечение сопряжения с телефонной сетью общего пользования, управление межлучевой коммутацией, прием и обработка телеметрии с борта спутника. Также, в системе Odyssey при подключении мобильных пользователей к телефонной сети общего пользования задержка сигнала, которая складывается из задержки спутникового канала (84 мс) и задержки речевого сообщения в каждом из обслуживаемых регионов Земли достаточно 7 станций. На каждой из них предусмотрена передача по четыре следящие параболические антенны диаметром около 7 м, три из которых будут использоваться для одновременной работы со спутниками, а четвертая — для передачи трафика от спутника к спутнику через станцию с учетом радиовидимости. Кроме того, эта антенна необходима для повышения надежности связи в случае неблагоприятных климатических условий.

Основные характеристики узловых станций

Таблица. 5.3.

Показатель	Направление связи	
	Прием	Передача
Диапазон частот, ГГц	29,1-29,4	19,3-19,6
Общая ширина полосы, МГц	300	300
Ширина полосы канала, МГц	2,5	2,5
Вид поляризации	RHCP	LHCP
Коэффициент усиления антенны, ДБ <sub>n</sub>	64,8	60,8
Ширина луча по уровню 3 дБ, °	2,2	0,17
Шумовая температура приемника, °К	666,5	-
Эквивалентная изотропная мощность излучения, дБВт	-	85,9

Бортовые антенны спутника имеют узкую диаграмму направленности, а приемные устройства спутников — высокую чувствительность, поэтому в абонентских станциях можно применять передатчики с малой выходной мощностью. Планируется выпустить две модификации абонентских терминалов, различающиеся выходной мощностью передатчика (0,5 и 5 Вт). В конструкции терминала предполагается использовать антенну типа «четырёхзаходная спираль» с коэффициентом усиления 2,5 дБ. Энергетический запас на линии связи составит 6—10 дБ.

#### Услуги системы Odyssey

Развертывание орбитальной группировки сети Odyssey производилось в 2 этапа. На первом этапе, услуги предоставляются только 6 КА. Они обеспечивают непрерывное обслуживание в основных регионах в течение 14 часов в сутки. На следующем этапе развернулась полномасштабная орбитальная группировка из 12 спутников. Были определены приоритетные зоны обслуживания: территория континентальной части США с прибрежными районами, Европа, Азия и акватория Тихого океана.

Пользователями системы будут частные лица и государственные структуры, нуждающиеся в непрерывной мобильной связи на значительных по площади территориях, а также население регионов с низким уровнем наземной инфраструктуры связи. В регионах, где отсутствуют альтернативные виды связи, использование каналов спутниковой связи позволяет расширить зоны действия сотовых сетей. Абонентам таких сетей предоставлена возможность глобального роуминга. Служба коротких сообщений предлагает услуги, аналогичные пейджинговому. Дополнительно предоставляются следующие услуги: определение местоположения клиента, голосовая почта, аварийные сообщения, перевод с одного языка на другой.

В 2005 г., после окончания развертывания системы ODYSSEY, число ее абонентов превысило 2 млн. На данный момент число пользователей составляет около 9 млн. человек. Цена одного абонентского терминала, составляет 350—1000 долларов,

размер ежемесячной абонентской платы — 25 долларов, а стоимость минуты телефонной связи в спутниковом канале — 0,75 долларов.

#### 5.8. Международная система ICO

##### Частотное обеспечение

Система ICO использует для связи L- и S-диапазоны частот, поддерживая цифровую обработку сигнала на борту спутника. В качестве базовой технологии определен метод многостанционного доступа с временным разделением каналов (TDMA).

При определении оптимальных полос частот для абонентских линий связи были рассмотрены несколько вариантов. Принимались во внимание следующие соображения. Диапазон 1,5/1,6 ГГц широко используемый для подвижных спутниковых служб (ПСС), очевидно, окажется чрезмерно перегруженным, что сильно ограничит потенциал служб ICO. Диапазон 1,6/2,4 ГГц выделенный службе ПСС на Всемирной административной конференции по радиосвязи (WARC-92), чреват серьезными проблемами координации с другими службами, которые применяют этот диапазон, например, для фиксированной наземной связи; кроме того, США намерены использовать его для национальных систем.

Наконец, были выбраны следующие диапазоны: «терминал-спутник» — диапазон 1980—2010 МГц, «спутник-терминал» — 2170-2200 МГц.

Для организации связи между КА и узловыми станциями предназначены фидерные линии. Для их работы Всемирная конференция по радиосвязи WRC-95 рекомендовала диапазон 5/7 ГГц («узловая станция—спутник» диапазон 5150-5250 МГц, «спутник—узловая станция» - 6975-7075 МГц).

##### Космический сегмент

Система ICO состоит из космического, наземного и пользовательского сегментов. Космический сегмент включает в себя 12 КА (10 рабочих и 2 резервных), запущенных на круговую орбиту высотой 10 355 км над поверхностью Земли. Стартовая масса спутника — 2750 кг, расчетный период эксплуатации — 12

лет. Спутники размещены в двух ортогональных плоскостях, по 6 КА в каждой. Угол наклона орбиты к плоскости экватора составит 45°.

Такая орбитальная группировка обеспечивает глобальный охват поверхности Земли, в том числе полярных районов. Вследствие перекрытия зон охвата в пределах видимости каждой точки зоны обслуживания одновременно находятся два—четыре КА. Один спутник обслуживать приблизительно 25% поверхности Земли (рис. 5.8). Первый спутник системы ICO был запущен в 1998 г.; ввод системы в эксплуатацию произведен в 2000 г.

Продолжительность обслуживания абонентов определяется следующими величинами:

- временем пролета одного спутника над зоной обслуживания;
- средним временем, затрачиваемым на переключение абонента с уходящего за горизонт КА на восходящий КА;
- продолжительностью установления соединения, определяемого схемой организации связи. Средняя продолжительность обслуживания абонентов составит 50 мин; максимальное время пребывания одного КА в зоне радиовидимости может достигать 1,5-2 ч.



Рис. 5.8. Диаграмма мгновенной зоны покрытия поверхности Земли системой ICO при использовании 10 КА.

В системе ICO применены, главным образом, уже известные и проверенные технические решения. Для изготовления спутников используется спутниковая платформа HS-601 корпорации Hughes Space and Communications (США), применяющаяся для создания крупногабаритных спутников на геостационарной орбите. В конструкцию внесены изменения, в частности переработанная программа ориентации бортовых антенн и панелей солнечных батарей, установлена упрощенная двигательная установка.

Чтобы исключить взаимовлияние системой ICO при использовании 10 КА трактов приема и передачи, на КА применяются отдельные антенны для каждого диапазона частот. Антенна L-диапазона имеет диаметр 2 м. Использование многолучевой диаграммообразующей схемы обеспечивает многократное назначение частот. Согласно проекту, в системе ICO для приема/передачи служат 163 отдельных луча (запас по энергетике составит 8—10 дБ); зона обслуживания одного КА — примерно 7 тыс. км (рис. 5.9). Спутники с установленными на них ретрансляторами С- и S-диапазонов одновременно поддерживают 4500 телефонных каналов.

В системе ICO не предусмотрена бортовая обработка сигнала в полном объеме. Однако управление назначением частот и маршрутизация сигнала осуществляются с помощью бортового процессора.

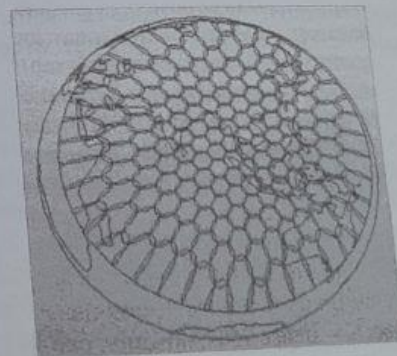


Рис. 5.9. Зона обслуживания одного КА (163 луча) системы ICO.

Применение арсенид-галлиевых батарей обеспечивает в конце эксплуатации потребляемую мощность 8700 Вт. В предварительном списке ракетоносителей, которые произвели запуск спутников системы ICO, числятся Atlas IIА, Delta III, «Протон» и «Зенит» (для запуска с морских площадок).

#### Наземный сегмент и организация связи

В состав наземного сегмента входят центр управления спутниковой группировкой SCC (Satellite Control Centre), центр управления наземной сетью (Network Management Centre) и наземная сеть ICONET (ICO network), (рис. 5.10).

NMS, центр управления наземной сетью ICONET, размещен в Японии, а центр SCC — в Лондоне. В функции последнего входят поддержание орбитальной группировки в работоспособном состоянии, сбор телеметрических данных об отдельных состояниях КА, контроль рабочих параметров и др. Службы SCC несут ответственность за запуск КА, управление и перераспределение частот между лучами КА.

Спутниковые каналы подключаются к существующим сетям связи через собственную сеть ICONET, которая на первом этапе внедрения состоит из 12 наземных станций — так называемых спутниковых узлов доступа SAN (Satellite Access Node). Узлы SAN служат «шлюзами» между спутниками ICO и абонентами наземных сетей общего пользования. Магистральные каналы с высокой пропускной способностью связывают узлы между собой.

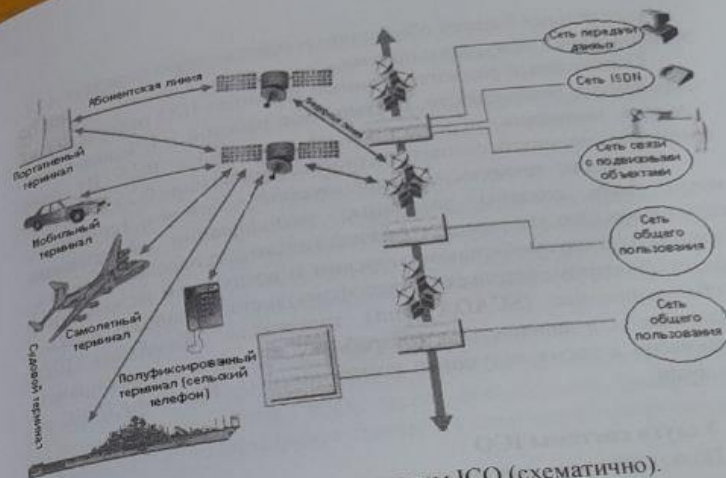


Рис. 5.10. Структура системы ICO (схематично).

Связь между абонентами (как и в существующей системе Inmarsat) организуется только через узлы SAN; непосредственная связь абонентов не поддерживается. Радиотелефонный терминал ICO работает в двух режимах — через КА системы ICO или наземные базовые станции сотовой связи — и совместим с ее основными стандартами. Для связи с подвижными объектами применяются специальные терминалы.

#### Терминалы пользователя

В спутниковой сети ICO в качестве базового используется портативный двухрежимный терминал, совмещенный с сотовым телефоном стандарта GSM (или CDMA, D-AMPS, PDC). Предполагается разработка одорежимного радиотелефонного терминала, работающего только через КА системы ICO.

Основные характеристики базового терминала:

- масса — менее 750 г,
- объем — около 500 см,
- стоимость — 750—1500 долларов,



- отдельная батарея обеспечивает одночасовую передачу и 24-часовой режим дежурного приема.

Портативный радиотелефонный терминал ICO отвечает всем требованиям безопасности, связанным с работой в ВЧ-диапазоне. Средняя мощность передатчика не превышает 0,25 Вт (для сравнения: мощность сотовых радиотелефонов равна 0,25—0,6 Вт).

На основе технологии, используемой в базовом терминале, могут быть созданы различные модификации абонентских терминалов. Это, например, терминал только для передачи данных, терминалы в автомобильном, морском и воздушном исполнении, полустационарные («сельский таксофон») и стационарные, а также необслуживаемые (SCADA unit) терминалы. Компания ICO заключила соглашение на разработку 3 млн. портативных терминалов с тремя ведущими компаниями — Panasonic, NEC и Mitsubishi.

#### Услуги системы ICO

Пользователям предоставлены следующие виды услуг: двусторонняя речевая связь, передача факсимильных сообщений группы 3, передача данных со скоростью 2,4 кбит/с. Качество речевой связи соответствует стандарту GSM для сотовых сетей. Предусмотрена пейджинговая связь с глубоким проникновением (т. е. с большим запасом по энергетике канала), а также дополнительные услуги — речевой вызов, связь с оплатой по кредитной карточке, отображение номера вызывающего абонента на встроенном в терминал индикаторе, определение местоположения абонента. При отсутствии КА в пределах прямой видимости имеется оповещение абонентов о вызове, о наличии сообщения электронной почты и отображение на дисплее номера вызывающего абонента.

Разработчики видят пять ключевых областей применения системы ICO:

- расширение спектра услуг для абонентов спутниковой связи в районах, уже охваченных сотовыми сетями;
- подвижная связь общего пользования через портативные радиотелефонные терминалы в районах, не охваченных сотовой связью или использующих несовместимые стандарты;

- специализированная подвижная связь для грузовых перевозок, а также обеспечение автомобильной, морской и воздушной связи;

- полужесткая связь для корпоративных пользователей (нефте- и газодобывающей промышленности, малого бизнеса (склады, большие магазины и др.);

- связь для государственных структур.

Пропускная способность системы составляет 1 млн. абонентов при средней продолжительности разговоров 60 мин/мес. Для сравнения: по прогнозам специалистов, в системе Iridium при тех же условиях число пользователей равно 600—800 тыс., а в Globalstar — 1 млн.

Разработка и изготовление 12 КА оцениваются в 1,3 млрд. долларов, а их запуск обойдется в 900 млн. долларов. Согласно расчетам специалистов ICO, цена абонентской аппаратуры составит 750-1500 долларов, а стоимость минуты разговора около 2 долларов.

#### 5.9. Сравнение систем Odyssey и ICO

В число наиболее крупных проектов создания систем глобальной персональной радиотелефонной связи входят (кроме рассмотренных выше систем Odyssey и ICO) низкоорбитальные системы Iridium и Globalstar, приведено в таблице 5.4. Предоставляя пользователям практически тот же набор телекоммуникационных услуг (речь, данные, пейджинг, короткие сообщения, определение местоположения), конкурирующие системы существенно различаются по своим характеристикам и наземным структурам. Так, для обеспечения глобальной связи в системах Odyssey/ICO требуются всего 7—12 узловых станций, а для обслуживания пользователей Globalstar — в 20 раз больше. Структура наземного сегмента сети Iridium несколько проще, чем в Globalstar (благодаря использованию межспутниковых линий связи).

Система ICO — единственная из четырех конкурирующих систем — пока не имеет лицензии США на коммерческое использование радиочастот. Однако организация прилагает все усилия для решения этой проблемы.

Сравнительная характеристика глобальных систем радиотелефонной связи. Таблица 5.4.

Показатель	Odyssey	ICO	Iridium	Globalstar
Тип орбиты	MEO	MEO	LEO	LEO
Число КА	12	12	66	48
Высота орбиты, км	10 354	10 355	780	1400
Наклонение орбиты, °	50	45	86	52
Масса КА, кг	2500	2750	690	450
Потребляемая мощность, Вт	4600	8700	1000	1200
Число лучей	51	163	48	16
Срок эксплуатации КА, лет	15	12	5	7,5
Метод многостанционного доступа	CDMA	TDMA	TDMA	CDMA/FDMA
Число узловых станций	7	12	25	150-210
Число каналов КА, эквивалентных 4,8 кбит/с	3000	4500	От 600	1300
Стоимость проекта млрд. долларов	2,5	2,8	От 3,5	2,0
Стоимость двухрежимного терминала, дол.	350	750	3000	750
Тариф, дол./мин	0,75	2	3	0,35-3

Наиболее важными для пользователя являются технико-экономические параметры, но эта информация нередко носит рекламный характер (т. е. является не вполне объективной), что объясняется жесткой конкурентной борьбой на рынке. Особо бурные споры вызывают цена терминалов и предлагаемые тарифы. Так, трудно объяснить, почему двухрежимный терминал Motorola, обеспечивающий практически те же характеристики, что и терминалы других фирм (например, терминалы Mitsubishi для

систем Odyssey и ICO), стоит в несколько раз дороже, чем они. Какими окажутся окончательные цены и тарифы, покажет время.

### 5.10. Расчет зависимостей характеристик спутников от параметров долготы и угла раскрыва антенны

В параграфе приводится расчет коэффициента усиления антенны бортового ретранслятора в зависимости от угла раскрыва, наклонной дальности, а также ширины диаграммы направленности по уровню половинной мощности.

#### Цель работы:

- Определить коэффициент усиления антенны бортового ретранслятора в зависимости от угла раскрыва.
- Определить наклонную дальность между ЗС и КС.
- Рассчитать ширину диаграммы направленности по уровню половинной мощности.

#### Методические указания к расчету:

1. По заданным величинам из таблицы 5.5 выписать исходные данные для расчета.

2. Определите коэффициент усиления антенны бортового ретранслятора по формуле:

$$G_6 = 44.4 - 10 \lg \Phi_0 - 10 \lg \Phi_1, \text{ дБ} \quad (1)$$

где  $\Phi_0, \Phi_1$  – углы раскрыва антенны ИСЗ (для симметрично раскрывающихся антенн  $\Phi_0 = \Phi_1$  (задается по варианту);

После расчета необходимо перевести значение  $G_6$  (дБ) в  $G_6$  (раз) по формуле:

$$G_6 (\text{раз}) = 10^{G_6 (\text{дБ})/10}, \text{ раз} \quad (2)$$

3. Наклонная дальность между ЗС и КС определяется по формуле:

$$d = 42644 \cdot \sqrt{1 - 0,295 \cdot \cos\psi}, \text{ км} \quad (3)$$

где  $\cos\psi = \cos\xi \cdot \cos\Delta\lambda$  (4)

$\xi$  – широта приемной станции (задается по варианту);  
 $\Delta\lambda$  – разность по долготе между ЗС и КС (задается по варианту).

4. Рассчитайте ширину диаграммы направленности по уровню половинной мощности.

$$\theta_{0,5} = \sqrt{4,9 \cdot 10^3 \cdot g/G_6} \quad (5)$$

где  $g = 0,5$  – коэффициент использования поверхности антенны

**Пример расчета:**

1. Исходные данные:

Углы раскрытия антенны ИСЗ  $\Phi_0 = \Phi_1 = 5^\circ$ .

Широта приемной станции  $\xi = 42^\circ$ .

Разность по долготе между ЗС и КС  $\Delta\lambda = 7^\circ$ .

2. Определение коэффициента усиления антенны бортового ретранслятора по формуле:

$$G_0 = 44,4 - 10 \lg 5 - 10 \lg 5 = 44,4 - 6,9 - 6,9 = 30,6 \text{ дБ}$$

Переведем значение  $G_0$  (дБ) в  $G_0$  (раз) по формуле:

$$G_0 (\text{раз}) = 10^{G_0 (\text{дБ})/10} = 10^{30,6/10} = 10^{3,06} = 1148 \text{ раз}$$

3. Для определения наклонной дальности между ЗС и КС сначала определим:

$$\cos\psi = \cos\xi \cdot \cos\Delta\lambda = \cos 42^\circ \cdot \cos 7^\circ = 0,743 \cdot 0,993 = 0,738$$

Тогда:

$$d = 42644 \cdot \sqrt{1 - 0,295 \cdot \cos\psi} = 42644 \cdot$$

$$\sqrt{1 - 0,295 \cdot 0,738} = 37720 \text{ км}$$

4. Расчет ширины диаграммы направленности по половинной мощности.  
 $L_{\text{полн}} = L_{\text{атен}} + L_0 + L_{\text{н}} + L_{\text{н}} = 0,15 + 0 + 0,3 + 0,05 = 0,5 \text{ дБ}$   
 Ослабление сигнала в свободном пространстве:  
 $\theta_{0,5} = \sqrt{4,9 \cdot 10^3 \cdot g/G_6} = \sqrt{4,9 \cdot 10^3 \cdot 0,5/1148} = 1,6^\circ$

Таблица 5.5.

Исходные данные по вариантам

Последняя цифра номера по журналу	$\Phi_0 = \Phi_1$	$\xi$	$\Delta\lambda$
1	17°	30°	3°
2	11°	32°	4°
3	5°	34°	5°
4	2,5°	36°	6°
5	17°	38°	7°
6	11°	40°	3°
7	5°	42°	4°
8	2,5°	44°	5°
9	11°	46°	6°
0	5°	48°	7°

**Контрольные вопросы.**

1. Из каких базисных частей состоит система спутниковой связи?
2. Перечислите главные компоненты спутника и поясните их функциональные задачи.
3. Какие типа ЗС существуют? Дайте их краткую характеристику.
4. Что представляет собой система Aloha?
5. Приведите и поясните схему организации связи в системе Odyssey.

## ГЛАВА 6. ИЗУЧЕНИЕ ПРИНЦИПОВ ПОСТРОЕНИЯ ЗЕМНЫХ СТАНЦИЙ VSAT

В настоящее время большое развитие получили земные станции спутниковой связи типа VSAT (Very Small Aperture Terminal), которые устанавливаются непосредственно у пользователей и не требуют для обслуживания постоянного высококвалифицированного персонала. Эти станции применяются обычно в выделенных сетях, частных и деловых, для передачи данных и телефонии в режиме только на прием или на прием и передачу. Они имеют антенны диаметром 1,8...3,5 м, скорости передачи в цифровом виде до 2 Мбит/с, мощность передатчика в несколько ватт и работают чаще всего в диапазонах частот FCC 6/4 и 14/12 ГГц.

В данное время в мире установлено более 900 тыс. терминальных станций VSAT, а темпы роста мирового рынка этого вида связи составляют в среднем около 15 % в год. Соответственно, ожидается, что количество VSAT терминалов в мире к 2015 г. достигнет 3 млн.

### 6.1. Отличие VSAT-сетей от локальных или наземных региональных компьютерных сетей

Поскольку рассматриваемые сети включают в свой состав космический сегмент линии связи, они существенно отличаются от наземных региональных компьютерных сетей, хотя во многом их назначение аналогично.

В то же время VSAT-сети характеризуются особыми свойствами:

1) большой задержкой времени распространения сигнала при передаче данных. Так эта задержка при однокачковой распространении сигнала в обратном канале (in route) от VSAT-станции на ретранслятор и от ретранслятора к другой VSAT-станции составляет 330 мс. Только одно это обстоятельство нарушает работоспособность любой наземной системы управления сетью, а в зависимости от метода доступа к ресурсам ретранслятора указанная временная задержка может удвоиться или утроиться;

2) различными методами доступа к ресурсам ретранслятора;

3) асимметричным обменом данными, когда прямой канал (route) от центральной станции к VSAT-станции использует совместно с большим числом других VSAT-станций этой сети широкую полосу частот. Обратный канал имеет меньшую полосу частот и совместно используется меньшим числом VSAT-станций; 4) недискретными отказами сети, когда сеть работает ненадежно. Сигналы в VSAT-сетях часто испытывают изменяющиеся во времени ослабления в атмосфере, например, при сильном дожде. Это приводит к понижению энергетических характеристик радиоканала, но не к полной потере связи, когда сеть продолжает функционировать с потерями до 70 % пакетов данных для пользователей;

5) пространственно-удаленным размещением VSAT-станций, поскольку часто они используются там, где локальная инфраструктура практически отсутствует. Это приводит к тому, что некоторые VSAT-станции получают электропитание от солнечных батарей. Удаленное размещение приводит к тому, что среднее время наработки на отказ (Mean Time Between Failure, MTBF) оборудования VSAT-станций должно быть достаточно большим и оно должно быть надежно защищено от внешних метеорологических воздействий, все устройства должны иметь дистанционное управление. Оборудование VSAT-станций должно быть устойчивым к отказам, в том числе к перебоям электропитания и преодолевать эти отказы автоматически. Механические устройства должны выполняться с высокой точностью, поскольку отклонение оси зеркала антенны на несколько десятых долей градуса от заданного направления может привести к снижению принимаемой мощности наполовину;

6) передачей сигналов управления сетью в том же радиоканале, в котором работает сеть. Почти всегда в VSAT-станциях не используются сигналы управления сетью с передачей по отдельному каналу (т. е. второе соединение для непрерывного контроля первого), которое может иметь собственные проблемы, например, необходимость обмена данными; 7) использованием специальных методов и сеансовым режимом работы сети. Большие временные задержки при использовании VSAT-технологии не позволяют использовать отдельные сеансовые протоколы. Полезное использование полосы

частот должно быть повышено до 90 % путем применения специальных методов, таких, как подавление сигналов подтверждения. Однако это может привести к проблемам в системах управления наземными сетями:

8) отсутствием стандартов, поскольку при использовании VSAT-технологии связи не существует промышленных стандартов, и каждый провайдер имеет свои собственные протокол и оборудование. Это оборудование сложно интегрировать с оборудованием других производителей, в том числе с системами управления сетью.

Эти отличия VSAT-сетей от локальных или наземных региональных компьютерных сетей относятся только к сравнительно простым VSAT-сетям. На самом деле большая часть сетей являются гибридными, с широким набором технологий и протоколов, таких, как ATM, Frame Relay, ISDN, VPN и Ethernet, и с использованием стандартного сетевого оборудования (маршрутизаторов, коммутаторов, центральных станций, брандмауэров и серверов). Система управления такой гибридной сетью способна работать со всем этим оборудованием, включая специфические условия VSAT-сетей.

### Определение класса земных станций VSAT

К классу земных станций VSAT относятся станции спутниковой связи, технические характеристики которых удовлетворяют следующим требованиям Рек. МСЭ-Р S.725 «Технические характеристики VSAT»:

- станции VSAT устанавливаются непосредственно у пользователей, причем плотность размещения их на ограниченной территории может быть весьма высокой;
- контроль и управление работой станций VSAT в сети осуществляются централизованно, но могут дополнительно использоваться и местные станционные системы контроля и управления;
- станции VSAT относятся к Фиксированной спутниковой службе (ФСС) и должны удовлетворять требованиям Регламента

радиосвязи (PP) и Рекомендациям МСЭ-Р, как и все земные станции ФСС;

- станции VSAT обычно применяются в так называемых выделенных сетях (частных, деловых) для передачи данных и телефонии (симплексе) или на прием/передачу (дуплексе);
- антенны VSAT обычно имеют диаметр 1,8...3,5 м, но в отдельных системах могут использоваться маломощный (диаметром до 6 м);
- в станциях VSAT используется маломощный радиопередатчик (обычно от 1 до 20 Вт) с обязательным ограничением излучаемой мощности в целях безопасности.

### 6.2. Типы сетей VSAT

По набору предоставляемых услуг все VSAT-сети условно разделяются на полнофункциональные с полным набором услуг, интерактивные, или сети обмена данными с многими услугами, и специализированные с несколькими отдельными услугами. Следует заметить, что полного набора услуг реально не представляет ни одна из сетей, так как набор таких услуг постоянно увеличивается.

Часто сети обмена данными называют интерактивными, и для этого имеются основания, поскольку реализуется взаимодействие пользователей. Считается, что полнофункциональные сети в отличие от специализированных сетей не ориентированы на предоставление только некоторых услуг, например, услуг сельской телефонии, технологической связи и т. п. В соответствии с этим условным делением самыми распространенными VSAT-сетями являются интерактивные сети (примерно 86 % всех VSAT-сетей, 10 % сетей — полнофункциональные и 4 % — специализированные). Однако полнофункциональные сети часто противопоставляют интерактивным, что неверно, так как полнофункциональные сети априорно интерактивные.

С технической точки зрения в основу классификации VSAT-сетей целесообразно положить характер обмена данными:

двухсторонний и односторонний (вещание), а также топологию сетей.

Спутниковые ретрансляторы VSAT-сетей обычно поддерживают одностороннюю передачу данных в виде вещания и приложения к нему в виде двухсторонней передачи данных (их часто называют интерактивными). Среди последних видов обмена выделяют интерактивные VSAT-сети с несимметричным обменом данными.

VSAT-сети с односторонней передачей данных не пользуются ширококвещательной спутниковой службой BSS.

Внедрение цифровых технологий предоставило провайдерам и пользователям большую гибкость в работе. Так, станция управления в составе потока ширококвещания посылает и сигналы, кодированные специальным образом. Патентованное программное обеспечение, предоставляемое провайдерами для терминалов группы пользователей, обеспечивает доступ (за соответствующую плату) к этой части ширококвещательного обмена данными. Эта форма выбора каналов получила наименование узкоквещания (narrowcasting). В зоне обслуживания ширококвещания может находиться множество групп пользователей, получающих за дополнительную плату услуги узкоквещания.

Интерактивные VSAT-сети с несимметричным обменом данными (Split-Two-Way или Split-IP) используются при недоступности высококачественного обратного канала, такого, как в системах ширококвещательной спутниковой службы Ku-диапазона частот, организующего обмен данными в сети Интернет. Относительно высокая скорость передачи данных приходящего потока не дополняется высокой скоростью передачи данных от VSAT-станции. Если прямой канал «борт — Земля» ширококвещательной спутниковой службы используется провайдером услуг Интернета только в качестве канала доставки данных до пользователя, то единственной возможностью организовать обратный канал для пользователя является телефонная сеть общего пользования. Следовательно, протокол сети Интернет разделяется на две части: между прямым спутниковым каналом «борт — Земля» и наземным телефонным каналом «Земля — борт». Преимущество такого подхода состоит в том, что VSAT-станция не используется в режиме передачи

данных, что существенно снижает ее стоимость, упрощает ее функциональную структуру, исключает возможность создания помех другим радиосредствам. Недостаток VSAT-сетей с несимметричным обменом данными состоит в том, что обычно доступ в телефонную сеть общего пользования реализуется с помощью модема, скорость передачи которого не превышает 56 кбит/с.

В случае интерактивных VSAT-сетей обратный канал вводится в состав сети таким образом, чтобы он устанавливался через ретранслятор того же спутника, что и прямой канал. Объединенная VSAT/WLL-сеть, рассмотренная выше, является интерактивной, позволяющей устанавливать двухсторонние соединения между центральной станцией (шлюзом) и любой VSAT-станцией. Следовательно, интерактивные VSAT-сети являются общим классом сетей этого типа, поэтому в дальнейшем будем рассматривать этот класс VSAT-сетей, подклассами которого являются VSAT-сети вещания с односторонней передачей данных и интерактивные VSAT-сети с несимметричным обменом данными.

### 6.3. Технологии, используемые в сетях VSAT для создания корпоративных сетей

Различные технологии построения сетей VSAT определяющие топологию, организацию множественного доступа, способ предоставления каналов, имеют свои особенности, влияющие на потребительские характеристики сетевого оборудования. Они определяют пропускную способность, задержки сигнала во времени, сложность технического обслуживания, стоимость проектирования. Эти особенности имеют также важное значение при принятии решений по развертыванию и внедрению спутниковых сетей связи.

Представляют интерес сетевые технологии SCPC (Single Channel Per Carrier) и MCPC (Multi Channel Per Carrier). Эти технологии активно применяются для построения небольших сетей с интенсивным трафиком. Каждая земная станция (ЗС), реализующая технологию SCPC или MCPC, имеет выделенный для нее постоянный сегмент емкости ретранслятора, через который

поддерживается постоянное соединение. Достоинство данных технологий в том, что они гарантируют необходимую пропускную способность канала спутниковой связи, а недостаток — в отсутствии возможности динамического перераспределения ресурса ретранслятора между узлами связи, в том числе тогда, когда необходимо передавать трафик со скоростью большей, чем скорость передачи несущей станции сети. Каналы SCPC просты в реализации, однако эффективность использования дорогостоящего космического сегмента в корпоративной сети на их основе ниже, чем в любой другой системе спутниковой связи с многостанционным доступом к одному и тому же частотному ресурсу (TDM, TDMA, DAMA и др.).

Так, многостанционный доступ (Demand Assigned Multiple Access, DAMA) — способ предоставления ресурса спутникового ретранслятора по требованию, когда канал выделяется пользователю только на время проведения сеанса связи, что обеспечивает экономное использование ресурса спутниковой емкости. В некоторых реализациях технологии DAMA предусмотрена возможность установления SCPC-соединений в зависимости от потребностей пользователей с разной пропускной способностью. Оборудование DAMA позволяет поддерживать полностью связную сетевую топологию.

TDM/TDMA (Time Division Multiplexing/Time Division Multiple Access) — комбинированная технология сетей с топологией типа «звезда». В сети с применением технологий TDM/TDMA центральная ЦС (ЦЗС) связывается со станциями пользователей с помощью одного или нескольких закрепленных каналов TDM с временным мультиплексированием. Передача информации в обратном направлении осуществляется по каналам TDMA с разделением по времени.

FTDMA (Frequency Time Division Multiple Access) — технология для сетей с разными топологиями (полносвязная — «каждый с каждым» или «звезда»), которая выбирается в зависимости от вида основного трафика (телефония или передача данных). В сети FTDMA ЦЗС организует связь для удаленных станций, предоставляя им свободные временные слоты, организованные на нескольких несущих полосах частот.

MF-TDMA (Multi Frequency-Time Division Multiple Access). Эта технология предоставляет множеству станций динамический доступ к общим частотным каналам с временным разделением. При этом может использоваться совокупность каналов с разной пропускной способностью, т. е. станция перестраивается не только по частоте, но и по скорости информационного потока. Технология MF-TDMA обладает двумя важными особенностями. Первая — это возможность динамического переназначения всего спутникового ресурса определенному соединению или даже направлению передачи трафика. Вторая — схожесть конфигураций и рабочих характеристик (диаметр антенны и мощность передатчика) ЦЗС и периферийных терминалов.

От терминалов ЦЗС отличается наличием системы управления сетью, осуществляющей мониторинг работы и имеющей средства изменения ее конфигурации. При необходимости эту систему можно развернуть на любом из терминалов. Таким образом, земной сегмент сети имеет высокую степень отказоустойчивости, однако ценой этому является более высокая стоимость оборудования терминалов.

Выбор той или иной технологии зависит от типа и назначения создаваемой корпоративной сети. Поскольку все многообразие потенциальных приложений систем спутниковой связи довольно полно отражается в режимах работы их периферийных или абонентских терминалов, корпоративные сети целесообразно классифицировать по уровню средней рабочей нагрузки терминала, поддерживающего передачу данных, телефонию и другие приложения, например, видеоконференцсвязь, в дневное время и в рабочие дни. При асимметричном трафике в расчет принимается только средняя скорость передачи данных по обратному каналу.

К основной категории относятся сети с низкой нагрузкой, в которых терминалу достаточно иметь пропускную способность не более 32 кбит/с, причем организовывать телефонную связь или не требуется, или же пользователям терминала необходима только одна телефонная линия. Основные функции таких сетей — сбор телеметрии, интернетизация школ, обеспечение работы сети банкоматов и др. Как правило, в таких сетях насчитывается от 50 до 10 тыс. терминалов, центральный узел сети в основном занимается обработкой информации, поступающей от терминалов. Для сетей с

низкой нагрузкой типичными являются протоколы передачи данных X.25 и IP.

### Топология сетей связи

Сети типа «звезда». Для организации сетей с топологией типа «звезда» и большим числом абонентских терминалов с низкой нагрузкой наиболее эффективными считаются технологии TDM/TDMA и FTDMA. В таких сетях все терминалы напрямую в один спутниковый скачок взаимодействуют только с ЦЗС (рис. 6.1).

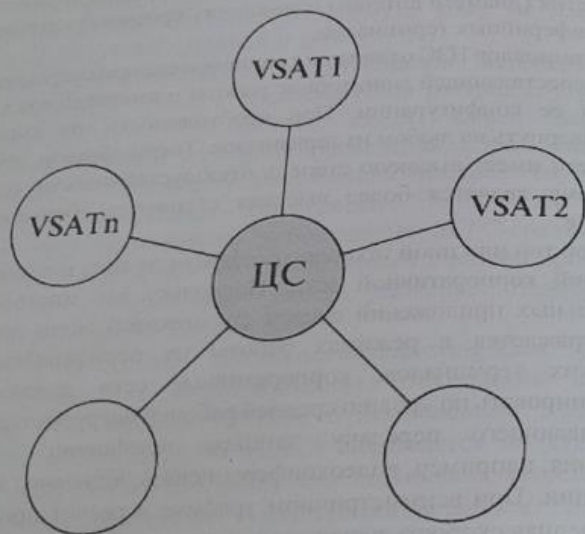


Рис. 6.1. Топология «звезда».

Благодаря этому появляется возможность применять маломощные и недорогие терминалы, компенсируя их низкую энергетику установкой на ЦЗС антенны большого диаметров свыше 5 м и более мощного. Так, например, для работы с отечественным спутником «Экспресс-6А» в Ku-диапазоне можно использовать терминалы TDM/TDMA или FTDMA с антенной, диаметром 1,8 м

и приемопередатчиком мощностью всего 1 Вт, что позволяет существенно снизить стоимость реализации проектов с большим числом терминалов. Помимо повышенных энергетических характеристик, ЦЗС должна иметь и высокий уровень надежности, поскольку от технического состояния этой станции зависит функционирование всей сети. Все это с учетом наличия средств управления сетью обуславливает высокую стоимость ЦЗС. Широко распространенными в мире спутниковыми системами, на основе которых строятся сети VSAT, являются системы PES (Personal Earth Station) фирмы Hughes Network System (США) и Skystar Advantage компании Gilat (Израиль), реализующие технологии TDM/TDMA и FTDMA соответственно.

«Смешанная» топология. Для построения больших сетей с низкой нагрузкой, предназначенных для передачи данных и организации телефонной связи между периферийными терминалами, используются системы, сочетающие в себе технологии TDM/TDMA для передачи данных и DAMA для телефонии, или системы на базе технологии FTDMA, обеспечивающие связь типа «каждый с каждым» (рис. 6.2).

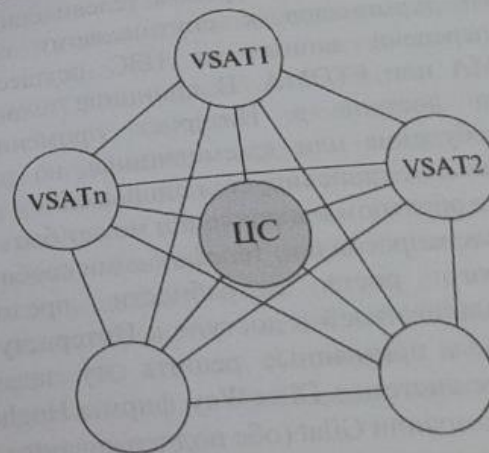


Рис. 6.2. Смешанная топология.



Терминалы таких систем связи стоят дороже, чем терминалы сетей с топологией типа «звезда», поскольку в них реализован дополнительный режим работы DAMA. К тому же для взаимодействия друг с другом они должны иметь более высокие энергетические характеристики. При использовании ретрансляторов Ku-диапазона российских спутников минимальный диаметр антенны для ЗС типа DAMA составляет 2,4 м. Системы HES (Hybrid Earth Station) фирмы Hughes Network System и Nextar AA/TDMA-BOD корпорации NEC поддерживают технологии TDM/TDMA и DAMA, а система FaraWay фирмы Gilat основана на технологии FTDMA.

*Сети широкополосного доступа в Интернет.* Отдельную группу образуют спутниковые сети связи для широкополосного доступа в Интернет. Как известно, они характеризуются значительной асимметрией трафика: объем информации, пересылаемой со всех абонентских терминалов на ЦЗС, в 3...10 раз меньше объема информации, передаваемой от самой ЦЗС к терминалу. В этих сетях для передачи высокоскоростного, до 40 Мбит/с, потока данных от ЦЗС к абонентским терминалам используется применяемая в цифровом телевидении технология DVB, а доступ терминалов к спутниковому сегменту для последующей передачи данных на ЦЗС осуществляется по технологии TDMA или FTDMA. В принципе технологии сетей широкополосного доступа в Интернет применимы и для дистанционного обучения или телемедицины, но только тогда, когда видеоизображение транслируется лишь в одном направлении — от ЦЗС, а связь в обратном направлении может быть ограничена передачей текстовых запросов или телефонными сообщениями.

Из-за быстрого роста потребности предприятий и индивидуальных пользователей в доступе к Интернету появились специальные системы, призванные решить эту задачу. К ним относятся следующие системы: DirecWay фирмы Hughes Network System и SkyBlaster компании Gilat (обе поддерживают технологию DVB), LinkWay.IP (реализуют технологию MF-TDMA) компании Comsat, вошедшей с недавних пор в состав фирмы ViaSat (США).

*Сети со средней нагрузкой.* Пользователями таких сетей обычно являются банки, производственные и торговые компании с разветвленной инфраструктурой филиалов и отделений, которым

для передачи данных и организации нескольких телефонных каналов, а возможно, и видеоконференцсвязи требуются ЗС с такой пропускной способностью 32...256 Кбит/с. Спутниковые сети с дистанционной топологией применяются для организации или IP, а их топология («звезда» или «иерархическая звезда») с числом станций от 10 до 100) используются в телемедицине. В сетях корпоративных и ведомственных сетей со средней нагрузкой и количеством удаленных терминалов до 15 часто применяют выделенные каналы SCPC и технологию MCPC. Эти сети могут иметь любую топологию, причем каждое направление связи на ЗС обеспечивается отдельным каналом.

На рынке представлены следующие системы связи с технологиями SCPC/MCPC: TRES фирмы Hughes Network System, SkyPerformer компании Clarent, а также станции спутниковой связи из отдельных модулей (антенна, приемопередатчик, спутниковый модем).

*Сети с большим числом станций.* Для организации сетей со средней нагрузкой с числом пользователей более 10 целесообразно использовать системы спутниковой связи на базе технологии MF-TDMA. Такая технология, построенная для полносвязной сети с большой пропускной способностью и равномерным трафиком передачи информации, является оптимальной. По общей совокупной стоимости содержания такие сети могут успешно конкурировать с наземными связными инфраструктурами. Тенденция развития корпоративных сетей указывает на повышенный спрос на эти системы в ближайшие годы. Типичным образцом систем MF-TDMA можно назвать SkyWAN фирмы ND SatCom (Германия) и LinkWay 2000 и ее модернизированный под VSAT-сети вариант LinkWay 2100 компании ViaSat.

Кроме спутниковых систем MF-TDMA, при большом числе терминалов и скорости передачи данных не выше 128 Кбит/с можно строить корпоративные сети на базе системы FaraWay компании Gilat или HES фирмы Hughes Network System. Эти системы обеспечивают полносвязную топологию сети и передачу данных

режиме DAMA с заданной скоростью передачи. Необходимость приобретения и установки дорогостоящей ЦЗС делает эти сети конкурентоспособными по сравнению с системами MF-TDMA только при большом числе абонентских терминалов.

*Сети с высокой нагрузкой.* Средняя рабочая скорость станции в сети с высокой нагрузкой составляет 256...2048 Кбит/с, а сами эти станции используются для передачи смешанного трафика (данные, Интернет, телефония, видеоконференцсвязь). Как правило, сети с высокой нагрузкой насчитывают небольшое число терминалов — от 5 до 25, между которыми организуется связь типа «каждый с каждым» с использованием технологии ATM или Frame Relay. Основные пользователи этого типа сетей — общероссийские холдинги или телекоммуникационные компании, использующие спутниковые каналы для объединения крупных региональных офисов в единую сеть, а также для резервирования основной наземной сети.

При построении спутниковых сетей с высокой нагрузкой для технологии MF-TDMA практически нет альтернативы, поскольку организация многочисленных высокоскоростных каналов SCPC/MCPC экономически не эффективна. Сети такого типа можно реализовать на основе MF-TDMA-систем SkyWAN фирмы ND SatCom, LinkWay 2000 компании ViaSat и VSAT Plus II фирмы NSI. Может быть также использована особенность системы LinkWay 2000, поскольку имеется возможность ее работы не только с протоколами IP и Frame Relay, но и с протоколами ATM и SS7/ISDN. Общая пропускная способность корпоративной сети, построенной на базе такой системы велика, поскольку скорость передачи информации может достигать 32 Мбит/с.

#### **Особенности организации ведомственной телефонной связи с помощью сетей VSAT**

*Прямой телефон.* При помощи спутниковых систем связи типа VSAT может быть организована прямая телефонная связь с абонентами. Это решение отличается особенностью использования телефонных аппаратов без набора номера. При снятии трубки в удаленном населенном пункте, где расположен вызывающий терминал абонента, по каналам спутниковой связи немедленно

поступает вызов на телефонный аппарат вызываемого абонента. Такой вид связи может быть использован в системах диспетчерской связи для оперативной связи медиков, пожарных, спасателей и там, где более сложная схема связи не нужна, например, для связи с охранником на удаленном объекте.

*Вынос абонентских линий офисной АТС в удаленные филиалы.* Это техническое решение используется компаниями, которые имеют центральный офис с высокоскоростным выделенным каналом в крупном городе, для связи с удаленными филиалами или подразделениями, расположенными в местах, где нет другой связи. Через спутник по технологии IP создаются абонентские внутренние телефонные линии для корпоративной АТС. В результате телефоны в удаленных точках получают внутренние простые номера, например, трехзначные. Все звонки осуществляются так же, как если бы эти телефоны физически находились в одном здании с центральным офисом. Сотрудники удаленных подразделений точно так же, как и сотрудники центрального офиса, могут воспользоваться выходом в городскую телефонную сеть набором дополнительного номера «9».

Такое решение особенно удобно тем, что для подключения удаленных подразделений не нужны дополнительные внешние телефонные линии. Таким образом компания может организовать для своих филиалов как внутреннюю, так и внешнюю телефонную связь без абонирования новых телефонных номеров.

*Связь двух офисных АТС и IP.* Это техническое решение используется для связи центрального офиса с относительно крупным филиалом, в котором недостаточно наличия одного-двух телефонов.

Две таких офисных АТС можно соединить друг с другом по IP через два шлюза FXO и сеть IP. Сотрудник центрального офиса для связи с удаленным офисом набирает внутренний номер, например, трехзначный номер шлюза. Шлюз «снимает трубку» и сотрудник слышит готовность для установления связи с дальней офисной АТС. После этого ему нужно набрать необходимый внутренний номер удаленного офиса, например, двузначный, с тем, чтобы связаться с нужным сотрудником этого офиса. При желании он может набрать код выхода на внешнюю линию (обычно «9») и позвонить любому абоненту местной, например, поселковой, АТС.

Звонок из удаленного офиса в центральный происходит аналогичным способом.

Телефонная связь между двумя абонентскими станциями можно применить в том случае, когда как удаленный филиал, так и центральный офис не имеют высокоскоростного «наземного» подключения к Интернету, но имеют подключение к спутниковому. «Двусторонний спутниковый Интернет» использует технологию, которая не позволяет напрямую связать через спутник две абонентские станции, однако связь возможна только через центральную станцию сети, «в два скачка». При такой связи имеет место довольно большая задержка сигнала во времени — от 500 миллисекунд до одной секунды, поэтому пользоваться телефонией с такими задержками очень неудобно. В подобных случаях можно использовать другие технологии, например, TDMA или SCPC. Они обеспечивают прямую связь «в один скачок», но стоят существенно дороже.

Частная сеть IP-телефонии. Этот вид телефонии может быть использован для нескольких небольших подразделений при наличии Интернета для организации связи по принципу «любой с любым», когда голосовые шлюзы могут выполнять функции небольшой «виртуальной АТС». Так можно создать частную телефонную сеть с внутренней нумерацией собственных абонентов, где связь будет осуществляться так же, как если бы все телефоны были физически подключены к одной офисной АТС без выхода в город. Телефоны, подключенные к Интернету через спутниковый канал связи, могут осуществить аналогичную процедуру, разница будет только в различии времени задержки.

#### 6.4. Общая структура сети VSAT для телефонии

##### Конфигурация периферийных станций VSAT

Типовой терминал VSAT для телефонии (ТЛФ), работающий в спутниковой телефонной сети (рис. 6.3), состоит из трех основных элементов:

- антенной системы (АС);

- блока наружной установки (БН), размещенного непосредственно на АС;
- блока внутренней установки (БВ), размещенного в помещении пользователя.

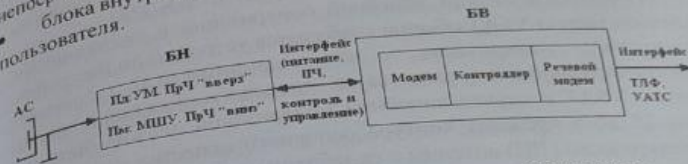


Рис. 6.3. Функциональная схема станции VSAT-ТЛФ

Где: АПКТ – аппаратура ПКТ; МШУ – маломощный усилитель;  
 Пд – передача; Пм – прием; ПрЧ – преобразователь частоты;  
 ПЧ – промежуточная частота; ТЛФ – телефон;  
 УМ – усилитель мощности; УАТС – учрежденческая АТС.

В состав антенной системы входит параболический рефлектор офсетного типа с облучающей системой и антенно-волноводным трактом (АВТ); БН размещается непосредственно на антенне. Производители выпускают широкую номенклатуру антенных систем станций VSAT с различными значениями добротности приемной системы ( $G/T$ ) и эквивалентной изотропно излучаемой мощности (ЭИИМ) для использования в спутниковых сетях с разными энергетическими характеристиками бортовых спутниковых ретрансляторов.

Для работы в диапазоне частот 14/11-12 ГГц (диапазон Ku) наиболее часто применяются малые антенны диаметром 0,75...1,8 м, хотя для регионов с высокой интенсивностью осадков могут применяться антенны большего размера. Офсетная конструкция обеспечивает минимальный уровень боковых лепестков, соответствующий огибающей  $G = 29-25 \log \theta$  дБ в соответствии с Рек. МСЭ-Р S.580-2 ( $\theta$  - угол относительно максимума диаграммы направленности антенны). В диапазоне частот 6/4 ГГц (диапазон C) антенны станций VSAT имеют несколько большие размеры

рефлектора (1,8...4,5 м) для лучшей пространственной избирательности.

Основные параметры антенных систем VSAT должны соответствовать требованиям Рек. МСЭ-Р S.727 и S.728.

При использовании линейной поляризации в диапазоне Ku антенна станции VSAT обычно снабжается устройством настройки плоскости поляризации на принимаемый сигнал. Кроссполаризационная развязка в антенно-волноводной части станции VSAT в случае линейной поляризации должна быть не менее 25 дБ в пределах контура основного лепестка диаграммы направленности (ДН) антенны с ослаблением 0,3 дБ и не менее 20 дБ в любом другом направлении (Рек. МСЭ-Р S.727 «Кроссполаризационная развязка для VSAT»).

Как правило, антенные системы станций VSAT не применяют систему слежения за спутником ввиду незначительного уровня потерь наведения при работе с ИСЗ с нестабильностью на ГО  $\pm 0,1^\circ$  в пределах основного лепестка ДН антенны. Однако ряд зарубежных производителей VSAT (HUGHES Network Systems, США, NEC Corporation, Япония) оборудуют станции VSAT системами наведения антенн с целью поставки таких станций на российский рынок для работы с существующими ИСЗ «Горизонт», характеризующимися недостаточно высокой точностью их удержания на ГО.

Наружный блок БН, реализующей функции приемопередатчика, состоит из двух основных частей: маломощного усилителя (МШУ) с маломощным приемным конвертером СВЧ/ПЧ (в англоязычной литературе Low Noise Block — LNB) в тракте приема и конвертера ПЧ/СВЧ в тракте передачи с усилителем СВЧ мощности (УМ), выполненными в герметичном всепогодном конструктиве.

Приемный блок БН обычно располагается непосредственно на облучателе антенны с целью уменьшения потерь в приемном СВЧ тракте до МШУ. Передающая часть БН (УМ и ПрЧ «вверх») монтируется на конструкциях АС, подключается к передающей СВЧ части АВТ и соединяется с внутренним блоком коаксиальным соединителем, по которому передаются сигналы ПЧ приема и передачи, электропитания наружного устройства постоянным током, сигналы контроля и управления блоком БН.

Большинство производителей станций VSAT выполняют БН в нерезервированном варианте, что упрощает конструкцию и удешевляет стоимость терминала VSAT, но предъявляет весьма высокие требования к надежности этого устройства. Типовые значения выходной мощности зарубежных БН в С/Ку диапазонах при использовании твердотельных транзисторных УМ (SSPA) составляют 2...30 Вт/1...16 Вт. При необходимости увеличения ЭИИМ станций VSAT используются УМ на основе лампы бегущей волны (БВ).

Современный МШУ в приемной части БН выполняется обычно на полевых GaAs HEMT транзисторах с минимальным коэффициентом шума (типичная эквивалентная шумовая температура современного приемника 200...220 К в диапазоне 11/12 ГГц и 50...60 К в диапазоне 4 ГГц). Для повышения надежности и удешевления оборудования VSAT используется технология гибридных монолитных СВЧ интегральных схем.

Для удобства размещения станции VSAT у пользователя максимальная длина соединительного коаксиального кабеля между БН и БВ может быть 100...200 м.

Излучение станций VSAT в сторону ГО и паразитные излучения жестко нормируются, причем в связи с возможностью размещения достаточно большого числа станций VSAT на ограниченной территории параметры их излучения должны быть ограничены более жестко, чем параметры больших ЗС ФСС.

Типовой блок внутренней установки БВ (см. рис. 6.3.) состоит из модема и компьютеризированного цифрового управляющего устройства (контроллера АПКТ), а также речевого кодека. БВ обеспечивает интерфейс с БН по ПЧ, питанию, дистанционному контролю и управлению и аналоговый интерфейс с необходимыми типами оконечного оборудования пользователя для передачи речевой информации, сигналов факса или телекса.

В варианте телефонной сети VSAT в составе БВ находится речевой кодек, обеспечивающий преобразование аналогового телефонного сигнала в цифровую форму; наиболее распространенным вариантом преобразования является адаптивная дифференциальная ИКМ (АДИКМ) со скоростью 32 кбит/с в соответствии с Рек. МСЭ-Р G.721, хотя в выделенных сетях для передачи речевой информации и сигналов факса по телефонному

каналу часто, используется АДКМ с более низкими; скоростями: 24 и 16 кбит/с. Помимо речевой информации в цифровую форму преобразуются и служебные сигналы сигнализации, передаваемые по абонентскому телефонному интерфейсу при установлении соединения.

Система с предоставлением каналов по требованию, действующая под управлением ЦУС сети VSAT, обеспечивает эффективное использование пропускной способности спутникового ретранслятора в режиме незакрепленных каналов, предоставляемых абонентам сети VSAT по требованию.

В состав модема VSAT включается дополнительный преобразователь частоты, позволяющий обеспечить частотное разделение при совместной передаче сигналов контроля и управления, а также передаваемого и принимаемого сигналов ПЧ по коаксиальному кабелю между наружным и внутренним блоками станции VSAT.

Скорость передачи информации цифровых модемов телефонных терминалов VSAT составляет 19,2...35,0 кбит/с с учетом передачи дополнительной служебной информации, модуляция – ФМ-2/ФМ-4. Практически во всех современных станциях модема входит цифровой кодек (кодер-декодер) помехоустойчивого кода с «прямым» исправлением ошибок. Наиболее распространенный способ кодирования - применение в тракте передачи кодера сверточного кода (СК) с относительными скоростями кодирования  $R=1/2$ ,  $3/4$  и  $7/8$ . В тракте приема на выходе когерентного демодулятора ФМ сигналов используется декодер СК, реализующий один из двух наиболее эффективных алгоритмов декодирования:

- 1) алгоритм Витерби (декодирование по методу максимального правдоподобия);
- 2) последовательный алгоритм в сочетании с «мягким» (квантованным) решением по каждому принимаемому символу.

Энергетический выигрыш от применения вышеупомянутых алгоритмов кодирования (ЭВК) при относительной скорости кода  $R = 1/2$  составляет 5,5...6,5 дБ при вероятности ошибки на выходе  $P_{\text{ош}} = 1 \cdot 10^{-6}$ . С увеличением относительной скорости кодирования до  $R = 3/4$ ,  $7/8$  ЭВК уменьшается соответственно на 1...2 дБ.

Дополнительное увеличение ЭВК на 2,5...3,0 дБ достигается при каскадном включении кодера СК и кодера кода Рида-Соломона, предназначенного для борьбы с пакетированием ошибок на выходе декодера СК. При использовании в модемах станций VSAT таких сигнально-кодowych конструкций должны выполняться весьма жесткие требования к возможности перескоков фазы тактовой и несущей частот в системах синхронизации когерентных ФМ демодуляторов ввиду весьма низкого отношения  $P_c/P_{\text{ш}}$  в рабочей полосе частот.

Генераторное оборудование аппаратуры VSAT содержит в составе блоков БВ или БН высокостабильный опорный генератор диапазона частот 10...100 МГц с весьма высокими требованиями к спектральной «чистоте» и долговременной стабильности частоты выходного сигнала, который используется для формирования гетеродинных частот в ПрЧ «вверх» и ПрЧ «вниз». Типовое значение долговременной стабильности частоты применяемых генераторов не хуже  $1 \cdot 10^{-7}$  в год.

4. Система контроля и управления, входящая в состав аппаратуры станции VSAT, должна соответствовать требованиям Рек. МСЭ-Р S.729 «Контроль и управление станциями VSAT». Согласно этой рекомендации каждая периферийная станция VSAT должна работать под постоянным контролем ЦУС, гарантирующим недопущение помех другим станциям сети и другим системам при возникновении нештатных ситуаций на необслуживаемых станциях VSAT. С этой целью в сетях VSAT должно быть предусмотрено дистанционное управление со стороны ЦУС по радиоканалу ЦУС-VSAT частотой и мощностью передачи станций VSAT в соответствии с сетевым трафиком, а также запрет на излучение мощности VSAT в аварийных ситуациях.

Во избежание нежелательного излучения в сторону соседних спутников при случайном смещении положения антенны необслуживаемой приемопередающей станции VSAT на каждой станции VSAT необходимо иметь систему защиты (контроля и управления), не допускающую излучения мощности до тех пор, пока не будет принят со спутника сигнал с центральной станции управления этой сетью VSAT.

Рассмотренный комплект оборудования станции VSAT обеспечивает организацию одного дуплексного телефонного

канала, предоставляемого в закреплённом режиме или по требованию. Как правило, БВ имеет модульную структуру для нескольких телефонных каналов и допускает наращивание числа абонентских комплектов оборудования для увеличения числа трафика. Интерфейс пользователя реализован в 2-проводном на прямое подключение телефонного аппарата или учрежденческой АТС (УАТС).

#### Конфигурация центральной управляющей станции телефонной сети

Центральная управляющая станция (ЦУС) телефонной сети VSAT содержит антенну большого диаметра с системой автоматического слежения за спутником, радиочастотное оборудование и оборудование полос модулирующих частот. Конфигурация ЦУС имеет модульную структуру, которая позволяет экономично наращивать объём сетевого трафика по мере развития сети и расширения номенклатуры услуг потребителям.

Антенна ЦУС имеет диаметр от 4,5 (6,0) до 11,0 м с целью экономии мощности передатчиков периферийных станций VSAT и энергетического ресурса спутникового ретранслятора.

Первичный контроллер АПКТ, являющийся ядром централизованной системы (ПКТ), выполняет функции контроля и управления сетью и предоставлением каналов по требованию, взаимодействует по общему каналу сигнализации (ОКС) с каждым вторичным канальным контроллером АПКТ терминалов VSAT.

В состав ЦУС, участвующей в трафике, дополнительно включаются блоки каналообразующего оборудования: модемы, вторичные канальные контроллеры АПКТ и речевые кодеки, модульно наращиваемые при увеличении емкости сети.

Система ПКТ рассчитана на обслуживание 256 дуплексных телефонных каналов и число обслуживаемых оконечных канальных блоков составляет 2000 шт.

#### 6.5. Мультисервисная DVB-RCS платформа для сетей VSAT

Компаниями MediaSputnik и EMS Satellite Networks разработана мультисервисная DVB-RCS платформа, которая обеспечивает высокоскоростной спутниковый доступ с приложениями реального времени (передача данных, голос, видео и т. д.), а также стандартные IP приложения (Интернет/Инtranет, электронная почта, передачи файлов и т. д.). Топология такой сети на базе мультисервисной DVB-RCS платформы чаще всего строится по типу «звезда» и подразумевает наличие двух трактов передачи:

- прямой канал — спутниковый канал от Центральной земной станции (ЦЗС/HUB) до удаленных спутниковых интерактивных терминалов (СИТ/SIT);
- обратный канал — спутниковый канал от терминала до Центральной земной станции.

Стандарт DVB-RCS утвержден Европейским Институтом Стандартизации в области Связи (ETSI) в 2000 г. Он предлагает прямой канал, основанный на формате данных DVB/MPEG 2, и обратный канал на основе режима множественного доступа с разделением по времени (MF-TDMA). Широкополосная несущая DVB/MPEG-2 может обеспечить скорость передачи в прямом канале до 110 Мбит/с, а режим MF-TDMA предусматривает скорость до 2...4 Мбит/с в обратном канале от каждого удаленного терминала.

Платформа DVB-RCS обеспечивает широкий спектр телекоммуникационных услуг, включая доступ к глобальной сети Интернет, построение географически распределенных LAN/WAN, передачу данных, организацию речевых каналов и видеоконференций по требованию.

При этом спутниковые терминалы могут использоваться для различных уровней потребителей услуг: от крупных предприятий и провайдеров услуг до конечных пользователей. Прямой канал системы соответствует требованиям стандартов MPEG2 и ETSI/DVB, регламентирующих цифровое телевидение. Трафик прямого канала мультиплексируется на Центральной земной станции в общий широкополосный поток DVB/MPEG-2 и ретранслируется через спутник на сеть станций VSAT. Этот поток

передается с модуляцией QPSK и кодированием Витерби/Рида-Соломона. Для корректного взаимодействия сети Интернет/Интранет с локальными сетями и передачи данных от станций пользователей до центральной станции используются широко известные сетевые стандарты и протоколы, в частности протоколы маршрутизации в среде Интернет (IP) и асинхронный режим передачи (ATM).

Данные о распределении спутникового ресурса, включая временные и частотные слоты, а также другие служебные сообщения также мультиплексируются в общий транспортный поток прямого канала.

По установленному стандарту алгоритму удаленные терминалы синхронизируются с прямым каналом, регистрируются на Центральной земной станции и для каждого терминала выделяется частотновременной план в терминах MF-TDMA слотов. При этом удаленные терминалы используют разноплановую MF-TDMA схему доступа в спутниковую сеть. Так, например, режим MF-TDMA обеспечивает группам терминалов первичную связь с ЦЗС с использованием слотированных пакетов ALOHA. Центральная же земная станция выделяет своим авторизированным и активным терминалам ряд пакетов, каждый из которых назначает частоту, полосу, время начала и длительностью временного слота.

Основанная на открытом стандарте ETSI EN 301 790 данная технология связи обеспечивает совместимость оборудования различных производителей в одной спутниковой сети.

Мультисервисная DVB-RCS платформа обеспечивает скорость передачи данных в одном прямом канале до 110 Мбит/с при использовании технологии DVB-S2 и скорость одного обратного канала до 4 Мбит/с с общим количеством до нескольких тысяч зарегистрированных в системе терминалов.

В стандарте DVB-RCS использование технологий DVB-S(S2) для прямого канала оправдывается прежде всего экономическими соображениями. Недорогие массовые DVB-S2 компоненты уже существуют и применяются, в то время как применение возможно более эффективных схем и фирменных стандартов потребуют повышенных расходов и большего времени на внедрение и развитие оборудования для центральных станций и терминалов.

Система DVB-RCS при использовании каскадного кодирования Рида-Соломона и Витерби обеспечивает более высокую эффективность полосы спутникового канала связи при заданном коэффициенте ошибок, а следовательно, увеличивает эффективность использования бортовой мощности. Используемые версии DVB-RCS системы еще более улучшают эти характеристики за счет применения турбокодирования, которое обеспечивают повышение энергетике более чем на 1 дБ в зависимости от длины пакета.

В DVB-RCS всегда можно выбрать тип пакетов трафика в обратном канале, поскольку IP пакеты могут передаваться как в ATM ячейках, так и в MPEG2 пакетах, несмотря на то что более короткие ATM пакеты в большинстве и с точки зрения задержек. В обратном канале DVB-RCS системы используется модуляция QPSK, которая признана в качестве оптимального компромисса между требуемой мощностью и эффективностью использования полосы для множественного доступа через спутниковые сети.

Мультисервисная платформа DVB-RCS поддерживает следующие сервисы и режимы доступа к данным. Доступ в режиме *Unicast (точка-точка)*. Система DVB-RCS через спутниковую сеть обеспечивает доступ в этом режиме для каждого абонентского терминала к ресурсам корпоративной мультисервисной сети или к сети Интернет. В обратном канале используется описанный в RFC-2684 метод LLC-SNAP инкапсуляции IP поверх AAL5/ATM.

Терминал VSAT может быть при этом непосредственно подключен к локальной компьютерной или телефонной сети, и всем станциям назначается статический IP-адрес.

Режим *Multicast*. Схема реализации режима *Multicast* разработана таким образом, чтобы не было необходимости передачи сигнализации IGMP, занимающей значительную часть пропускной способности сети. Данные при этом передаются с использованием основанного на рекомендациях RFC 1112 (передача IP Multicast по PID таблице) соответствия между IP multicast адресами и MAC адресами multicast трафика. На стороне пользовательских станций параметры DVB и IP multicast

терминалов настраиваются на соответствие IP multicast адресам, локальным PID фильтрам и фильтрам MAC адресов. Для приема multicast трафика на спутниковом терминале не требуется использование обратного канала.

**Поддержка SLA.** Система DVB-RCS может поддерживать процедуру управления соглашением об уровне обслуживания (Service Level Agreement, SLA), представляющее собой контракт между клиентом и провайдером услуг, между внешней сетью, спутниковой сетью и локальными сетями. Функция управления SLA в прямом канале и обратном канале заложена в базовой конфигурации системы.

**Поддержка VPN.** VPN — виртуальные персональные сети — ориентированы на организацию связи с удаленными друг от друга подразделениями. Применение VPN сокращает эксплуатационные расходы, поскольку инфраструктура принадлежит сетевому оператору, распределяющему ресурсы физической сети между корпоративными VPN. По существу, технология VPN выполняет строгое разделение между различными сетями этого вида в целях обеспечения конфиденциальности и качества обслуживания.

**Поддержка виртуальной ЛВС (VLAN)** представляет собой логическую группу сегментов ЛВС, не зависящую от физического местоположения и организованную на основе общего набора критериев.

**IP-телефония.** Решение по IP-телефонии базируется на комбинации внешних VoIP шлюзов или Softswitch платформ со спутниковой сетью доступа DVB-RCS.

**Поддержка топологии «вложенная звезда» (Multistar).** Мультисервисная DVB-RCS платформа в топологии «вложенная звезда» состоит из главного шлюза и нескольких меньших ведомых шлюзов, расположенных в пределах зоны охвата главного шлюза. Каждый из ведомых шлюзов может автономно обслуживать свою подсеть терминалов.

Сеть на основе мультисервисной DVB-RCS платформы состоит из центральной наземной станции, множества удаленных пользовательских терминалов и спутника, обеспечивающего каналы связи в прямом и обратном направлении.

Центральная земная станция обеспечивает:

- передачу трафика прямого канала через ретранслятор на удаленные терминалы;
- прием от ретранслятора и маршрутизацию трафика обратного канала от удаленных терминалов;
- сетевую синхронизацию станций;
- распределение спутниковых ресурсов для спутниковых станций;
- подготовку и процедуру вещания данных и мультимедиа в различных режимах;
- аутентификацию абонентов и учет трафика;
- локальное или удаленное управление оборудованием ЦЭС и вспомогательным сетевым оборудованием.

**Подсистема прямого канала** предназначена для передачи данных через широкополосный спутниковый канал в направлении пользовательских терминалов VSAT. Данная технология основана на Европейском индустриальном стандарте ETSI EN 301 190, спецификации которого определяют механизмы инкапсуляции блоков IP данных в DVB поток и транспортировки частных данных в MPEG2 транспортном информационном потоке.

Мультиплексированный транспортный поток передается на DVB модулятор и после этого в радиочастотный тракт прямого канала, где осуществляется вставка PCR пакетов.

IP/DVB инкапсулятор/мультиплексор выполняет следующие основные функции:

- инкапсуляцию IP TCP, UDP, ICMP пакетов в транспортный поток MPEG2;
- мультиплексирование внешнего транспортного MPEG2 потока непосредственного вещания, трафика IP данных и информации сигнализации прямого канала в общий выходной транспортный поток со скоростью передачи данных до 110 Мбит/с;

- обеспечение прозрачной передачи сформированного транспортного потока для организации аудио-видеовещания в режиме DVB/MPEG2;

- назначение скорости передачи данных для каждого IP адреса приемника в диапазоне от 256 Кбит/с до 110 Мбит/с.

**Модулятор DVB-S(S2)** кодирует, модулирует и преобразовывает с повышением частоты транспортный MPEG поток прямого канала. Модулятор обеспечивает реализацию



нескольких возможных скоростей упреждающего исправления ошибок (FEC), основанного на коде Рида-Соломона. Исходящий сигнал после этого модулируется с использованием фазовой QPSK или 8-PSK модуляции. Модулированный сигнал преобразуется в диапазон L (950...1450 МГц) и передается по радиочастотный тракт ЦЗС.

Сервер QoS является средством многоуровневого управления и контроля трафика, распознающим тип проходящего через сеть трафика. Этот сервер управляет трафиком на основе правил (или «политик»), которые устанавливаются в зависимости от специфических требований оператора.

Коммутатор DVB-ASI обеспечивает «горячее» переключение на резервный комплект подсистемы прямого канала по команде с сервера контроля или автономно.

Подсистема обратного канала принимает абонентский трафик и информацию сигнализации от пользовательских станций, а также готовит решения на запросы удаленного доступа (разрешение на вход в систему, распределение полосы и временных интервалов). Эта подсистема принимает, преобразовывает с понижением частоты, демодулирует и декодирует IP трафик абонента, инкапсулированный в ячейки ATM и передаваемый на MF-TDMA несущей со скоростью до 2 Мбит/с.

В обратном канале используется схема многостанционного доступа типа MF-TDMA (множественный доступ с частотно-временным разделением каналов). Существуют четыре типа радиочастотных пакетов обратного канала: трафик (TRF), захват синхронизации (ACQ), синхронизация (SYNC) и канал общей сигнализации (CSC). Пакеты трафика используются для передачи в обратном канале необходимых данных, в этом случае полезная нагрузка представляет собой 53-байтовые ячейки ATM.

Все типы пакетов передаются после защитного временного интервала, вводимого в целях снижения мощности передачи и компенсации ошибок синхронизации.

Эта схема является более эффективной по сравнению с традиционно используемыми для сетей VSAT, такими, как FDMA/TDMA или SCPC (single channel per carrier — один канал на несущую), где станции на время сеанса присваивается одна несущая. В системах FDMA/TDMA применяется процедура

перераспределения информации в целях выравнивания пропускной способности среди ряда несущих. Планировщик составляет планы работы спутниковых станций в двумерном частотно-временном пространстве исходя из условия невозможности одновременного проведения передачи каждой станцией более чем на одной частоте. Пропускная способность сети используется при этом практически полностью, избегая появления фрагментации информации и неиспользуемых временных «дыр», что может происходить при работе схем SCPC и FDMA/TDMA.

Протокол планирования радиочастотного ресурса представляет собой один из вариантов коллективного доступа на основе комбинации свободного распределения ресурса и распределения ресурса по требованию. Протокол разработан для работы со всеми категориями запросов с целью обеспечения высокой пропускной способности. Особенностью данного протокола является отсутствие столкновений при доступе к каналу после первоначального входа при использовании протокола случайного доступа Slotted ALOHA.

Подсистема обратного канала также управляет станциями для балансировки нагрузки, компенсации затуханий в атмосфере, регулирования мощности передачи спутниковых станций. В соответствии со стандартом DVB-RCS используется схема управления мощностью передачи с замкнутым циклом, на основе измерения на ЦЗС отношения  $E_b=N_0$  в обратном канале.

Подсистема обратного канала включает следующие основные модули:

- мультичастотный MF-TDMA демодулятор (Multi-Carrier Demodulator) с модулем преобразования частоты (MIF);
- трафик-процессор (Traffic Processor) с модулем вставки PCR;
- процессор сигнализации (Signaling Processor);
- процессор управления (OAM Processor).

Мультичастотный MF-TDMA демодулятор принимает пакетный трафик из радиочастотного тракта приемного канала земной станции и производит демодуляцию MF-TDMA пакетов. После демодуляции приемник отделяет ATM ячейки от ячеек SAC и CSC пакетов трафика сигнализации. Демодулированные ATM пакеты направляются трафик-процессору, а CSC пакеты передаются процессору сигнализации для обеспечения начальной

инициализации станции при входе в сеть. Демодулятор также собирает статистику измерений параметров демодулированных пакетов и пересылает данную информацию процессору сигнализации для последующей обработки.

Трафик-процессор принимает упакованные в формат ячейки ATM данные от демодулятора и направляет их в наземную сеть через коммутатор, а если в состав подсистемы обратного канала входит более одного демодулятора, то к граничному маршрутизатору. Трафик-процессором поддерживается операция вставки эталонов времени в транспортный поток прямого канала.

Процессор сигнализации имеет в своем составе контроллер терминалов (SIT Controller), планировщик (Scheduler) и контроллер сигнализации прямого канала (FSH).

Процессор управления в подсистеме обратного канала управляет процессором сигнализации каждого соответствующего демодулятора. Каждый процессор управления может управлять двумя процессорами сигнализации, каждый из которых приписывается к определенному банку демодуляторов.

Технические данные, приведенные ниже дают представление об основных параметрах типовой станции VSAT, использующей технологию DVB-RCS.

Спутниковые интерактивные терминалы (СИТ) обеспечивают доступ в спутниковые сети с топологиями типа «звезда», «вложенная звезда» и «полносвязная» (опция).

### 6.6. Расчет параметров приемника ЗС

В параграфе приводится расчет шумовой температуры приемного тракта ЗС, мощность шумов на входе приемника ЗС, мощность сигнала на входе приемника ЗС, предельно допустимой пропускной способности ствола и выбор стандарта цифрового вещания.

#### Цель работы:

- Определить шумовую температуру приемного тракта ЗС.
- Определить мощность шумов на входе приемника ЗС.
- Определить мощность сигнала на входе приемника ЗС.

- Определить предельно допустимую пропускную способность ствола.
- Выбрать стандарт цифрового вещания.

#### Методические указания к расчету:

1. По заданным величинам из таблицы 6.1 выписать исходные данные для расчета по варианту.
2. Шумовая температура приемного тракта определяется по формуле:

$$T_{\text{пр}} = T_0 \cdot (K_{\text{ш}} - 1), \text{ К} \quad (1)$$

где  $K_{\text{ш}}$  – коэффициент шума приемника ЗС (задается по варианту)  
 $T_0$  – шумовая температура АВТ ЗС ( $T_0 = 290 \text{ К}$ )

Для учета помех от других систем связи необходимо увеличить  $P_{\text{ш}}$  на 20%.

3. Суммарная шумовая температура:

$$T_{\Sigma} = T_A + T_0 \left( \frac{1 - \eta}{\eta} \right) + T_{\text{пр}} / \eta, \text{ К} \quad (2)$$

где  $T_A$  – шумовая температура антенны ЗС (задается по варианту);  
 $\eta$  – КПД АВТ ЗС (задается по варианту);

4. Определение мощности шумов на входе приемника ЗС:

$$P_{\text{ш}} = k \cdot T_{\Sigma} \cdot \Delta f_{\text{ств}}, \text{ Вт} \quad (3)$$

где  $k = 1,38 \cdot 10^{-23}$  – постоянная Больцмана;  
 $T_{\Sigma}$  – суммарная шумовая температура, К;  
 $\Delta f_{\text{ств}}$  – эффективная полоса частот ствола (задается по варианту).

5. Определение мощности сигнала на входе приемника ЗС.

Для спутниковых систем связи:  $P_c/P_{ш} = 10 \dots 12$  дБВт  
 Для корректных расчетов необходимо мощность шумов, определенную из уравнения 3, перевести из Вт в дБВт. Для этого можно воспользоваться уравнением:

$$P \text{ (дБВт)} = 10 \log_{10}(P \text{ (Вт)} / 1 \text{ (Вт)}) \quad (4)$$

$$P_{c \text{ пр}} = P_{ш} + P_c/P_{ш}, \text{ дБВт} \quad (5)$$

6. Определение предельно допустимой пропускной способности ствола:

$$C = \Delta f_{\text{ств}} \cdot \log_2(1 + P_c/P_{ш}), \text{ бит/сек} \quad (6)$$

7. Существует несколько стандартов цифрового телевидения:

4:4:4 – 324 Мбит/сек;

4:2:2 – 216 Мбит/сек;

4:1:1 – 162 Мбит/сек.

Стандарт цифрового телевидения выбираем по величине пропускной способности ствола.

**Пример расчета:**

1. Исходные данные:

Коэффициент шума приемника ЗС  $K_{ш} = 8$ .

Шумовая температура антенны ЗС  $T_A = 40 \text{ К}$ .

КПД АВТ ЗС  $= 0,8$ .

Эффективная полоса частот ствола  $\Delta f_{\text{ств}} = 72 \text{ МГц}$ .

2. Шумовая температура приемного тракта:

$$T_{пр} = T_0 \cdot (K_{ш} - 1) = 290 \cdot (8 - 1) = 2030 \text{ К}$$

3. Суммарная шумовая температура:

$$T_{\Sigma} = T_A + T_0 \cdot ((1 - \eta) / \eta) + T_{пр} / \eta = \\ = 40 + 290 \cdot ((1 - 0,8) / 0,8) + 2030 / 0,8 = 2650 \text{ К}$$

4. Определение мощности шумов на входе приемника ЗС:

$$P_{ш} = k \cdot T_{\Sigma} \cdot \Delta f_{\text{ств}} = 1,38 \cdot 10^{-23} \cdot 2650 \cdot 72 \cdot 10^6 = 2,63 \cdot 10^{-12} \text{ Вт}$$

Для учета помех от других систем связи необходимо увеличить  $P_{ш}$  на 20%.

$$P_{ш} = 2,63 \cdot 10^{-12} \cdot 1,2 = 3,16 \cdot 10^{-12}$$

5. Определение мощности сигнала на входе приемника ЗС.  
 Примем, что  $P_c/P_{ш} = 10$  дБВт

Для корректных расчетов необходимо мощность шумов перевести из Вт в дБВт. Для этого воспользуемся уравнением:

$$P \text{ (дБВт)} = 10 \log_{10}(P \text{ (Вт)} / 1 \text{ (Вт)}) =$$

$$= 10 \cdot \log_{10}(3,16 \cdot 10^{-12}) = -115 \text{ дБВт}$$

$$P_{c \text{ пр}} = P_{ш} + P_c/P_{ш} = -115 + 10 = -105 \text{ дБВт}$$

6. Определение предельно допустимой пропускной способности ствола:

$$C = \Delta f_{\text{ств}} \cdot \log_2(1 + P_c/P_{ш}) = 72 \cdot 10^6 \cdot \log_2(1 + 10) = 252 \cdot 10^6 = \\ = 252 \text{ Мбит/сек}$$

7. Из приведенных стандартов цифрового телевидения по пропускной способности подходит стандарт 4:2:2.

Таблица 6.1.

Исходные данные по вариантам

Последняя цифра номера по журналу	$K_{ш}$	$T_A$ К	$\eta$	$\Delta f_{\text{ств}}$ МГц
1	6	30	0,75	60
2	7	40	0,8	64
3	8	50	0,85	68

4	9	60	0,9	72
5	6	70	0,95	74
6	7	30	0,95	70
7	8	40	0,9	60
8	9	50	0,85	64
9	7	60	0,8	62
0	8	70	0,75	68

### Контрольные вопросы.

1. Что собой представляет земная станция спутниковой сети связи VSAT?
2. Какое основное оборудование содержит абонентский терминал сети VSAT?
3. Какие основные топологии используются при организации спутниковой сети связи VSAT?
4. Как конструктивно отличаются земные станции в зависимости от вариантов топологии?
5. На какие классы делятся земные станции VSAT по пропускной способности?
6. На основе каких наиболее распространенных спутниковых систем строятся сети VSAT?
7. Какие известные мировые фирмы и компании поставляют оборудование для сетей VSAT?
8. Какие основные технологии многостанционного доступа реализованы в оборудовании этих разработчиков?
9. Какие виды ведомственной телефонной связи могут быть организованы с помощью сетей VSAT?
10. Назовите основные особенности мультисервисной DVB-RCS платформы для спутниковых сетей связи VSAT. Какие типы многостанционного доступа используются в сетях VSAT?
11. Приведите функциональную схему станции VSAT-ТЛФ.
12. Приведите функциональную схему ЦУС/ЦЗС телефонной сети VSAT.

### ГЛАВА 7. Изучение принципов построения передающей части аппаратуры временного уплотнения каналов

Временное уплотнение каналов имеет ряд преимуществ перед частотным. Аппаратура временного уплотнения значительно проще, чем частотного, в ней отсутствуют сложные полосовые фильтры.

Временное уплотнение канала связи заключается в том, что сигналы различных сообщений передаются последовательно во времени. Установленные на передающей и приемной сторонах коммутаторы, работающие синхронно, обеспечивают подключение соответствующих сообщений к своим приемным устройствам. Выбор частоты переключения определяется из условия точности восстановления сообщений с учетом ограниченной полосы частот, пропускаемых каналом связи.

Особенность временного уплотнения канала связи заключается в том, что сигналы различных сообщений передаются последовательно во времени. Установленные на передающей и приемной сторонах коммутаторы, работающие синхронно, обеспечивают подключение соответствующих сообщений к своим приемным устройствам. Выбор частоты переключения коммутаторов определяется из условия точности восстановления сообщений с учетом ограничений по полосе частот, пропускаемых каналом связи. Восстановление сообщения по зарегистрированным дискретным сообщениям в темпе приема сигналов осуществляется с помощью фильтра нижних частот.

Системы с временным уплотнением каналов сравнительно не чувствительны к нелинейности трактов передачи, так как последняя не создает здесь, как в системах с частотным уплотнением, переходных помех, а лишь искажения в отдельных каналах; переходные помехи возникают лишь тогда, когда из-за переходных процессов в селективных системах от импульса одного канала в следующем канале действуют остаточные напряжения.

### 7.1. Структурная схема имитационного макета аппаратуры уплотнения

Основой устройства является модулятор, позволяющий объединить в себе, с целью эффективной визуализации, три вида модуляции – АИМ, ШИМ, ФИМ, а также получить групповой 4-х канальный сигнал по этим видам с учетом правил ВРК. Возможность наблюдения с помощью осциллографа всех характерных сигналов, учитывается выведение на переднюю панель макета соответствующих контрольных точек.

На рисунке 7.1 изображена структурная схема макета.

Схема содержит:

- генератор тактовых импульсов 1;
- счетчик-делитель 2;
- формирователь импульса запуска ЛИН 3;
- формирователь маркерного импульса синхронизации 4;
- генератор ЛИН 5;
- сравнивающее устройство (компаратор) 6;
- аналоговый коммутатор 7;
- преобразователь нормализатор (ПН) 8;
- формирователь входных сигналов 9;
- суммирующее устройство 10;
- компьютер 11;
- узел питания 12;
- осциллограф 13;
- контрольные гнезда (Г1..Г8) 14.

Тактовые импульсы с периодом следования  $T_T$  вида  $U_1$  от выхода генератора тактовых импульсов (контрольное гнездо Г1, далее Гп), поступают к инверсному счетному входу четырехразрядного двоичного счетчика 2. На выходах разрядов ( $2^0...2^3$ ) счетчика появляются двоично-зависимые импульсные последовательности  $U_2, U_3, U_4$  (Г2, Г3, Г4), а также импульсная последовательность ( $2^3$ ), предназначенная для синхронизации начала развертки в осциллографе (на диаграмме не показана). Задним фронтом импульсов вида  $U_2$  с периодом следования  $2T_T$  (Г2), по инверсному входу, запускается формирователь 3 и на его инверсном выходе (Г5) появляется отрицательный перепад напряжения вида  $U_5$ , или отрицательный импульс длительностью

дл., который в свою очередь способствует запуску (началу действия) генератора линейно изменяющегося напряжения ГЛИН. Это напряжение вида  $U_6$  (гнездо Г6), далее подается на первый вход второй вход компаратора (инверсный) поочередно, с периодом  $2T_T = T_k$ , подаются входные значения напряжения каждого из четырех каналов ( $U_{вх1}...U_{вх4}$ ) с выхода аналогового коммутатора 7. Графики линейно изменяющегося напряжения каждого из пересечения в моменты времени  $t_1...t_4$ , появления которых зависит от величины  $U_{вхi}$ . В результате сравнения этих напряжений, на выходе компаратора формируется сигнал ШИМ.

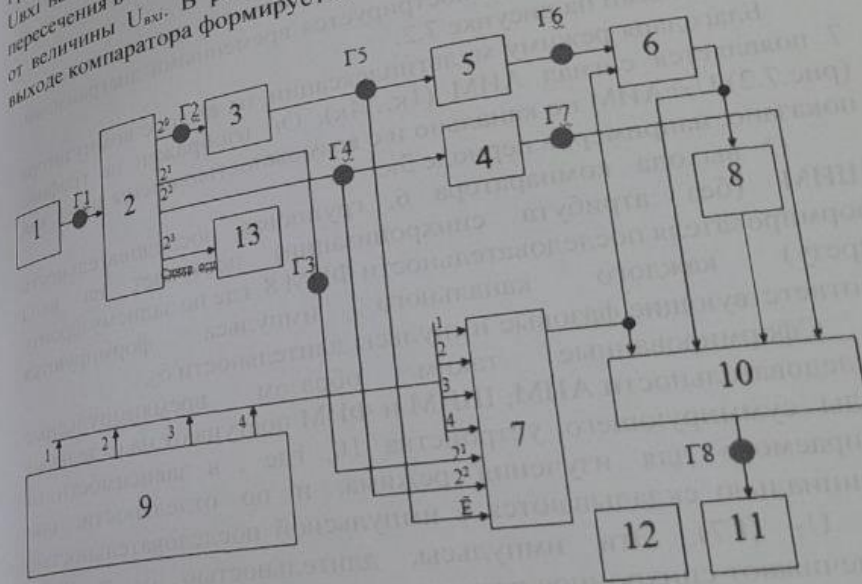


Рис. 7.1. Структурная схема аппаратуры уплотнения.

Режим поочередного сравнения входных напряжений каждого из каналов обеспечивается путем автоматической коммутации входных канальных аналоговых сигналов, поступающих из формирователя входных сигналов 9 (1...4) через входы 1...4

аналогового коммутатора 7, работающего в режиме мультиплексора – на выход коммутатора. Этот режим обуславливается наличием в составе коммутатора встроенного дешифратора управляемого разрядами счетчика  $2^1$  и  $2^0$ . При этом вход Е коммутатора служит для отключения (подавления) его информационных входов на интервал времени  $\delta_1$ .

Это способствует, во-первых, устойчивому возобновлению процесса ЛИН, во-вторых – созданию, так называемого защитного межканального интервала времени  $\delta_1$ , в течении которого, например, может происходить перестройка приемной части на прием очередного канала.

Работа структурной схемы имитационного макета аппаратуры уплотнения (рис. 7.1) иллюстрируется временными диаграммами, изображенными на рисунке 7.2.

Благодаря режиму мультиплексации, на выходе коммутатора 7 появляется сигнал АИМ (1к...4к). Он изображен на графике (рис. 7.2)  $U_7$ +АИМ по канално и с возможностью смены знака, как показано, например, в периоде  $3к$ .

С выхода компаратора 6, групповая последовательность ШИМ (без атрибута синхронизации) поступает на вход формирователя последовательности ФИМ 8, где по заднему фронту (срезу) каждого канального импульса формируются соответствующие фазовые импульсы длительности  $\delta_2$ .

Сформированные таким образом времяимпульсные последовательности АИМ, ШИМ и ФИМ поступают на отдельные входы суммирующего устройства 10, где, в зависимости от выбираемого для изучения режима, и по отдельности, они потенциально складываются с импульсной последовательностью вида  $U_7$  ( $\Gamma 7$ ). Эти импульсы, длительностью  $\delta_3 \approx 0,5\delta_1$ , обеспечивают синхронное разграничение групповых периодов  $T_1$  и по времени действуют в первой половине защитного интервала перед каждым первым канальным интервалом  $T_k$ . Синхроимпульсы  $U_7$  поступают на отдельный вход суммирующего устройства 10 от прямого выхода формирователя импульсов групповой синхронизации 4, который, в свою очередь запускается через воздействие на его инверсный вход заднего фронта импульсов последовательности  $U_4(\Gamma 4)$  третьего разряда (22) - счетчика

делителя 2. Действие синхроимпульсов  $U_7(\Gamma 7)$  происходит с периодом времени  $8T_1 = T_1$ , являющимся групповым периодом. Диапазон изменения времяимпульсной характеристики преобразования ШИМ и ФИМ сигналов обозначен на диаграмме как  $\Delta t$ , а диапазон изменения амплитудной характеристики преобразования АИМ сигнала обозначен как  $\Delta U$ .

Сформированные таким образом канальные импульсные последовательности АИМ, ШИМ, ФИМ групповых сигналов совместно с синхросигналом могут быть по отдельности поданы на вход осциллографа 13 или на информационный аналоговый вход компьютера 11, для их визуализации и изучения, как в статическом так и динамическом режимах работы макета.

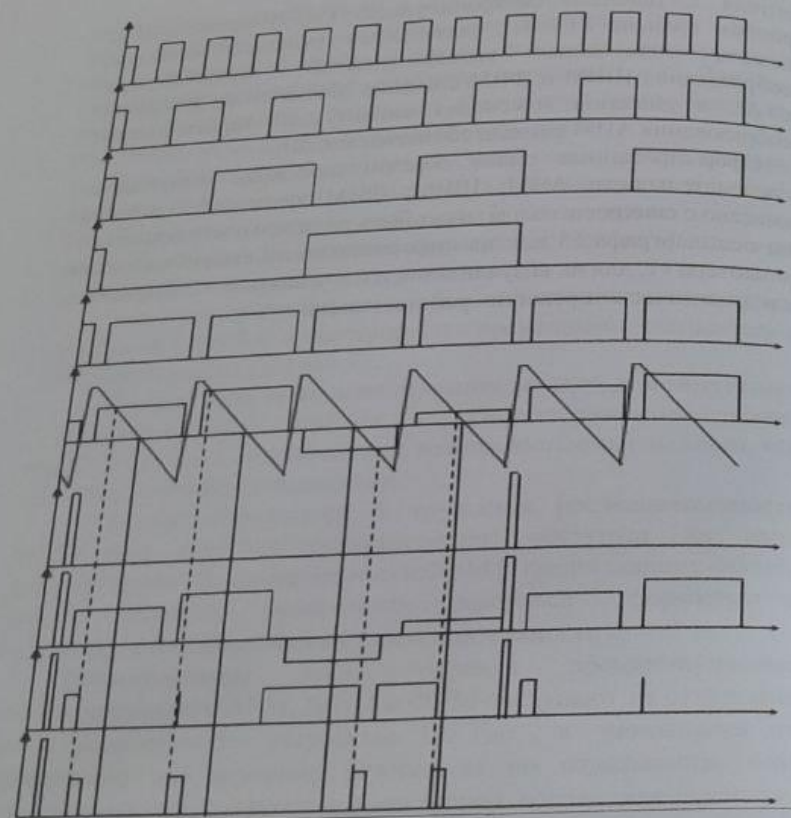


Рис. 7.2. Временная диаграмма.

## 7.2. Основные положения системы связи с ВРК

### Принцип распределения каналов

В системах с временным разделением каналов общий тракт связи предоставляется поочередно каждому абоненту на время  $T_k$ , называемое каналным интервалом. Каждый канал подключается к тракту, периодически, с периодом  $T_i$ , и посылает в групповой тракт свой каналный сигнал (КС). При наличии  $N$  каналов в группе справедливо, что  $T_k \leq T_i/N$ , и чем больше число каналов в группе ( $N$ ), тем короче длительность КС, т.е. тем меньше время отведенное для обработки каждого из сигналов.

Таким образом в системах ВРК передача осуществляется циклами или периодически, группами из  $N$  различных каналных сигналов (КС). Длительность цикла  $T_i$  включает в себя помимо  $N$  каналных интервалов  $T_k$  и интервалы вспомогательных сигналов, например, цикловой синхронизации  $T_{цс}$ , а также интервал служебной связи  $T_{сс}$ .

### Виды преобразования сигналов в системах ВРК

Сигналы в системах с ВРК подвергаются преобразованиям с целью их подготовки для ввода в канал через соответствующий линейный (или каналный) обработчик). Различают следующие основные виды преобразования:

а) дискретизация — замена непрерывного сигнала  $S(t)$  последовательностью дискретных отсчетов его мгновенных значений;

б) импульсная модуляция — формирование импульсных каналных сигналов КС, несущих информацию об отсчетах  $S_k(t)$ . Эта операция названа первой ступенью модуляции.

в) уплотнение во времени всех КС, несущих информацию или размещение на групповом временном интервале  $NT_k$  группового импульсно-аналогового (временнo-импульсного) сигнала  $U_{гр}(t)$ .

Далее в канальном обработчике, как правило, и в основном для передачи в эфир, этим сигналом модулируется высокочастотная несущая.

В приемнике производятся обратные преобразования.

г) выделение  $U_{гр}(t)$  из принятого радиосигнала.

д) разделение сигнала  $U_{гр}(t)$  на отдельные каналные сигналы;

е) преобразование каждого КС для восстановления соответствующего отсчета  $S_k(t)$ .

ж) интерполяция передаваемых сигналов по последовательности их отсчетов  $S_k(t)$ .

В некоторых случаях последние две операции могут объединяться.

Интерполяция (пункт ж), как наиболее ответственная и сложная операция восстановления сигнала по его отсчетам  $S_k(t)$  имеет более глубокие корни предыстории научных споров и сомнений, которые были в достаточной мере разрешены с появлением в 1933 году доказательства теоремы отсчетов или теоремы В.А. Котельникова. Данная теорема дает обоснование выбора значения частоты дискретизации сигнала с ограниченным спектром: сигнал  $S(t)$  с ограниченным спектром полностью определяется через мгновенные отсчеты (значения) взятые через интервал времени  $T \leq 1/2F_v$ . При этом  $S(t)$  для любого  $t$  определяется рядом, учитывающим взаимодействие отдельных гармоник сигнала отсчетов появляющихся в спектре:

$$S(t) = \sum_{k=-\infty}^{\infty} F(kT) \frac{\sin 2\pi F_v(t - kT)}{2\pi F_v(t - kT)}$$

В зарубежной литературе данную теорему называют теоремой отсчетов или теоремой выборки, а частоту  $F_i = 2F_v = 1/T_i$  - частотой Найквиста.

Руководствуясь положением этой теоремы можно оптимизировать систему связи по частоте дискретизации и тем самым предать ей наилучшую экономичность при требуемой эффективности, например, в вопросе быстродействия. Так, исходя именно из этой теоремы на практике, если для стандартного телефонного канала  $F_v = 3,4$  кГц и  $F_i \geq 2F_v = 6,8$  кГц. Однако, с целью облегчения реализации интерполятора (ФНЧ) и для повышения точности самой интерполяции, в современных системах связи принято

$F_i = 8$  кГц,  $T_i = 125$  мкс.. Эти значения рекомендованы МККР для всех международных линий связи с временным уплотнением.

### 7.3. Уплотнение и модуляция

При передаче сигналов  $N$  числа каналов по каналу связи, импульсы всех каналов равномерно распределяют внутри тактового периода. Для этого необходимо чтобы тактовые частоты всех каналов были равны и строго синхронны, а между ними должны быть постоянные фазовые сдвиги, равные  $3600/N$ , что соответствует временному интервалу или групповому периоду  $T_i/N$ .

При передаче 6-ти каналов временной интервал между импульсами равен  $125/6 = 20,83$  мкс.

Из сказанного следует, что в линии связи, где применяется временное уплотнение, частота повторения импульсов группового сигнала или групповая частота равна:

$$F_{\text{групп}} = NF_i$$

Например, для передачи 6-ти-х каналов  $F_{\text{групп}}$  равна:

$$F_{\text{групп}} = 6 \cdot 8 = 48 \text{ кГц.}$$

На рисунке 7.3 показаны отдельные последовательности немодулированных импульсов 6-ти каналов соответственно сдвинутые по фазе, а также групповой сигнал всех шести каналов таким, каким он подается в линию связи после сложения сигналов всех каналов.

Если бы импульсы всех каналов были одинаковыми, на приёмной стороне было бы невозможно узнать какой импульс несёт информацию какого канала и, распределительное устройство не смогло бы правильно распределять импульсные сигналы соответствующим корреспондентам. В связи с этим аппаратура временного уплотнения должна выделить импульсы одного из каналов по какому-то отличительному признаку, однозначно определяемому на приемной стороне. Такой импульс называется маркерным импульсом или импульсом синхронизации (СИ). Синхроимпульсом может быть любой из передаваемых импульсов, но всегда заранее оговоренный; синхронизирующий канал (СК) служит исходным для отсчета номеров каналов в распределительном устройстве аппаратуры разделения каналов на приемном конце линии связи. Обычно в группе маркерным каналом служит первый канал. Маркерный импульс передается один раз за тактовый период. Это пример наиболее простого вида маркерного



импульса, применяемого во многих радиорелейных станциях. Маркерный импульс при этом называется «широтный» так как отличается от других импульсов своей длительностью (рис. 7.3). В нашем случае маркерный импульс отличается от канальных амплитудой.

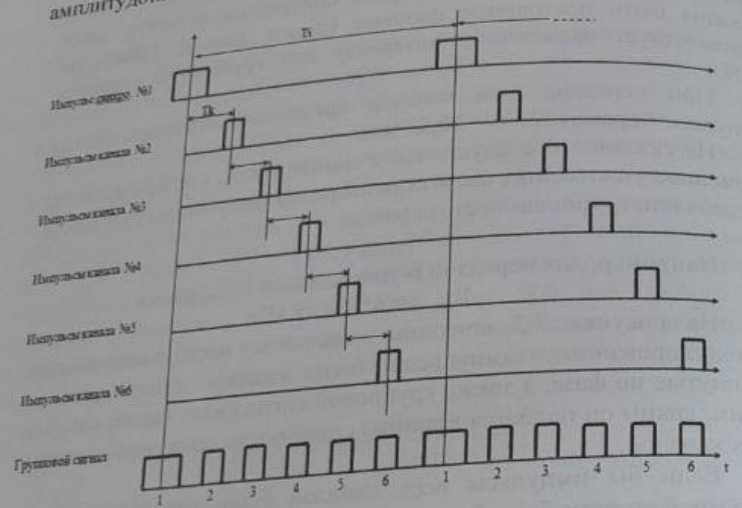


Рис. 7.3. Образование группового сигнала при ВРК.

**Фазовая стабильность и защитный интервал между каналами.**

Распространена и эффективно используется малоканальная связь с фазоимпульсной модуляцией (ФИМ), т.к. это наиболее устойчивый вид импульсной модуляции наряду с ИКМ и дельта модуляцией.

ФИМ, также позволяет построить очень компактную и не дорогую аппаратуру временного уплотнения. Но ФИМ обладает некоторым недостатком. При этом виде многоканальной импульсной модуляции каждый канальный импульс во время модуляции занимает значительную часть временного интервала между импульсами других каналов. От этого зависит

помехоустойчивость ФИМ. Поэтому, если не принимать меры при построении аппаратуры, импульсы одного канала во время модуляции могут заходить в область, выделенную для импульсов другого канала, в результате чего могут возникать сильные взаимные помехи между каналами. Кроме того, станет невозможным на приемном конце надежное разделение импульсов. Для каждого отдельного канала необходимо выделить интервал времени  $\Delta t_k$ , называемый канальным интервалом, за пределы которого импульсы данного канала не должны выходить при любых значениях модулирующего сигнала. Между канальными интервалами необходимо оставить защитный интервал  $\Delta t_z$ , необходимый для обеспечения надёжного выделения импульсов различных каналов на приёмной стороне. Защитный интервал нужен также по техническим причинам. Существует большое количество внешних факторов, стремящихся сдвинуть импульсы со своих номинальных фаз.

**Спектр импульсов, модулированных по амплитуде**

На рисунке 7.4 а показана последовательность прямоугольных импульсов одного из каналов, модулированных по амплитуде сигналов низкой частоты F.

Как известно, последовательность немодулированных импульсов (рис. 7.4 б) может быть разложена в ряд Фурье вида:

$$A_i(t) = \frac{A_0}{q} \left[ 1 + 2 \sum_{n=1}^{\infty} \frac{\sin x}{x} \cos n \Omega t \right],$$

где  $x = \frac{n \Omega t \tau}{2} = \frac{n \tau}{q}$

и  $q = \frac{T_i}{\tau}$  - скважность этих импульсов.

На основе этого уравнения можно построить амплитудный спектр синусоидальных гармонических составляющих, из суммы которых состоит вышеуказанная последовательность. Этот спектр показан на рис 7.5.а. Огибающая этого спектра имеет вид функции  $\sin(x)/x$ , которая равна нулю в точках, где  $F=k/\tau$  (k -любое целое число). Из этого следует, что основная часть спектра сосредоточен

в области частот  $\Delta F = 1/\tau$ . Из этого также следует, что скважность импульсов  $q$  численно равна количеству гармоник тактовой частоты  $F_0$ , находящихся внутри полосы частот  $\Delta F$ .

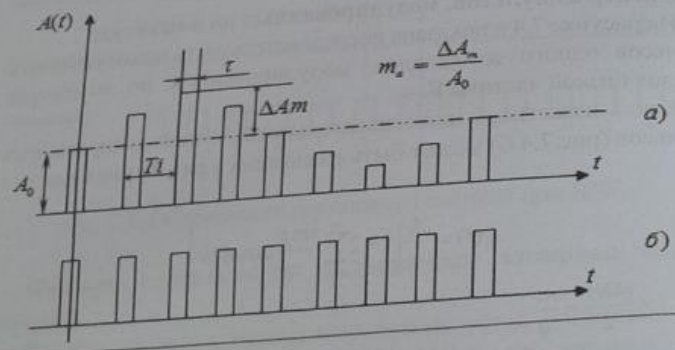
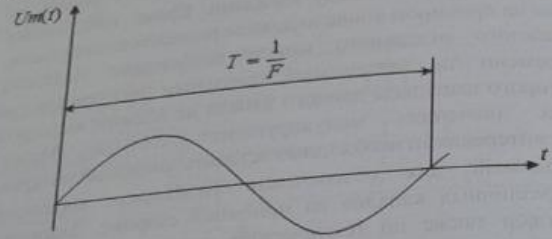


Рис. 7.4. Последовательность прямоугольных импульсов одного канала, модулированных по амплитуде.

Так как длительность обычных рабочих импульсов в системах с ВРК приблизительно равна или меньше одной микросекунды, то  $q \geq 100$ . Поэтому амплитуды первых нескольких составляющих кратных тактовой частоте спектра (рис 7.5 а), практически равна

между собой и равны  $2A_0/q$ , а постоянная составляющая будет в 2 раза меньше (рис. 7.5 б).

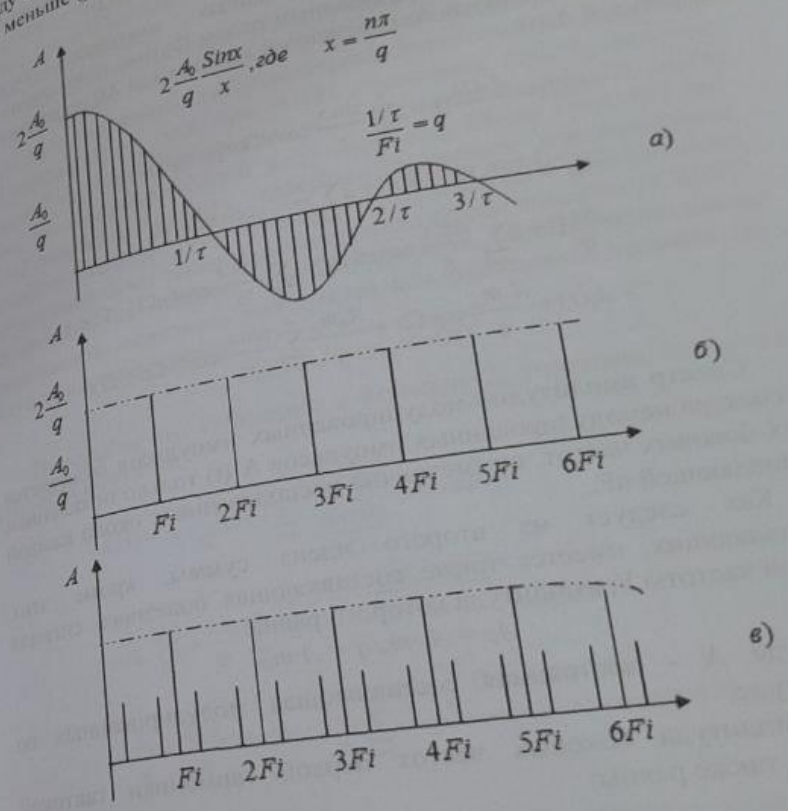


Рис. 7.5. Спектр импульсов.

Если при отсутствии модуляции амплитуда всех импульсов была постоянной и равно  $A_0$ , то при модуляции импульсов по амплитуде синусоидальным сигналом с частотой  $F$  и с