

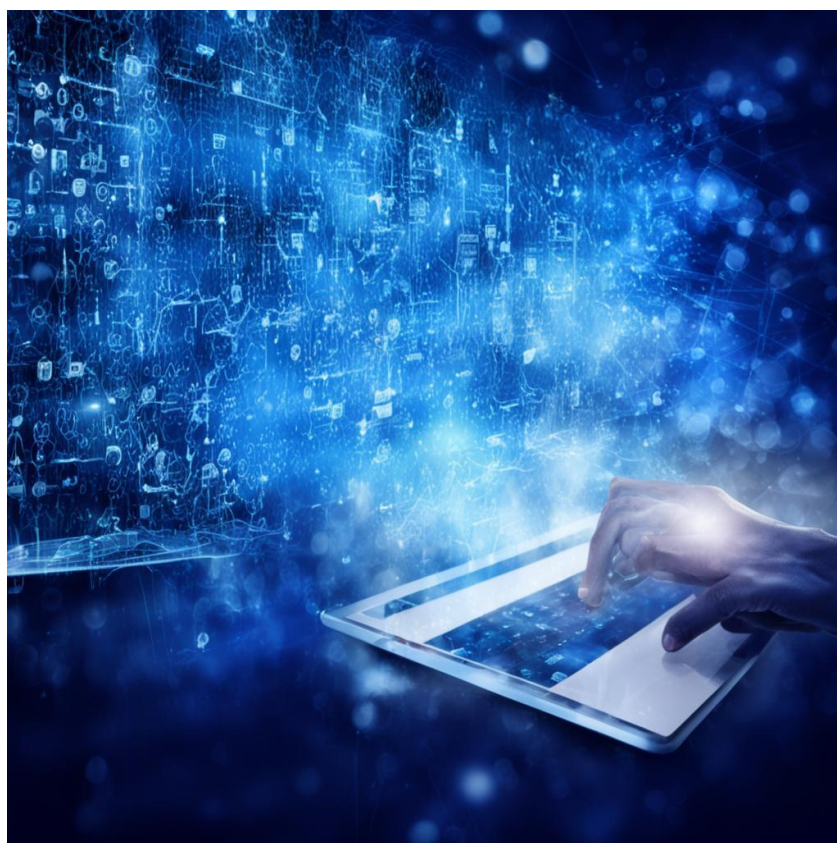
**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЛЕСОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ С. М. КИРОВА**

М.Р. ВАГИЗОВ, М.А. НОВИКОВА, С.А. ИВАНОВ, Н.В. МЕЛАМЕД

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ

ЧАСТЬ 3

УЧЕБНОЕ ПОСОБИЕ



Министерство науки и высшего образования РФ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Санкт-Петербургский государственный лесотехнический
университет имени С. М. Кирова»

М. Р. ВАГИЗОВ, М. А. НОВИКОВА, С. А. ИВАНОВ, Н. В. МЕЛАМЕД

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ

ЧАСТЬ 3

Учебное пособие
Для студентов высших учебных заведений

Санкт-Петербург
2023

УДК 004.9
ББК 16я73
В23

Рассмотрено и рекомендовано к изданию кафедрой Информационных систем и технологий
Института леса и природопользования Санкт-Петербургского государственного
лесотехнического университета имени С. М. Кирова

Авторы:

Вагизов Марсель Равильевич — кандидат технических наук, доцент, зав. кафедры информационных систем и технологий, Институт леса и природопользования, ФГБОУ ВО Санкт-Петербургский государственный лесотехнический университет имени С. М. Кирова.

Новикова Мария Александровна — кандидат сельскохозяйственных наук, доцент кафедры информационных систем и технологий, Институт леса и природопользования, ФГБОУ ВО Санкт-Петербургский государственный лесотехнический университет имени С. М. Кирова.

Иванов Сергей Александрович — кандидат технических наук, доцент, доцент кафедры информационных систем и технологий, Институт леса и природопользования, ФГБОУ ВО Санкт-Петербургский государственный лесотехнический университет имени С. М. Кирова.

Меламед Наталья Владимировна — кандидат технических наук, доцент кафедры информационных систем и технологий, Институт леса и природопользования, ФГБОУ ВО Санкт-Петербургский государственный лесотехнический университет имени С. М. Кирова.

Рецензент:

Бурлов Вячеслав Георгиевич — доктор технических наук, профессор, профессор кафедры информационных технологий и систем безопасности, Институт информационных систем и геотехнологий, ФГБОУ ВО Российский государственный гидрометеорологический университет.

Вагизов, М. Р.

В12 Информационные технологии и системы : учеб. пособие для студентов высш. учеб. зав. Ч. 3 / М. Р. Вагизов, М. А. Новикова, С. А. Иванов, Н. В. Меламед ; Минобрнауки России. ФГБОУ ВО «СПбГЛТУ им. С. М. Кирова». — Санкт-Петербург, 2023. — 109 с.

ISBN 978-5-00125-897-1

В учебном пособии приведены базовые понятия в области информации, информатизации, информационных технологий и систем. Рассматриваются программные комплексы и среды, базы данных, аппаратно-программное обеспечение, языки программирования, информационные ресурсы и геоинформационные системы. Для студентов, бакалавров, магистрантов, специалистов, аспирантов, преподавателей, изучающих и интересующихся вопросами подготовки информационных технологий.

УДК 004.9
ББК 16я73

© М. Р. Вагизов, М. А. Новикова,
С. А. Иванов, Н. В. Меламед, 2023
© СПбГЛТУ, 2023

ISBN 978-5-00125-897-1

СОДЕРЖАНИЕ

ГЛАВА 1. ПРОЦЕССОРЫ.....	4
ГЛАВА 2. МОБИЛЬНЫЕ ОПЕРАЦИОННЫЕ СИСТЕМЫ.....	16
ГЛАВА 3. ТЕХНОЛОГИЯ БЛОК-ЧЕЙН.....	30
ГЛАВА 4. СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ.....	41
ГЛАВА 5. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ.....	66
СПИСОК ЛИТЕРАТУРЫ.....	106

ГЛАВА 1. ПРОЦЕССОРЫ

Центральный процессор (ЦПУ) (*от англ. central processing unit, CPU*) - устройство, составная часть аппаратного обеспечения современной вычислительной техники, является главным элементом производящим вычислительные операции, реализует код поступающий от программ.

На сегодняшний день сложно представить себе современную вычислительную технику без упоминания такого важного её компонента как вычислительный процессор. Процессоры входят в состав большинства устройств микро и обычной электроники, от высокопроизводительных компьютеров, до калькуляторов, бытовой и автомобильной техники. По характеристикам процессоров определяют производительность персональных компьютеров и некоторых мобильных устройств. Процессоры прошли относительно небольшой исторический путь от момента их появления и до сегодняшнего времени, этот путь характеризуется сложными инженерно-техническими и научными достижениями, знание которых, позволит вам приблизиться к пониманию принципов их работы. Основная информация данного раздела учебного пособия будет сведена сначала к устройству процессоров, затем к историческому срезу нескольких этапов развития процессорных элементов вычислительной техники, в заключение данного раздела будут приведены в сравнении некоторые наиболее значимые линейки процессоров с их характеристиками.

Среди характеристик процессоров выделяют:

1) производительность - количество операций выполняемых за секунду, измеряется в флопсах (FLOPS).

2) тактовая частота - активный сигнал или синхросигнал, частота выполняющая синхронизацию поступающих импульсов на электронную схему. Такт - промежуток между двумя импульсами генерации электрических импульсов. В современных процессорах тактовая частота может изменяться в зависимости от нагрузки на вычислительный процесс.

3) техпроцесс - технологический процесс по производству полупроводников, основного элемента в процессоре. Техпроцесс отражает реальный масштаб производимых процессоров на фотолитографическом оборудовании. Определяется в микро и нанометрах, указывает о плотности стояния транзисторов на кристалле, чем меньше техпроцесс, тем плотность стояния транзисторов выше друг к другу. Например

процессоры Intel 8086 1979 года производства имели техпроцес 3 мкм, а процессоры Эльбрус-8С 2015 года производства уже 32 нм.

4) архитектура процессора - структура системы адресации команд и способа их вычисления. Архитектуры процессоров принято выделять на определенные семейства близкие по технологическому способу их производства, например архитектура процессоров AMD Athlon отличается от архитектур процессоров семейства Intel Core или Байкал. Однако и те и другие производят вычислительные операции. Так же существуют архитектуры процессоров с программной точки зрения, такие как RISC (ARM) или CISC (x86)

Внутреннее устройство процессоров. Основным элементом любого процессора выступает транзистор - прибор регулирующий электросигнал. Историю развития вычислительной техники вы можете найти в первой части цикла учебных пособий авторов [1]. Генерацию электрического импульса и пропуск тока в транзисторе можно проследить на рис. 1 (а, б, в), где а) затвор закрыт подачи тока нет, б) затвор открыт на половину, частичная подача тока в) затвор открыт, ток проходит. Стоит отметить, что разработке первого транзистора предшествовала 10 летняя работа физиков.

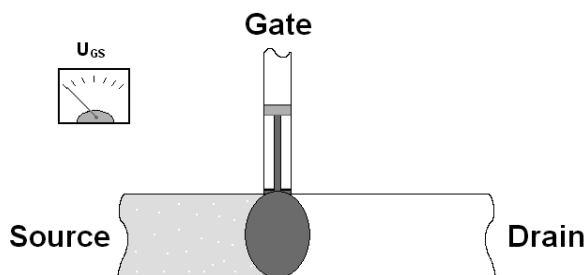


Рис.1 а - затвор закрыт, подачи тока нет.

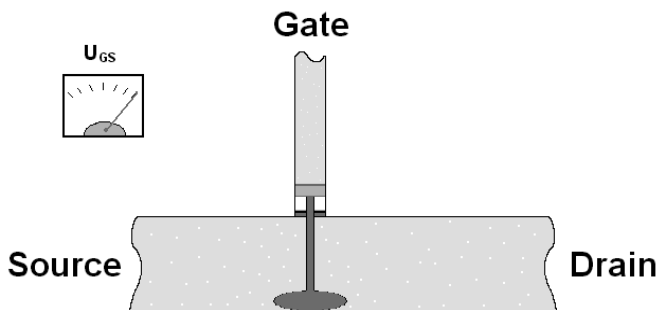


Рис.1 б - затвор открыт, ток проходит.

Количеством транзисторов размещенных на кристалле, определяет их производительность и сложность производства, например в первых процессорах количество транзисторов определялось сотнями, позднее тысячами штук. В современных процессорах на 2023 год, количество транзисторов на кристалле достигает несколько миллиардов штук (рис.2). Однако для размещения такого большого числа транзисторов требуется специальное высокотехнологическое оборудование в виде литографов, которыми обладают ограниченное количество компаний. Стандартное размещение процессора на материнской плате персонального компьютера представлено на (Рис. 1.1)

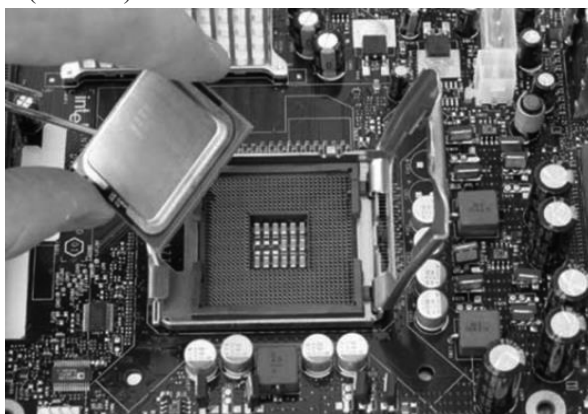


Рис. 1.1 Размещение процессора на материнской плате.

В 1960-х годах получила свое развитие технология изготовления интегральных схем, что позволило создавать микрочипы с расположенными на них транзисторами. Сам процессор состоял из

нескольких таких схем. Такие элементы именовали кристаллами, которые выполнены из полупроводникового материала, обеспечивающего необходимую электропроводность, в качестве химического основного химического элемента используется кремний, реже германий. С течением времени технологии позволили размещать все большее количество транзисторов на кристалле, в связи с чем количество используемых в CPU интегральных схем сокращалось [4].



Рис. 2 Снимок кристалла процессора Intel Kaby Lake

Хронологический анализ развития процессорной техники. Для наглядного понимания скорости развития, увеличения технических характеристик и введения модификаций в устройства процессоров. Можно провести срез развития техпроцессов (табл.1). Стоит отметить, что сокращение мкм и нм, расшифровывается как, мкм - микрометр, равен 0,001 миллиметру, нм - нанометр, равен 10^{-9} метра, 1000 нм равно 1 мкм. В таблице 1 представлены только некоторые разработанные разными компаниями процессоры, так же в списке, сокращено количество существующих техпроцессов. Разнообразие процессоров определяется их сложностью, производством для конкретных задач и типов устройств. Разные компании за различные периоды, производили уникальные модификации и технологические решения. Физически процессор размещается на материнской плате в специальном разъёме, который обозначается как сокет, отличаются сокеты количеством контактов, способом расположения и размерами, наиболее точные характеристики и историю развития сокетов вы можете найти во второй части учебного пособия авторов [3].

Таблица 1 Развитие техпроцессов.

Размер	Процессор	Год	Частота процессора
20 мкм	RCA CD4000	1968	1 Гц - 2 МГц
10 мкм	Intel 4004	1971	740 КГц-750 КГц
6 мкм	Intel 8080	1976	2-4 МГц
3 мкм	Intel 8086	1979	4-16 МГц
1 мкм	Intel 80286	1982	4-25 МГц
0,8 мкм	Intel 80486	1989	16-150 МГц
0,6/0,5 мкм	Power PC 601	1994	66 МГц
350 нм	Pentium II, AMD K5	1997	233-450 МГц
250 нм	AMD K6, Pentium III	1999	1000 - 1400 МГц
180 нм	Pentium IV, AMD Athlon XP	2000, 2001	1,3 - 3,8 ГГц
130 нм	Intel Xeon	2002	400 МГц - 4,4 ГГц
90 нм	AMD 64 X2 Эльбрус (2010)	2006	1,6 МГц - 2,4 ГГц
65 нм	Intel Core 2 Duo, AMD Athlon 64 Xbox 260 Jasper	2006 2007	2,2 - 3,33 ГГц
28 нм	Intel Sandy Bridge AMD Bulldozer Snapdragon	2011	3 - 4,4 ГГц
16 нм	AMD Ryzen Эльбрус-16С Intel Coffe Lake	2017 2021 2017	3,1 - 4,2 ГГц
12 нм	Apple A11 Intel Ice Lake	2017, 2019	2,39 ГГц
6 нм	Apple A14	2020	3,1 ГГц
3 нм	Apple A17 Pro	2023	3,78 ГГц
2 нм	Samsung 2GAP	2024	-
1,4 нм	-	2027	-

На производительность процессоров влияет не только их размер, но и архитектура по которой они произведены, а также технологические особенности инженерных решений, такие как, например многоядерность при их производстве. Далее будут рассмотрены некоторые типы архитектур процессоров. В 1971 году вышел процессор Intel 4004, особенность процессора Intel 4004 состоит в том, что это первый в истории доступный микропроцессор, который при стоимости 200 долларов реализовывал на одном кристалле функции процессора большой ЭВМ. Среднее количество операций в секунду составляло порядка 60 000 тысяч.

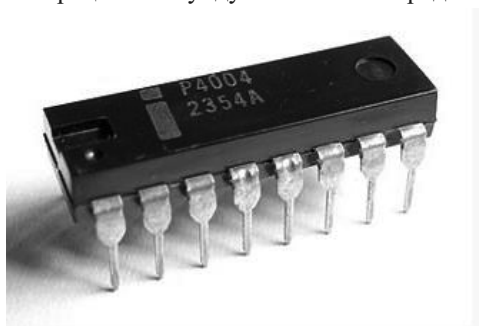


Рис. 3 Процессор Intel 4004

Производство процессоров. Для производства процессоров используются специальное фотолитографическое оборудование или литографы. Которые посредством лазера различной длины волны формируют микроскопические структуры на поверхности кремниевых чипов и других полупроводниковых материалов.



Рис. 3 Работа фотолитографической машины

Основной принцип фотолитографии заключается в передаче узора с маски на поверхность кремния или другого материала с использованием светочувствительного резиста. Сначала поверхность материала покрывается фоточувствительным резистом. Затем через маску проецируется ультрафиолетовое (UV) излучение на покрытую поверхность. Там, где свет достигает резиста через маску, происходит химическая или физическая реакция, приводящая к изменению свойств резиста. После экспозиции проводится процесс развития, который удаляет либо вымывает часть резиста, создавая узор, соответствующий узору на маске (Рис.4). Для производства процессоров по наименьшим стандартам в 10 и менее нанометров используется экстремальная ультрафиолетовая литография, длина проецирующей волны составляет 13,5 нм.

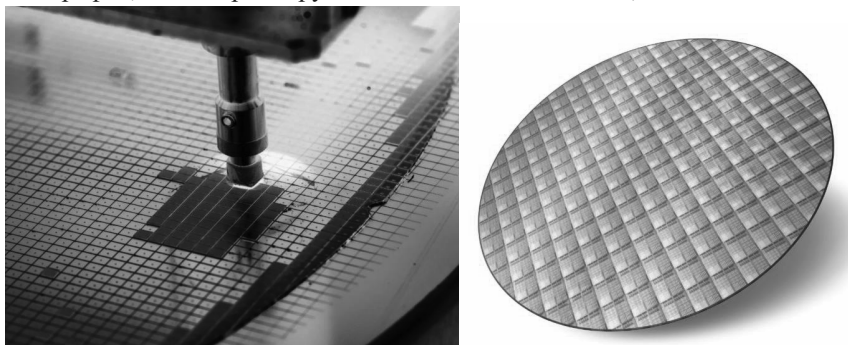


Рис.5 Процесс нанесения маски на кремниевую пластину.

Аппаратная архитектура процессора - качественные свойства процессора, базирующиеся на применение определённой технологии в ряде линейки (семейства) процессоров.

Программная архитектура процессора - совместимость с определённым набором команд, структурой команд (системы адресации, набора регистров) и способа исполнения (счётчика команд).

Все интегральные схемы делятся на две большие категории: RISC (Reduced Instruction Set Computer «компьютер с сокращённым набором команд») и CISC (Complex Instruction Set Computer «компьютер с полным набором команд»). Вторые работают со сложными инструкциями, а первые — с упрощёнными.

Если вы читаете учебное пособие со смартфона, то для вас работает процессор на архитектуре ARM, а если с ноутбука, для вас трудится чип на архитектуре x86. ARM процессоры используют упрощённый набор

инструкций — RISC. ARM - чаще всего используются в мобильной технике. x86 процессоры - используют сложный набор инструкций, используется в персональных компьютерах и ноутбуках. Процессоры x86 потребляют намного больше энергии, чем ARM. Но на потребление энергии также влияет размер самого транзистора.

RISC — архитектура процессора, в котором быстродействие увеличивается за счёт упрощения инструкций: их декодирование становится более простым, а время выполнения — минимальным [4].

CISC — тип процессорной архитектуры, в первую очередь, с нефиксированной длиной команд, а также с кодированием арифметических действий в одной команде и небольшим числом регистров, многие из которых выполняют строго определенную функцию. Различия между архитектурами можно посмотреть на схеме (Рис.4).

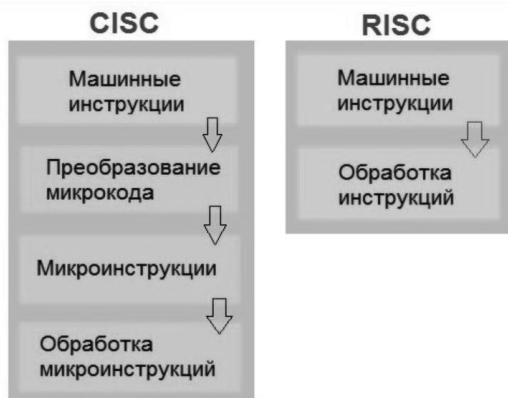


Рис.4 Архитектура CISC и RISC

Так же существуют другие типы архитектур используемые в различных устройствах.

MISC (англ. Minimal Instruction Set Computer — «компьютер с минимальным набором команд»). Элементная база состоит из двух частей, которые либо выполнены в отдельных корпусах, либо объединены. Более простая архитектура, используемая в первую очередь для ещё большего уменьшения итоговой цены и энергопотребления процессора. Используется в IoT-сегменте и недорогих компьютерах, например, роутерах.

VLIW (англ. Very Long Instruction Word — «очень длинная машинная команда») — архитектура процессоров с несколькими

вычислительными устройствами. Характеризуется тем, что одна инструкция процессора содержит несколько операций, которые должны выполняться параллельно.

VLIW-архитектура представляет интерес для полупроводниковой промышленности по двум причинам. Первая — теперь на кристалле больше места может быть отведено для блоков обработки, а не для блока предсказания переходов. Вторая причина — VLIW-процессор может быть высокоскоростным, так как предельная скорость обработки определяется только внутренними особенностями самих функциональных модулей. VLIW при определенных условиях может реализовать старые CISC-инструкции эффективнее RISC.

Технологические решения в процессорах. Многоядерные процессоры – Содержат несколько процессорных ядер в одном корпусе (на одном или нескольких кристаллах). – Процессоры, предназначенные для работы одной копии операционной системы на нескольких ядрах, представляют собой высокоинтегрированную реализацию мультипроцессорности (Рис.5).

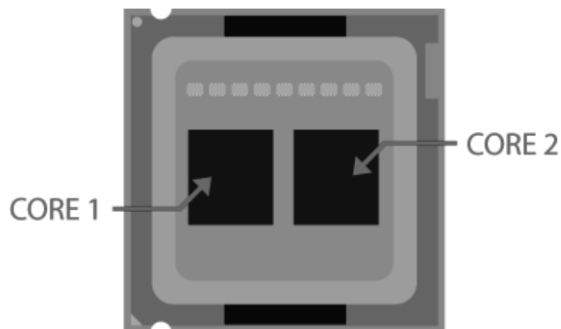


Рис. 5 Двухъядерный процессор

Первым процессором, предназначенным для массового использования, а не для встроенных систем, стал POWER4 с двумя ядрами PowerPC на одном кристалле, выпущенный компанией IBM в 2001 году. Позднее в 2005 году AMD выпустила 2-ядерный процессор Opteron архитектуры AMD64, предназначенный для серверов. 14 ноября 2005 года компания Sun microsystems выпустила восьмиядерный UltraSPARC T1, каждое ядро которого выполняло 4 потока. В 2006 году Intel представила двухъядерный процессор на одном кристалле Core Duo, для мобильной

платформы. Практика реализации многоядерных процессоров росла, массовое производство четырёхядерных процессоров произошло в 2010 году, позднее процессоры становились 8 ядерными, позднее 16 ядерными. В августе 2011 года компанией AMD были выпущены первые 16-ядерные серийные серверные процессоры Opteron серии 6200. В феврале 2020 года компания AMD выпустила в продажу первый 64-ядерный процессор для домашних компьютеров AMD Ryzen Threadripper 3990X. Так же необходимо отметить что, существовали экспериментальные проекты 80 и 100 ядерных процессоров, для применения их в суперкомпьютерах.

Технология Hyper Threading - это технология, разработанная компанией Intel, которая позволяет одному физическому процессору эмулировать наличие двух логических процессоров для повышения производительности. Пока одна инструкция завершит выполнение, процессор начинает выполнять другую инструкцию на втором "виртуальном" процессоре внутри себя. Когда одно вычисление задерживается, процессор переключается на другое, что позволяет использовать пространственный и временной параллелизм для более эффективного использования ресурсов процессора (Рис.6).

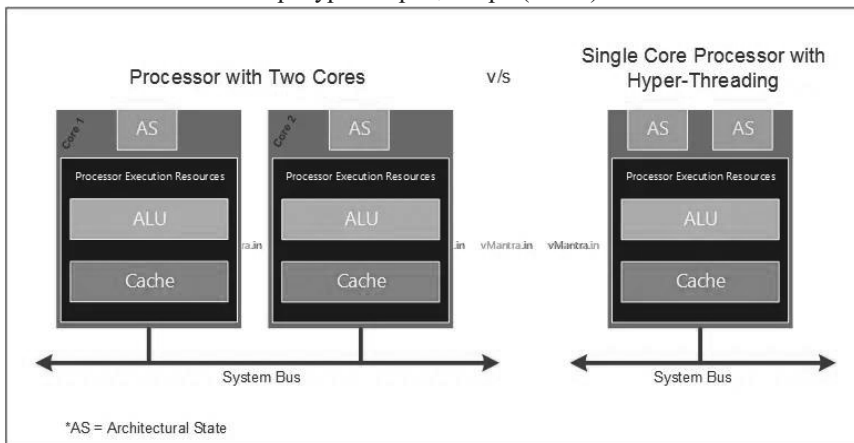


Рис. 6. Технология Hyper-Threading и двухядерный процессор

Преимуществами технологии считаются:

- возможность запуска нескольких потоков одновременно (многопоточный код);
- уменьшение времени отклика;
- увеличение числа пользователей, обслуживаемых сервером;

- уменьшение площади кристалла.

Технология Intel Sandy Bridge. Sandy Bridge (можно перевести как «песочный мост») — микроархитектура центральных процессоров, разработанная фирмой Intel. Основана на 32-нм технологическом процессе, содержит встроенный видеоускоритель [5]. Sandy Bridge принесла с собой значительное улучшение в графическом интегрированном ядре, которое было встроено в процессоры. Это означало, что пользователи могли получить приличное качество графики даже без установки отдельной видеокарты.

Структуру чипа Sandy Bridge можно условно разделить на следующие основные элементы:

- процессорные ядра;
- графическое ядро;
- кэш-память L3
- «Системный агент».

Все перечисленные элементы объединены с помощью 256-битной межкомпонентной кольцевой шины, выполненной на основе новой версии технологии QPI. Одним из процессоров построенных на технологии Sandy Bridge является процессор Intel Core i7 2700K с тактовой частотой 3,5 ГГц. Усовершенствования в управлении кэш-памятью также считались одним из ключевых компонентов этой архитектуры, что сказалось на общей производительности процессоров.

Стоит отметить, что существует большое количество различных технологий процессоров, разные компании ежегодно развивали, модернизировали и производили уникальные технологические решения в области микроэлектроники, производства полупроводников и архитектурных решений в производстве процессорной техники. Особым подходом при производстве процессоров является формирование многоядерной архитектуры сочетающей в себе CPU и GPU. Например, компания Apple разработала новый тип процессоров на ARM-архитектуре на 5-нанометровов техпроцессе, под наименованием Apple M1 Ultra в котором содержится 20 ядер ARM, 48 графических ядра и 32-ядра встроенного ИИ-ускорителя. Что означает, что в структуре архитектуры процессора присутствуют специальные блоки для ускорения его работы, которые основанные на системах искусственного интеллекта, одним из первых таких процессоров являлся процессор от компании AMD - Ryzen 7040HS.

Особенности закона Мура. В 1965 году один из основателей компании Intel Гордон Мур сформулировал закономерность в которой утверждается, что количество транзисторов размещаемых на микропроцессорах, будет удваиваться каждые 18-24 месяца. Не смотря на физические пределы различных элементов используемых при производстве полупроводников, на данный момент, с учётом различных инженерно-технологических решений, данная закономерность выполняется.

В заключении данного раздела можно отметить, что процессоры и их развитие оказывает колоссальное влияние на все сферы человеческой деятельности, в прикладном аспекте, данные элементы используются как в повседневных устройствах, смартфонах, так и в специализированных устройствах, такие как системы управления космических кораблей и суперкомпьютеры. Однако нельзя не отметить возможную будущую технологическую смену, полупроводниковых процессоров на квантовые технологии и квантовые компьютеры, принцип работы которых и вычислительные способности многократно превышают процессоры на кристалле.

Глава 2. Мобильные операционные системы

Мобильные операционные системы в современном мире существует огромное разнообразие мобильных устройств. Мобильный — значит переносимый, т.е. сюда мы относим не только сотовые телефоны, но и планшеты, плееры и прочие устройства

Самый первый мобильный телефон был выпущен в 1983 году (имеется в виду именно коммерческий аппарат, появившийся в продаже). Это был Motorola Dyna TAC 8000X.С 1980 года, на рынке были коммуникаторы **Radio Shack Pocket Computer TRS – 80**.

Данный коммуникатор работал на интерпретаторе языка BASIC, он позволял создавать свои простые программы без участия персонального компьютера.

В начале 1991 года, появились первые PDA. PDA расшифровывается как Personal Digital Assistant, PDA — «личный цифровой секретарь». В России они именовались КПК – карманный персональный компьютер. Первым представителем PDA были коммуникатор HP-95LX (рис.1)



Рисунок 1. HP-95LX

В 1992 году была создана операционная система Newton OS. Программа для неё была написана на C++. Также это был первый сенсорный телефон

В октябре 1993 года был создан Zoomer. Компьютеры Zoomer работали под управлением объектно-ориентированной, графической, многозадачной и многопоточной версии операционной системы GEOS. Данная система работала до 100 часов при смешанном режиме работы – одна минута ввода текста на 10 минут пассивной работы. Но были и

недостатки такие, как не удобная система рукописного ввода, слабый процессор, отсутствие средств обмена данными с компьютером и достаточно высокая цена.

В 1996 году появился коммуникатор Nokia 9000, который работал на новой операционной системе GEOSTM 3.0. Данный КПК не имел сенсорного ввода, но был одним из первых представителей класса смартфонов.

Также 1996 год можно назвать основополагающим годом для развития мобильных операционных систем. Так как в этом же году появилась система Windows CE, которая долгие годы была одной из основных операционных систем для огромного количества карманных компьютеров и даже смартфонов. Первыми устройствами, выпущенными на этой системе были КПК Palm стали PalmPilot 1000 (рис.2) и PalmPilot 5000.

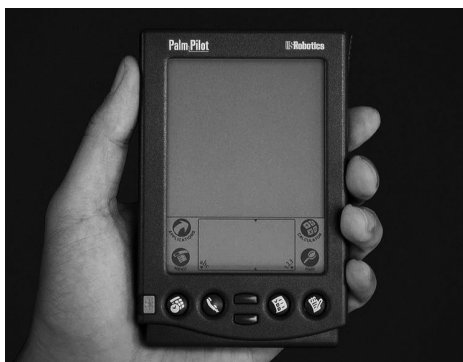


Рисунок 2. КПК PalmPilot 1000

На операционной системе Windows CE выходило большое количество устройств вплоть до 2000 года. Систему в своих устройствах использовало большое количество производителей: Nokia, Sony, Casio, Sharp, Siemens и другие.

В 2000 году появилась система Windows CE 3.0 Pocket PC Edition. Эта система уже являлась полноценной операционной системой для карманных компьютеров и также устанавливалась на огромное количество устройств. В ней была реализована поддержка цветных экранов, а также поддержка карт памяти (Compact Flash и Multimedia Card).

В 2003 году Windows CE разделила свою операционную систему на три направления для различных устройств: с сенсорным и не сенсорным

вводом, классические карманные персональные компьютеры (Pocket PC for Pocket PC, Windows Mobile Classic), версия для коммуникаторов с сенсорным дисплеем (Pocket PC Phone Edition, Windows Mobile Professional) и версия для несенсорных устройств с классической телефонной клавиатурой (Pocket PC for Smartphone, Windows Mobile Standard).

В 2000 году появилась еще одна мобильная операционная система Symbian. Ее разработчиком была компания Ericsson. И первым устройством, выпущенным на рынок с ОС Symbian, была Ericsson R380 (Рис.3).



Рисунок 3. Ericsson R380

Позже Nokia взяв на вооружение эту операционную систему, начали выпускать одно за другим различные карманные компьютеры и смартфоны. Первым смартфоном Nokia вышедшим на Symbian в 2002 г был Nokia 7650

В 2002 году, появилась операционная система BlackBerry OS. Данная система не была массовой, так как изначально позиционировалась для корпоративного сегмента, и ставила упор на работу в безопасной среде и внутри корпоративных систем.

В 2007 году появилась первая версии IOS и в этот же год появляется первый Iphone. Позже каждый год эта система совершенствовалась и каждый год выходили новые версии вплоть до 2021 года

В 2008 году появляется первая версия операционной системы Android. Данная операционная система только набирала обороты и поклонников.

2010 год. На рынок выпущена ОС Windows Phone. Первым аппаратом, выпущенным на этой операционной системой, был HTC 7 Mozart (Рис.4)



Рисунок 4. HTC 7 Mozart

В 2011 году Samsung совместно с Intel выпустили свою операционную систему под названием Tizen. И выпустили на рынок несколько своих аппаратов под этой системой. Но самым примечательным был Samsung Z1, который массово продавался на множестве рынков.

С 2015 года основная масса операционных систем это Android и IOS.

Далее рассмотрим подробнее самые распространенные мобильные операционные системы.

Современные операционные системы для мобильных устройств: Android, Kai OS, Lineage OS, Fire OS, Flyme OS, iOS, Sailfish OS, Tizen, MIUI, RemixOS, FuchsiaOs, Windows Phone, Аврапа, а также Sailfish.

Операционная система Android.

Android — операционная система, разработанная компанией Google для мобильных устройств. Она является самой распространенной операционной системой в мире и используется на множестве смартфонов и планшетов.

Основная цель разработки Android — предоставить пользователям удобную и мощную операционную систему, которая обеспечивает широкие возможности для запуска приложений и доступа к интернету. Android поддерживает множество функций, включая мультизадачность, уведомления, синхронизацию с облачными сервисами и многое другое.

Одно из главных преимуществ Android — его открытость и гибкость. Это означает, что разработчики могут создавать приложения для Android на различных языках программирования и распространять их через Google Play Store. Также пользователи могут настраивать свои устройства под свои нужды, устанавливая приложения, изменяя интерфейс и многое другое.(28)

Android является самой распространенной операционной системой для мобильных устройств, благодаря своей функциональности, гибкости и широкой поддержке от производителей и разработчиков.

Положительные качества Android:

1. Любой пользователь может создать свое приложение, благодаря открытому исходному коду. Приложение основано на Linux
2. Существует абсолютное равноправие всех приложений, как встроенных, так и установленных. Выбор программы осуществляется по умолчанию простым нажатием кнопки.
3. Операционная система многозадачна
4. Личная информация в безопасности

Недостатки Android:

1. Много актуальных версий, но новая версия либо выходит поздно, либо не выходит вовсе. В связи с этим разработчики создают приложения, ориентируясь на более старые версии
2. Возможны хакерские атаки из-за открытости кода
3. Высокий расход батареи
4. Требуется доработки

Операционная система Lineage

Lineage OS – крупнейшая кастомная прошивка для Android. Она безопасна, конфиденциальна и не требует использования Google сервисов (5), была создана в далеком 2016 году в качестве стандартного пользовательского ПЗУ для Android (30).

Операционная система интегрирует и улучшает функции Android, и предлагает:

- Полную поддержку тетеринга, Wi-Fi и Bluetooth
- Поддержку аудиокодека FLAC
- Список расширенных имен точек доступа (APN), что представляет собой настройки телефона для подключения SIM-карты к Интернету
- Повышение производительности, включая разгон процессора
- Управление интерфейсом и доступом к приложениям

Lineage OS поддерживает почти 200 устройств. (29) Функции Lineage OS:

1. Воспроизведение аудио

2. Использование жестов. Жесты — это то, что резонирует со всем брендом OnePlus. Жесты позволяют выполнять определенные задачи из удобства вне экрана. Например, чтобы включить фонарик нарисовать букву V.

Операционная система iOS

Разработана специально для устройств iPhone и iPad.

Главная особенность данной системы заключается в том, что она разработана и может использоваться только на телефонах и планшетах фирмы Apple.

Преимущества iOS:

1. Высокий уровень безопасности. Реализована многофакторная система аутентификации, защита конфиденциальных данных пользователей.
2. Оптимизация и быстроедействие.
3. Синхронизация с другими устройствами Apple. Возможность обмениваться данными между различными фирменными гаджетами: смартфонами, телеприставками, компьютерами и т.д.
4. Пользовательский интерфейс.
5. Высокая автономность. Батарея достаточно долго может функционировать без дополнительной подзарядки.

Недостатки iOS:

1. Закрытость системы. Пользователь не имеет доступа к системным папкам операционной системы, а значит, ограничен в своих возможностях настроить ее под свои предпочтения.
2. Слабая многозадачность. Большинство приложений в фоновом режиме быстро прекращают работать.
3. Невозможность расширить память. Объем памяти строго ограничен и его невозможно расширить с помощью карты памяти (32).

Операционная система Kai

Kai OS – это операционная система, разработанная для телефонов с физическими клавиатурами. Телефоны на этой ОС – это как обычный кнопочный телефон с возможностями смартфона.(33)

Kai OS поддерживает 4G LTE, VoLTE (голосовые звонки через LTE), GPS, Wi-Fi и может работать на устройствах с ограниченным

объемом памяти. Включает в себя такие функции, как доступ к интернету, удобный интерфейс, поддержка 4G-связи и многие другие.

Преимущества Kai OS:

1. Доступность и низкая стоимость. Телефоны на Kai OS представляют собой более доступную альтернативу смартфонам. Они способны предоставить базовые функции связи, а также использовать приложения, предназначенные для Kai OS.

2. Длительное время работы от одной зарядки. Телефоны на KaiOS обычно имеют более длительное время работы от одной зарядки, поскольку они потребляют меньше энергии, чем смартфоны на других операционных системах.

3. Простой интерфейс. Телефоны на Kai OS имеют простой и интуитивно понятный интерфейс, что делает их легкими в использовании даже для пользователей, не знакомых с современными смартфонами

4. Более долгий срок службы. Благодаря своей простоте и отсутствию большого количества сложных компонентов, телефоны на Kai OS обычно имеют более длительный срок службы, чем смартфоны.(34)

Недостатки Kai OS:

1. Ограниченные возможности. Некоторые функции, такие как многозадачность или сложные игры, могут быть недоступны или иметь ограниченную поддержку.

2. Ограниченный выбор приложений. Такие популярные приложения или игры, как WhatsApp, Instagram или Fortnite, могут быть недоступны для установки на телефоне на Kai OS, так как операционная система все еще не поддерживает их полностью.

3. Ограниченная коммуникационная функциональность. Несмотря на то, что телефоны на Kai OS обеспечивают базовые функции связи, такие как звонки и SMS, они могут быть ограничены в некоторых функциях, например, в видеочатах или передаче медиафайлов.

4. Ограниченные обновления и поддержка. Проблемы с обновлениями и ограниченной поддержкой производителей могут возникнуть, поскольку операционная система Kai OS все еще находится в развитии и может не получать обновления так часто, как другие операционные системы.(33)

Операционная система Fire

Fire OS — мобильная операционная система с закрытым исходным кодом, основанная на базе Android, созданная Amazon для собственных смартфонов Fire Phone, электронных книг Kindle Fire, планшетов, иных устройств типа Amazon Fire (37).

Из особенностей этой операционной системы можно выделить:

- Максимальная интеграция со всеми сервисами компании Amazon.
- Возможность подключать мобильные устройства к телевизорам или игровой приставке Sony
- Наличие специальной кнопки, предоставляющей технической службе полноценный доступ к устройству, для решения тех или иных проблем.

Кроме того, устройства с Fire OS считаются очень эргономичными, так как специальные инструменты позволяют очень чутко регулировать энергопотребление смартфонами или телефонами (38).

Плюсы:

- Дешевый. Подходит для детей
- Работает как умный динамик на беспроводной зарядной док-станции. Позволяет настраивать индивидуальные профили для всех в семье.
- Много премиум-контента.
- Родительский контроль
- Многопользовательская учетная запись

Минусы:

- Медленный
- Тусклый, посредственный экран
- Нельзя установить приложения Google без вспомогательных программ (9).
- Невозможно расширить память

Система Flyme OS

Flyme OS — операционная система на базе Android для смартфонов и планшетов с открытым исходным кодом. Представляет собой сильно модифицированную систему Android, разрабатываемую и поддерживаемую компанией Meizu Technology Co., Ltd.. Первая версия, на тот момент ещё оболочки вышла в 2012 году (39).

Обладает сильно изменённым (по сравнению с оригинальной ОС) пользовательским интерфейсом — в котором, в частности, отсутствует меню приложений (39).

Преимущества Flyme OS:

1. Уникальный дизайн и пользовательский интерфейс: Он отличается яркостью и легкостью использования, и при этом предлагает множество настраиваемых опций и тем.

2. Богатый набор функций: позволяет пользователям получить больше возможностей в использовании своего телефона. Эти функции включают в себя различные жесты и движения, темы, анимацию, управление звуком и многое другое

3. Хорошая оптимизация и быстродействие.

4. Большой выбор настроек и персонализация: пользователи могут настроить различные аспекты интерфейса, такие как темы, обои, значки приложений и т.д. Также имеется возможность настройки различных системных уведомлений, звуков, жестов и движений.

5. Удобное управление и навигация: на устройствах с Flyme можно использовать жесты вместо кнопок навигации, что облегчает использование телефона одной рукой

6. Безопасность: устройства с Flyme имеют встроенный антивирус, который регулярно обновляется и проверяет установленные приложения на наличие угроз.

7. Регулярные обновления и поддержка: это позволяет пользователям быть в курсе последних разработок и получать новые возможности на своих устройствах.

8. Совместимость с другими устройствами: поддерживает множество стандартов и протоколов, таких как Miracast, DLNA, AirPlay, Bluetooth, Wi-Fi и другие (40).

Операционная система Windows Phone

Windows Phone — мобильная операционная система, разработанная Microsoft, вышла 11 октября 2010 года.

Преимущества Windows Phone

1. Плавность интерфейса. Перемещаться по спискам, меню и плиткам очень удобно и легко.

2. Небольшой «вес» приложений. Размер большинства программ под Windows Phone – 1-2 Мб. Это очень важно, если нужно срочно что-то скачать и установить.

3. Наличие Zune – специальной программы для синхронизации музыки и видео между компьютером и телефоном.

4. Резервное копирование всех файлов на компьютер.

5. Красиво оформляется текст

6. . регулярные обновления

Недостатки Windows Phone 8

1. Небольшое число предустановленных приложений.

2. Закрытость интерфейса. Поменять нельзя практически ничего, разве что только цвет интерфейса.

3. Отсутствие поддержки Flash в браузере (41).

Операционная система Symbian

Операционная система для различных мобильных устройств, которую начали разрабатывать в далеком 1998-ом году. О серьезном значении этого продукта в свое время говорит тот факт, что над его доведением до ума трудились такие известные на весь мир компании, как: Motorola, Ericsson, Nokia, Siemens, Sony, Samsung и другие.

Фактически большая часть первых поколений смартфонов работали на основе операционной системы Symbian. В2006-ом году почти 73% всей мобильной техники использовали эту операционную систему, но в начале 2013-ого года было принято решение полностью прекратить выпуск устройств, работающих на базе Symbian (42).

Преимущества:

1. Высокая степень настраиваемости. Пользователи могут настраивать интерфейс и функциональность своих устройств по своему усмотрению с помощью множества тем, виджетов и приложений (43).

2. Широкие возможности для разработчиков приложений. Множество средств разработки позволяют создавать приложения быстро и эффективно, но кроме этого, большое сообщество разработчиков могут обмениваться опытом и помогать друг другу создавать качественные продукты (43).

3. Высокая степень совместимости с различными устройствами.

Недостатки:

1. Устаревший интерфейс.

2. Ограниченная поддержка приложений. По сравнению с другими системами меньше разнообразие приложений доступных для скачивания.

3. Относительно низкая производительность. Операционная система может работать медленнее на старых устройствах или иметь проблемы с запуском требовательных приложений и игр.

4. Ограниченная поддержка. Как результат, пользователи могут испытывать затруднения с получением обновлений и исправлений для операционной системы, что может повлиять на безопасность и функциональность устройства (43).

Операционная система Cyanogen Mod.

Cyanogen Mod – своеобразный тюнинг стандартной системы Android, который исправляет некоторые недостатки, делает ее более эффективной, безопасной и интересной в эксплуатации (42). **Cyanogen Mod** - бесплатная операционная система для смартфонов и планшетов с открытым исходным кодом, основанном на ОС Android. Предназначена для замены проприетарных версий Android, предустановливаемых поставщиками мобильных устройств. Открытый исходный код и отсутствие пакета предустановленных сервисов Google обеспечивают безопасность и конфиденциальность (44).

Преимущества:

1. Быстрый цикл обновления.

2. Отсутствуют «мусорные» приложения, есть масса настроек, но нет перегруженности функциями и приложениями. Иконки приложений только что установленной прошивки занимают чуть более половины одной страницы в меню, и среди них только самое важное: камера, галерея, браузер, файловый менеджер и прочее (45).

3. Защищенный режим или AppGuard. Позволяет сделать две вещи: запретить приложениям доступ к тем или иным данным или датчикам (отозвать полномочия) либо включить своего рода конфиденциальный режим, когда при запросе личных данных пользователя (список контактов, местоположение, сведения о владельце и так далее) приложение получает случайным образом сгенерированную информацию (45).

4. Черный список. Блокиратор номеров реализован на уровне системы, поэтому он никогда не дает сбоев, не пропускает рингтоны и в

целом работает идеально. Кроме явно заданных номеров, он позволяет блокировать номера с помощью регулярных выражений, блокировать скрытые и неизвестные номера (45).

Недостатки:

1. многие пункты в меню обходятся без нужной локализации
2. ярлыки различаются размерами, компактный стиль меню отображается некорректно
3. дата или время сбиваются (46)

Операционная система BlackBerry OS

Телефоны, выпускаемые под торговой маркой BlackBerry, предназначены для серьезных людей, которым подобное устройство необходимо, прежде всего, для работы или подтверждения собственного статуса.

Среди многообразия различных возможностей BlackBerry OS следует отметить два очень интересных и важных. Первый – подобная система считается одной из самых защищенных в мире, даже более распространенным аналогом до ее функционала очень далеко в этом смысле. Кроме того, эта операционная система позволяет использовать на телефонах, смартфонах и фирменных планшетах приложения, выпущенные под ОС Android, а также игры для Sony Playstation. Для этой цели используется специально встроенный конвертор, преобразующий файлы в нужный формат (42). В 2017 году Роскомнадзором были заблокированы сервисы BlackBerry в России. Связано это с тем, что компания отказалась предоставлять данные о своих пользователях(48).

Преимущества:

1. Интерфейс очень удобен. В особенности это проявляется при просмотре различных электронных документов.
2. Легкий в управлении.
3. Удобна для работы с документами.
4. Удобна для использования электронной почтой.
5. очень надёжна и работает стабильно

Недостатки:

1. мультимедиа-функции слабо развиты
2. плохо поддерживает функции мессенджеров (47).
3. Система подходит для бизнес-смартфона.

Операционная система «Аврора»

«Аврора» — российская мобильная операционная система с открытым исходным кодом на базе ядра Linux, разработанная в 2016 году компанией «Открытая мобильная платформа». Включена в Единый реестр российского программного обеспечения Минцифры, зарегистрирована в Роспатенте, имеет сертификаты ФСТЭК и ФСБ России. На данный момент используется и государственными и корпоративными заказчиками (49).



Рисунок 5. ОС «Аврора»

ОС «Аврора» имеет встроенные средства верификации загрузки и запуска программного обеспечения, а также политики безопасности. Дистанционно осуществляется тотальный контроль и мониторинг всех функций мобильного устройства.

Преимущества:

- повышение уровня безопасности,
- снижение расходов на администрирование корпоративных мобильных устройств,
- поддержка работы до 500 000 мобильных устройств одновременно,
- защищённость корпоративных данных,
- повышение уровня мотивации сотрудников,
- эксплуатация платформы в закрытом контуре без взаимодействия со сторонними приложениями и сервисами,
- работа с группами пользователей и динамическими группами устройств через Microsoft Active Directory,
- внесение обновлений на устройства,
- проведение диагностик и аудитов,
- в случае потери гаджета удаление данных и блокировка устройства в удалённом режиме,

- интеграция с LDAP-каталогом,
- управление правами доступа (49).

Недостатки:

- Непривычное управление. Если Android ориентирован больше на тапы, то в «Авроре» в ходу свайпы — нажимать на экран приходится редко.
- Посредственная интеграция приложений друг с другом. На Android, скажем, снимок из галереи можно отправить куда угодно — хоть в мессенджер, хоть в «облако».
- Отсутствие центра управления — интерфейс не позволяет быстро запустить Wi-Fi или Bluetooth (50).

Встроенные механизмы безопасности не позволяют установить ОС в любое устройство, но некоторые модели можно приобрести уже со встроенной ОС «Аврора». Все изделия предназначены для офисного и корпоративного использования:

- КПК Aquarius NS M11,
- планшет Aquarius CMP NS220 v5.2,
- смартфон Масштаб TrustPhone T1,
- смартфон F+ R570,
- планшет БайтЭрг MBK-2020,
- планшет F+ Life Tab Plus,
- КПК Aquarius NS M12,
- смартфон Qtech QMP-M1-N IP68,
- планшет Aquarius CMP NS220,
- смартфон Qtech QMP-M1-N,
- планшет Aquarius Cmp NS208.

ГЛАВА 3. ТЕХНОЛОГИЯ БЛОК-ЧЕЙН

Технология *blockchain*

Блокчейн – это программный продукт, который позволяет хранить и преобразовывать величины или данные при помощи интернета защищенным и прозрачным способом, не имея при этом центрального управляющего органа.

Блокчейн (англ. *blockchain*, изначально *block chain* — цепь из блоков) — выстроенная по определённым правилам непрерывная последовательная цепочка блоков (связный список), содержащих информацию. Блокчейн — это технология баз данных, которая лежит в основе надежного хранения и обмена ценностями в сети: криптовалютой, произведениями искусства и другими цифровыми активами. Блокчейн позволяет без посредников заключать любые сделки в сети.

Например, можно найти в блокчейне: суммы, даты или участники сделки, заверенные специальными электронными подписями.

Существуют два типа блокчейна:

1. открытый, доступный каждому с личного компьютера и не поддающийся контролю;

2. закрытый, зависящий от администратора.

Но на данный момент появились несколько версий классификации блокчейнов: канадская, британская и российская. Рассмотрим их подробнее:

Канадская версия основывается на видении создателя блокчейн-платформы Ethereum канадца Vitaly Buterin. Его классификация предполагает наличие 3 видов блокчейна:

- **Public blockchain** (публичный блокчейн) — это цепочка блоков, которую может «прочитать» любой человек в мире. Таким образом любой человек может отправлять транзакции, ожидать их включения, если они действительны, и участвовать в процессе консенсуса, т.е. процессе для определения, какие блоки добавляются в цепочку и каково текущее состояние сети.

- **Consortium blockchain** (блокчейн консорциума) — это блокчейн, в котором процесс согласования контролируется заранее выбранным набором узлов. Право на чтение блокчейна может быть общедоступным или ограниченным для участников.

- Fully private blockchain (полностью приватный блокчейн) — это блокчейн, характеризующийся ограниченным уровнем доступа к данным. Подтверждение транзакций в таких сетях, проведение аудита, управление базами доступно четко определенному кругу лиц. Если говорить о праве на чтение данных, то оно может быть как общедоступным, так и полностью ограниченным.

Британская версия основывается на докладе главного научного советника правительства Великобритании Mark Walport. В своем докладе Distributed Ledger Technology: beyond block chain по распределенным реестрам и потенциалу блокчейна в сфере государственного управления он разделил блокчейн на 3 вида:

1. Unpermissioned public ledgers — открытые публичные реестры.
2. Permissioned public ledgers — закрытые публичные реестры.
3. Permissioned private ledgers — закрытые частные реестры.

Данная классификация идентична той, которую приводил Vitaly Buterin, и в ней аналогом Public Blockchain в британской версии является Unpermissioned public ledgers, аналогом Consortium Blockchain является Permissioned public ledgers, а аналогом Fully private Blockchain выступает Permissioned private ledgers. Более того, в докладе был предложен небольшой тест «Классификация распределенных реестров», позволяющий самостоятельно определять, к какому виду относится тот или иной блокчейн.

Российская версия основывается на видении заместителя председателя Центробанка РФ Ольги Скоробогатовой, которая классифицировала блокчейны на 3 вида: открытые, закрытые и гибридные. Так, выступая с докладом «Инновационные технологии и подходы к их применению» на конференции «Блокчейн и открытые платформы — 2016», Ольга Скоробогатова отметила: «Общепризнано на данный момент два вида систем. Это открытая и закрытая сети. В открытой сети участники не проходят реальной идентификации, допуск к сети не ограничен, статус процессоров не закреплен за конкретными участниками, и нет никаких видимых инстанций, управляющих правилами. Если говорить о закрытой сети, то она организована практически с точностью до наоборот: участники сети идентифицированы, допуск к участию в сети ограничен и у него есть правила, статус процессоров закреплен за определенными контрагентами и обычно есть некий орган, который эти правила устанавливает, регламентирует, меняет. Эти два вида — открытые

и закрытые сети. Через три месяца исследований стало понятно, что есть третий вид систем — гибридные. Гибридная сеть — это сеть, которая может сочетать часть элементов из открытой сети и часть элементов из закрытой сети».

Опираясь на вышеизложенное, можно утверждать, что видение специалистов из разных стран относительно классификации blockchain очень похоже, и в основу такой классификации положена степень открытости блокчейна для его участников. Так, блокчейны можно разделить на две большие группы — публичные и приватные. Степень публичности (открытости) блокчейна может зависеть от нескольких факторов. Первый — доступность исходного кода протокола блокчейна. Исходные коды публичных блокчейнов, как правило, открыты с самого начала разработки, и их изменения ведутся на одном из наиболее популярных интернет-ресурсов, например, на GitHub. Блокчейны и платформы для корпоративного применения (приватные блокчейны) могут быть как с открытым, так и с частично или полностью закрытым исходным кодом.

Если рассматривать блокчейн-разработки с точки зрения их применения в государственном секторе, то и количество таких разработок и продуктов, используемых в реальных процессах, еще слишком мало. При этом следует ожидать, что большинство блокчейнов также будут закрытыми или приватными.

Второй и наиболее важный показатель — это возможность любого пользователя свободно подключиться к сети без получения каких-либо разрешений. Именно это является определяющим отличием публичного блокчейна от приватного. Подавляющее большинство известных на данный момент блокчейнов публичны — для подключения к ним достаточно скачать совместимое с текущей версией протокола клиентское программное приложение и установить связь с другими равноправными узлами сети.

Чтобы полноценно участвовать в работе сети, проверять и ретранслировать транзакции других пользователей или участвовать в создании блоков, необходимо запустить клиентское программное приложение с функциональностью полного узла. В остальных случаях достаточно клиентского программного приложения с ограниченными возможностями. Однако, в публичных блокчейнах уровень участия пользователя всегда определяется самостоятельно и зависит только от его

личных (финансовых или аппаратных) ресурсов. Кроме того, никто не может отключить пользователя от распределенной сети, так как все участники публичного блокчейна равноправны. В отдельных случаях они могут, например, игнорировать или блокировать пользователя, рассылающего некорректные транзакции или пытающегося передать не соответствующую протоколу информацию, но такие инициативы носят исключительно саморегулируемый характер и не устанавливаются на уровне протокола.

В частных блокчейнах за подключение к сети новых пользователей, а также за возможность их отключения могут отвечать выделенные доверенные узлы или группы узлов, имеющие более высокий уровень полномочий по сравнению с остальными пользователями. Частные блокчейны представляют собой иерархические структуры, состоящие из двух или более уровней. Пары ключей, предоставляющие доступ к системе, выдаются и управляются специальными административными узлами и при необходимости могут быть отозваны. Таким образом, частные блокчейны не полностью реализуют основные принципы технологии — децентрализацию и равноправие участников, так как для корпоративных систем их наличие может быть обусловлено существенными рисками.

Следующим критерием, который создает еще одну ступень в классификации блокчейнов, является уровень управления блокчейнами. По данному критерию блокчейны можно разделить на четыре группы:

1. Публичные децентрализованные блокчейны.
2. Публичные блокчейны с делегированным управлением.
3. Частные контролируемые блокчейны.
4. Государственные блокчейны.

Большинство современных публичных блокчейнов имеют одноуровневую структуру. В них все участники равноправны и консенсус достигается путем опосредованного голосования узлов, выполняющих функции создания блока. Публичные децентрализованные сети не накладывают каких-либо ограничений на участие в управлении, и возможности участников определяются только долей имеющихся у них ресурсов от общего количества.

За более чем десятилетнюю историю развития блокчейна можно сделать вывод, что полная децентрализация в саморегулируемых, а точнее стихийно регулируемых, сетях на практике почти невозможна — все

публичные блокчейны рано или поздно сталкиваются с одной из форм централизации. В связи с этим была сделана попытка внедрения элементов централизации для улучшения функций управления и других показателей блокчейна. Это привело к тому, что в 2015 году возникли первые публичные блокчейны с двухуровневой структурой, где ведущую роль играли узлы с расширенными полномочиями. Именно признак наличия двух и более уровней управления в сети блокчейн, с разной степенью полномочий для каждого, является основным для публичных блокчейнов с делегированным управлением.

Приватные контролируемые (корпоративные) блокчейны представляют собой технологические решения для корпоративных потребностей. В таких системах каждый узел имеет предварительно назначенный ему уровень доступа и, в отличие от публичного блокчейна, данные не всегда общедоступны даже для чтения. Управление в таких блокчейнах осуществляется с помощью специальных узлов, обладающих повышенными полномочиями, именно они отвечают за политику распространения данных и идентификацию пользователей, а также удостоверяют внесение данных в блокчейн.

Распределенные реестры для государственного применения в целом незначительно отличаются от корпоративных блокчейнов и также требуют контролируемого доступа к информации. Однако у государственных ведомств есть особые требования к блокчейну — максимальный уровень неизменности уже добавленной информации и самый строгий контроль над ее добавлением. При этом уже имеющаяся в блокчейне информация во многих случаях может быть публичной, поскольку государственные органы должны стремиться к повышению прозрачности своей работы. Так, можно сказать, что государственные блокчейны представляют собой частный случай корпоративных блокчейнов со своими специфическими особенностями, но при этом относятся к отдельной группе.

Классификацию блокчейна, основанную на уровне доступа к информации, можно представить в виде двухуровневой структуры, в которой первый уровень определяет критерий публичности, а второй — уровень управления блокчейна.

Применение *blockchain*

Во многих видах деятельности можно успешно использовать блокчейн, это: банковский сектор, сектор здравоохранения, транспорт и

недвижимость. Блокчейн можно использовать даже для проведения выборов.

В государственном секторе рассматриваются возможности блокчейна как официального реестра для государственных и частных активов, таких как здания, дома, транспортные средства и патенты. Блокчейн способен также упростить процессы голосования, понизить уровень мошенничества и оптимизировать такие функции бэк-офиса, как закупки. Эта технология идеально подходит для сценариев использования в государственном секторе из-за жестких нормативных требований, которые необходимо проверять и подтверждать — и блокчейн устраняет вопросы о надежности таких процессов.

Применения в сфере энергетики и ЖКХ: прямые продажи солнечной энергии между соседями, торговля электроэнергией между энергетическими корпорациями, автоматизированное выставление счетов для автономных заправочных станций для электромобилей и т. д.

Применение в борьбе с подделками. Технология блокчейн может оказаться чрезвычайно эффективной при обнаружении контрафактной продукции. Чаще всего от подобной проблемы страдают производители предметов роскоши, например, модельеры. Потребитель также не всегда уверен в подлинности товара. Аутентификация продукта с помощью блокчейна защищает как бренд, так и покупателя. Это решение использовалось LVMH (французский концерн Louis Vuitton), Microsoft и Consensus для создания платформы Aurora, помогающей индустрии роскоши в мониторинге и отслеживании товаров.

Группа LVMH использует технологию блокчейн почти в 60 своих брендах и планирует расширить функциональность блокчейна за счет защиты интеллектуальной собственности и предотвращения рекламного мошенничества. Многие другие люксовые бренды последовали их примеру, предоставив своим продуктам доказательства подлинности, что позволило их быстро идентифицировать. Таким образом, сфера продаж может продвигать и проверять этическое производство, одновременно доказывая его устойчивость.

Блокчейн-аутентификация также используется в индустрии драгоценных камней. Она была представлена платформой Tracr, предназначенной для алмазной промышленности. Используя блокчейн, платформа позволяет определять подлинность природных алмазов, присваивая им индивидуальные номера. Зарегистрированный алмаз с

таким цифровым номером можно идентифицировать, проверив его происхождение и связанные с ним транзакции.

Применяется как криптовалюта в электронной коммерции. Криптовалюты вызывают множество вопросов, которые имеют решающее значение для достоверности экономических транзакций с использованием таких активов, в частности, в отношении их правовой природы и связанных с этим налоговых вопросов. Тем не менее, они очень удобны для покупателей и владельцев интернет-магазинов: платежи криптовалютой в интернет-магазинах уже становятся распространенным явлением.

Преимущества и недостатки технологии блокчейн

Использование технологии блокчейн в торговле имеет множество преимуществ:

- удовлетворение потребностей современного потребителя;
- обеспечение доступа к достоверным знаниям о продукте;
- неизменность;
- стабильность;
- возможность отслеживать весь производственный процесс;
- обеспечение прозрачности бренда;
- аутентификация продукта;
- содействие укреплению доверия к производителю и лояльности к бренду;
- защита бренда и потребителей от контрафактной продукции;
- использование в логистике.

Блокчейн в IoT

Децентрализация вычислений – неотвратимая тенденция развития интернета вещей (IoT). С другой стороны – блокчейн также основан на принципе децентрализации, поэтому он очень органично вписывается в архитектуру интернета вещей. Комбинация блокчейна и Интернета вещей (IoT) часто рассматривается как один из наиболее перспективных сценариев применения технологии за пределами непосредственно криптовалют. Blockchain и IoT — это технологии, у которых есть большой потенциал и перспектива в будущем. И то, и другое уже широко и с энтузиазмом используется, как в промышленности, так и в государственном секторе. Их объединение может стать ключом к

минимизации рисков безопасности для бизнеса, который всегда идет рука об руку с технологическими новинками.

Основными трендами развития информационных технологий будут следующие:

1. Развитие экосистем смартфонов и носимых устройств

2. Облачные технологии

a) Инфраструктуры вычислений и хранения (IaaS, PaaS), такие как AWS, Azure

b) Облачное ПО для бизнеса (SaaS)

c) Услуги для бизнес-экосистем предоставляемые через стандартные API

3. «Уберизация» всего, бизнес в социальных сетях и социальных платформах.

4. Искусственный интеллект на основе «больших данных» в режиме онлайн.

a) Аналитика искусственного интеллекта: аналитика с использованием машинного обучения

b) Вещи с искусственным интеллектом: автономные самообучающиеся вещи (автомобили, роботы, подключённые домашние приборы, сенсоры)

c) Интерфейсы с ИИ: компьютерное зрение, голосовые интерфейсы, виртуальная и дополненная реальность (VR/AR), цифровые помощники и двойники: Alice, Siri, Google, Alexa

5. IoT/IIoT Интернет вещей и промышленный интернет для Индустрии 4.0: интегрированная цепочка поставок в режиме онлайн, планирование ресурсов предприятий (ERP), операционные и промышленные цифровые помощники (digital twins).

6. Blockchain: цифровые активы и «валюты», деловые взаимодействия без доверия сторон (trustless interactions), «умные контракты» и Децентрализованные Автономные Организации (DAO).

Все эти тенденции взаимосвязаны и взаимообусловлены, и часто случается, что польза одной тенденции проявляется только при наличии другой, как в случае Blockchain и IoT.

Большинство облачных сервисов, в том числе приложений IoT, имеют централизованную систему управления, например, фотохостинг, облачные системы хранения, денежные транзакции (банковская система) и многие другие.

Основным препятствием развития интернета вещей является централизация облачных сервисов. Огромные массивы данных, собираемых с датчиков и сенсоров IoT, приходится:

1. передавать в центральное облако (перегрузка транспортных сетей связи),
2. обрабатывать в централизованном облаке (требуются огромные мощности серверов и систем хранения),
3. необходимо передавать обратно полученные результаты (задержки при управлении онлайн).

Поэтому, появилась новая технология – Fog Computing (т.н. «туманные вычисления»), которая устраняет указанные недостатки.

Таким образом, децентрализация вычислений – неотвратимая тенденция развития интернета вещей.

С другой стороны – блокчейн также основан на принципе децентрализации. Поэтому, он очень органично вписывается в архитектуру интернета вещей.

В условиях значительного роста IoT, появилось несколько критичных проблем, которые необходимо решить, чтобы решения IoT можно было без проблем масштабировать, и адаптировать к быстрому росту подключённых вещей. Кроме того, проблемы безопасности и защиты данных также достаточно остры как в отношении устройств IoT, так и в отношении данных, которые они собирают. Вот некоторые из этих проблем:

1. **Масштабирование:** в централизованных облачных платформах IoT наблюдается очень плотный трафик передачи сообщений. Здесь кроется самое «узкое место» в масштабировании решений IoT для большого числа устройств.

2. **Безопасность:** большой объем данных, собранных от миллионов устройств остро ставит вопрос информационной безопасности как для отдельных пользователей, так и для предприятий и организаций. Как показали происходящие DoS-атаки на устройства IoT, большое число недорогих устройств, подключённых к интернету непосредственно – это главная проблема обеспечения безопасности в IoT.

3. **Недостаточная проработка стандартов для данных и унификация данных:** мир идёт к открытым стандартам данных, но унифицированного подхода пока нет. Есть ряд протоколов, но нет единой платформы для соединения между собой устройств от всех возможных

производителей. Совместимость устройств – это основная задача, которую нужно решить, для повсеместного распространения решений IoT.

4. **Стоимость:** решения IoT как правило предусматривают множество подключённых устройств и сетевого оборудования, их соединяющего, а также аналитических платформ, которые обрабатывают данные IoT. Все это уже стоит довольно дорого и последующий рост будет только увеличивать эту стоимость.

5. **«Архитектура»:** централизованные облачные платформы остаются «узким местом» в комплексных решениях IoT. Любая неисправность или ошибка в этом месте может повлиять на всю сеть.

Проблемы роста IoT:

- Не отлаженные бизнес-модели
- Недостаток функционала
- Отсутствие перспектив
- Недостаток защиты информации
- Высокая стоимость

Преимущества технологии блокчейн для IoT

1. Децентрализация данных.

Информация, собираемая IoT-устройствами, часто хранится и управляется централизованными серверами, что открывает возможность несанкционированной передачи конфиденциальных данных третьим лицам. Кроме того, при возникновении единой точки отказа вся сеть может сразу отключиться.

2. Безопасные обновления оборудования.

Повышенная безопасность и неизменность блокчейна позволяет разработчикам решить проблему устаревшего программного обеспечения IoT, так как они смогут безопасно передавать код на IoT-устройства.

3. Повышенная конфиденциальность.

Блокчейн может защитить и даже скрыть связь между устройствами, обеспечивая валидацию транзакций по передаче данных без привлечения третьих сторон. Кроме того, блокчейн также может оптимизировать протоколы IoT, обеспечивая сквозное шифрование. Это устраняет риски утечки данных и несанкционированного вмешательства в работу IoT-сетей.

4. Улучшенное управление данными.

IoT-сети должны в режиме реального времени передавать большие объемы информации через устройства, системы и платформы, что создает сложности для эффективного управления данными. Блокчейн позволяет устройствам обмениваться информацией напрямую, без участия сервера, облачного хранилища или локальной базы данных, что сокращает количество транзакций как минимум на треть (устройство → устройство вместо устройство → сервер / облако /, локальная сеть → устройство). При этом смарт-контракты могут автоматизировать большую часть подобных взаимодействий между IoT-устройствами.

5. Улучшенная масштабируемость.

Децентрализованные сети блокчейна распределяют рабочие нагрузки на множество узлов, что позволяет обеспечить более быструю обработку транзакций и улучшает координацию между миллиардами подключенных IoT-устройств. Возможность распределенного хранения информации на множестве устройств также способствует повышению масштабируемости.

6. Более строгая аутентификация.

Для идентификации участников большинство блокчейн-платформ используют децентрализованный подход к криптографической аутентификации инфраструктуры открытых ключей (PKI) с генератором закрытых и открытых ключей. Такой подход, в отличие от централизованных PKI, имеет значительно больший уровень безопасности за счет того, что закрытый ключ остается только у пользователя и используется для подтверждения его личности, в то время как открытый ключ передается сетевому провайдеру. И поскольку оба ключа генерируются криптографически, взломать их невероятно сложно.

7. Автоматизация

Предприятия могут использовать подробные показания датчиков устройств IoT в сочетании со смарт-контрактами для автоматизации и оптимизации существующих бизнес-процессов. Например, в промышленных условиях разработчики могут интегрировать смарт-контракты с датчиками IoT для проведения предиктивного обслуживания. IoT-устройства на производственном предприятии могут периодически передавать диагностические данные машин в блокчейн, а смарт-контракты могут быть запрограммированы на отключение отдельных машин, если они отклоняются от рекомендуемых параметров работы.

ГЛАВА 4. СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Теория поддержки принятия решений

Задача принятия решения – это каждодневная деятельность любого человека, специалиста в организации, предприятия. Потребители должны решать, какие продукты покупать, врачи, какие методы лечения назначать, отделы кадров, – каких кандидатов на должность назначать, присяжные, – признавать или оправдывать обвиняемого, инвестиционные организации, – какие проекты финансировать, а законодательные органы – какие законы принимать. Deskриптивная теория принятия решений направлена на то, чтобы дать объяснения и прогноз последствий выбора, который на самом деле делают индивиды и группы, сталкивающиеся с подобными задачами. Нормативная теория принятия решений, с другой стороны, обращается к вопросу о том, какие решения они должны принимать и как они должны их принимать: как они должны оценивать стоящие перед ними альтернативы, какие критерии они должны использовать и каким процедурам они должны следовать.

В более общем случае, если A_1-A_m – это m вариантов, возможных для лица, принимающего решение (ЛПР), S_1-S_n – n возможных состояний мира (они должны быть взаимоисключающими и исчерпывать все возможности), а $C_{11}-C_{mn}$ – $m \times n$ следствий, которые могут вытекать из выбора, то задача решения может быть представлена матрицей состояний-последствий следующего вида (таблица 1):

Таблица 1. Матрица состояний-последствий

	S1	S2	...	Sn
A1	C11	C12	...	C1n
...
Am	Cm1	C21	...	Cmn

Выбор должен быть сделан при столкновении с проблемой принятия решений такого рода, которая будет зависеть от обстоятельств, с которыми сталкивается ЛПР, и, в частности, от объема информации, которой он располагает о различных его характеристиках. Стандартные представления различают условия определенности, когда известно истинное состояние мира и, следовательно, результат действия; риск или неопределенность, когда либо известны вероятности исходов, связанных с

опционом (риском), либо ЛПР может прийти к суждению о том, насколько они вероятны, на основе имеющейся у него информации (неопределенность).

Когда лицо, принимающее решение, знает истинное положение дел в мире, теория принятия решений говорит, что он должен выбрать тот вариант, который он считает лучшим. Когда нет уверенности в действительном состоянии мира, необходимо вынести суждение о том, насколько вероятно, что каждое из возможных состояний действительно имеет место, и выбрать вариант, ожидаемая выгода от которого наибольшая по сравнению с этими вероятностными суждениями.

Современная теория принятия решений в условиях неопределенности уходит своими корнями в дискуссии 18-го века о рисках в азартных играх, когда Даниэль Бернуэлли (Daniel Bernoulli) дал самую раннюю точную формулировку чего-то похожего на принцип максимизации ожидаемой полезности. Первый аксиоматический вывод ожидаемой полезности предпочтений принадлежит Фрэнку Рэмси, чья трактовка во многом превосходит трактовку более поздних авторов.

Но современная теория принятия решений восходит к работам Леонарда Сэвиджа, а не к Рамси, и именно в его книге «The foundations of statistics», где впервые приводится строгий одновременный вывод субъективных вероятностей и полезностей из того, что явно является условиями рациональности предпочтения. Именно Сэвиджу мы обязаны представлением проблемы принятия решений, с которой сталкиваются ЛПР в условиях неопределенности, что в настоящее время является стандартом в теории принятия решений. Сэвидж выделяет три типа объектов: состояния, последствия и действия. Лица, принимающие решения, должны полностью охватывать все возможные факты, которые могут преобладать в ситуации принятия решения, влияющие на исход действий. Последствия, с другой стороны, – это особенности мира, которые имеют значение для лица, принимающего решения. Действия являются связующим звеном между ними, средством, с помощью которого достигаются различные последствия в различных состояниях мира. Формально для Сэвиджа это просто функции от состояний к последствиям.

Хотя это трехчастное различие естественно и полезно, Сэвидж накладывает некоторые довольно строгие условия на эти объекты и отношения между нашими отношениями к ним. Во-первых, состояния

должны быть каузально независимы от действия, в то время как последствия каузально зависят как от действия, так и от состояния мира. Во-вторых, желательность каждого следствия, по мнению Сэвиджа, не зависит от состояния мира, в котором они реализуются, и от наших представлений о них, и наоборот (известный специалист в области теории игр Кеннет Бинмор называет это принципом Эзопа). Оба эти условия должны выполняться, если представление задачи решения с помощью матрицы состояния-следствия, приведенной в таблице 1, может быть преобразовано в матрицу вероятности-полезности.

Особенность представления задач принятия решения с помощью вероятностных матриц полезности требует уточнения. По мнению Сэвиджа, неуверенность ЛПР в том, что делать, полностью проистекает из его неуверенности в том, каково состояние мира. Эта «фундаментальная» неопределенность фиксируется вероятностной функцией состояний мира, измеряющей степень, в которой ЛПР верит в то, что каждое состояние является фактическим. Есть два критических замечания к этому взгляду на неопределенность, которые следует рассмотреть.

Во-первых, модель Сэвиджа, по-видимому, игнорирует другие формы неопределенности и, в частности, неопределенность, которую мы можем испытывать в отношении того, какое значение придавать последствиям, и неопределенность, которую мы можем иметь в отношении того, какие действия доступны. Во-вторых, по-видимому, существует существенная разница между неуверенностью в том, когда кто-то придет, потому что у нас нет точной информации о времени отправления, условиях маршрута, маршруте, по которому он проехал и т.д., и абсолютно непонятным, когда он прибудет, потому что мы не знаем, когда он уехал и уехал ли вообще. Идут ли они пешком или едут на машине и собираются ли они вообще приехать. В первом случае информация, которой располагают, такова, что позволяет определить разумные вероятности прибытия человека в различные промежутки времени. В последнем случае вообще нет никаких оснований для приписывания вероятностей, что является ситуацией радикальной или полной неопределенности. Возможно, мы редко бываем в полной неопределенности, но ситуации частичной неопределенности (или двусмысленности), в которых лицо, принимающее решение, не в состоянии определить вероятности для всех соответствующих случайностей, являются распространенными и важными.

Роль и место систем поддержки принятия решений в процессах компаний

Система поддержки принятия решений (СППР)– это программное обеспечение, которое оказывает поддержку лицам, принимающим решения (ЛПР) в принятии и обосновании решений во многих отраслях.

СППР является компонентом производственных систем и систем бизнес-аналитики и обычно включает в себя базу данных, которая относится к конкретной отрасли. Системы поддержки принятия решений могут также включать в себя файлы, модели компаний, информацию о состоянии организации, данные о продажах, прогнозы, маркетинговые данные, базы знаний и т.д. Такие отрасли, как здравоохранение, сельское хозяйство и маркетинг часто используют преимущества систем поддержки принятия решений. Базовая структура СППР представлена на рисунке 1.

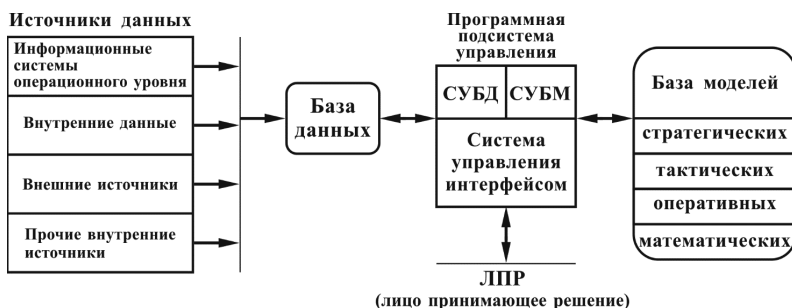


Рисунок 1 – Базовая структура СППР

Можно выделить три общих компонента в системах поддержки принятия решений.

База данных.

Каждая система поддержки принятия решений имеет базу данных, хотя некоторые категории могут иметь более обширные (центральные) базы данных или больше полагаться на них. База данных обычно является основным элементом системы поддержки принятия решений, что делает ее полезной для организации, поскольку программа может исследовать большие хранилища данных при поддержке решений намного быстрее, чем это может сделать один человек или группа. Информация в базе данных зависит от категории системы и отрасли, для которой предназначена система. Некоторые базы данных могут содержать статистику, тогда как другие больше ориентированы на документы.

Например, отдел маркетинга внутри компании может использовать систему поддержки принятия решений для поддержки своих маркетинговых решений. База данных в этой системе, вероятно, содержит внешнюю информацию о текущих тенденциях и рыночных факторах. Она также содержит внутренние знания о предыдущих решениях компании и их результатах.

Пользовательский интерфейс.

Пользовательский интерфейс является точкой доступа для тех, кто использует систему поддержки принятия решений. Эффективные системы поддержки принятия решений используют гибкие и интуитивно понятные пользовательские интерфейсы, которые позволяют специалистам получать доступ к необходимой им информации и управлять системой без обширных технических знаний. Например, пользовательский интерфейс, который использует финансовый эксперт, может иметь четкие указания о том, как выполнять прогнозы и запрашивать модели из системы.

Модели.

Системы поддержки принятия решений также создают модели для поддержки принятия решений с учетом различных параметров, которые влияют на их ситуацию. Модели в системе поддержки принятия решений – это прогнозы или траектории, которые программа определяет как правдоподобные. Эти модели информируют специалистов о том, как их решения могут повлиять на их ситуацию, их клиента или организацию. Система поддержки принятия решений опирается на базу данных для создания точных моделей. Эти модели могут представлять переменные в бизнес-планах организации, действия конкурентов или профессиональные отношения.

Например, компания может принять решение о приобретении производственного предприятия для производства своей продукции. Руководство может принять решение либо о приобретении уже существующего производственного предприятия и произвести его реинжиниринг в соответствии со своими потребностями, либо организовать новое предприятие в соответствии со своими стандартами. Их система поддержки принятия решений может моделировать эти две ситуации и их результаты, чтобы помочь им решить, что лучше.

Модели или инструменты стратегического планирования представляют собой подготовленные структуры планирования, которые помогут в создании планов действий, стратегий и планов проектов. Эти

модели дают специалистам сформулированные методы создания планов определенным образом, но есть возможность гибко модифицировать эти модели в соответствии с потребностями компании. Многие крупные компании внедряют модели планирования, чтобы дать своим менеджерам более согласованную структуру для создания стратегий и целей, что может сделать бизнес-процессы более согласованными между отделами. Существуют различные типы инструментов стратегического планирования, из которых можно выбрать необходимые практически под любые потребности бизнеса, и каждый из них имеет свой собственный способ планирования и подготовки проектов. Место систем поддержки принятия решений в общей структуре предприятия представлено на рисунке 2.



Рисунок 2 – Место СППР в структуре компании

Постановка задачи принятия решения. Неопределенность

Задача принятия решения состоит в формировании множества возможных вариантов, обеспечивающих разрешение проблемной ситуации при существующих ограничениях, и выделении среди этих вариантов одного лучшего или нескольких предпочтительных вариантов, удовлетворяющих предъявляемым к ним требованиям. Формально задачу принятия решения D можно записать в следующем обобщенном виде [9]:

$$D = \langle F, A, X, G, P \rangle,$$

где:

F– формулировка задачи принятия решения, которая включает в себя содержательное описание стоящей проблемы и при необходимости ее модельное представление, определение цели или целей, которые должны быть достигнуты, а также требования к виду окончательного результата;

A – совокупность возможных вариантов (альтернатив), из которых производится выбор. Это могут быть реально существующие варианты, в качестве которых в зависимости от контекста задачи выступают объекты, кандидаты, способы достижения цели, действия, решения и т.п., либо гипотетическое множество всех теоретически возможных вариантов, которое может быть даже бесконечным. Выбор возникает только тогда, когда имеется не менее двух возможных вариантов решения проблемы;

X– совокупность признаков (атрибутов, параметров), описывающих варианты и их отличительные особенности. В качестве признаков выступают, во-первых, объективные показатели, которые характеризуют те или иные свойства, присущие вариантам, и которые, как правило, можно измерить; во-вторых, субъективные оценки, которые обычно даются по специально отобраным или сконструированным критериям, отражающим важные для участников выбора черты вариантов;

G– совокупность условий, ограничивающих область допустимых вариантов решения задачи. Ограничения могут быть описаны как содержательным образом, так и заданы в виде некоторых формальных требований к вариантам и/или их признакам. Например, это могут быть ограничения на значения какого-либо признака или различная степень характерности (выраженности) признака для тех или иных вариантов, или невозможность одновременного сочетания определенных значений признаков для реально существующих вариантов;

P– предпочтения одного или нескольких ЛПР, которые служат основой для оценки и сравнения возможных вариантов решения проблемы, отбора допустимых вариантов и поиска наилучшего или приемлемого варианта. Достаточно часто для упрощения постановки задачи принятия решения часть информации, описывающей предпочтения ЛПР, превращается в ограничения[9].

Большинство научных статей и учебно-методических пособий по поддержке принятия решений и исследованиям операций, вводя формулировку проблемы, которая изучается, начинают с утверждения

типа «дан набор альтернатив А». И исследователи, и практики знают, что в действительности множество А никогда не бывает «данным». На самом деле оно конструируется (синтезируется) в процессе принятия решений и в большинстве случаев определяется несколько раз в течение того же процесса.

Множество альтернатив А может быть фактически трех типов:

– подмножество векторного пространства, где альтернативы описываются как точки (векторы) n-мерного пространства решений (часто каждое измерение связано с «переменной решения»), $A \subseteq R_n$;

– подмножество комбинаторной структуры, где альтернативы описываются как комбинации переменных решения, имеющие конечное и дискретное число возможных значений;

– явное перечисление объектов, возможно, описываемых одним или несколькими признаками или атрибутами.

При принятии решений в условиях абсолютной неопределенности лицо, принимающее решение, не имеет никаких знаний о каких-либо исходах естественных состояний и/или получение необходимой информации обходится дорого. В таких случаях принятие решений зависит только от типа личности лица, принимающего решение.

Анализ принятия решений в целом предполагает, что лицо, принимающее решение, сталкивается с проблемой принятия решения, в которой он или она должны выбрать по крайней мере один вариант из набора вариантов. В некоторых случаях это ограничение можно преодолеть, сформулировав процесс принятия решений в условиях неопределенности как игру с нулевой суммой. При принятии решений в условиях абсолютной неопределенности лицо, принимающее решения, не имеет представления о том, какое естественное состояние «наиболее вероятно» может произойти. ЛПР вероятностно не информированы в отношении состояния природы. Поэтому, ЛПР не может быть оптимистом или пессимистом. В таком случае лицо, принимающее решение, ссылается на соображения безопасности. Необходимо заметить, что любая методика, используемая при принятии решений в условиях абсолютной неопределенности, подходит только для решений в личной жизни. Более того, специалист должен обладать некоторыми знаниями о состоянии природы, чтобы прогнозировать вероятности различных состояний природы. В противном случае лицо, принимающее решение, не способно принять разумное и обоснованное решение в данном случае.

Риск подразумевает некоторую степень неопределенности и неспособность полностью контролировать результаты или последствия такого действия. Однако в некоторых случаях устранение одного риска может привести к увеличению других рисков. Эффективное управление риском требует его оценки и последующего влияния на процесс принятия решения. Процесс принятия решений позволяет лицу, принимающему решение, оценить альтернативные стратегии перед принятием какого-либо решения. Процесс выглядит следующим образом[1]:

1) Определяется проблема и рассматриваются все возможные альтернативы. Оцениваются возможные результаты для каждой альтернативы.

2) Результаты обсуждаются на основе затрат и прибыли по отношению к активам или времени.

3) Различные неопределенности количественно оцениваются в терминах вероятностей.

4) Качество оптимальной стратегии зависит от качества суждений.

Лицо, принимающее решение, должно определить и изучить оптимальную стратегию по отношению к ключевым факторам. Всякий раз, когда лицо, принимающее решение, обладает некоторыми знаниями о естественных состояниях, он/она может назначить субъективные оценки вероятности наступления каждого состояния. В таких случаях проблема классифицируется как принятие решений в условиях риска. Лицо, принимающее решения, может определить вероятности, основанные на наступлении естественных состояний.

Процесс принятия решений в рамках риск-процесса выглядит следующим образом:

а) Необходимо использовать имеющуюся информацию, чтобы соотнести убеждения (называемые субъективными вероятностями) относительно каждого состояния природы, $p(S)$;

б) Каждое действие имеет выигрыш, связанный с каждым из состояний природы $X(a, S)$;

в) Вычислить математическое ожидание, также называемое доходом (R), для каждого действия:

$$R(a) = \sum_{i=1}^n X(a_i, s_i) * p(s_i)$$

г) Определить принцип, согласно которому необходимо минимизировать (или максимизировать) ожидаемую отдачу,

д) выполнить действие, которое минимизирует (или максимизирует) $R(a)$.

Модели рационального и интуитивного принятия решений

Модель принятия решений – это структурированный процесс, который помогает командам принимать решения. Каждая модель, которую может применить лицо, принимающее решение, использует разные методы, чтобы помочь проанализировать и преодолеть конкретную проблему. Поскольку модели используют разные подходы, они полезны для людей с разными стилями обучения или временными ограничениями.

Модели принятия решений могут помочь командам упростить процессы принятия решений и повысить эффективность совместной работы. Модели включают в себя конкретные этапы решения поставленных задач, которым нужно следовать, чтобы синтезировать решения.

Когда все в команде понимают используемую модель принятия решений, им легче внести свой вклад в процесс принятия и вместе прийти к сбалансированному и оптимальному решению.

Можно выделить ряд моделей принятия управленческих решений без поддержки информационных систем, среди которых наиболее популярны модель рационального решения и интуитивная модель принятия решений.

Модель рационального решения.

Модель рационального принятия решений фокусируется на использовании логических шагов для достижения наилучшего возможного решения. Она часто включает в себя анализ нескольких решений одновременно, чтобы выбрать то, которое обеспечивает наилучший результат.

Команды обычно используют модель рационального решения, когда у них есть время для встреч и исследований, что позволяет им составить список потенциальных решений и обсудить плюсы и минусы каждого из них. Можно выделить следующие этапы рациональной модели принятия решений:

1. определение цели и ограничений;
2. определение необходимых для решения поставленной задачи данных: на этом этапе необходимо определить возможность делегирования исследовательских задач членам команды, проведения мозгового штурма во время собрания команды. Необходимо определить,

какая информация о цели или ограничениях имеет отношение к поиску решения;

3. синтез множества вариантов (альтернатив): используя соответствующую информацию, команда может создать список возможных вариантов решений. Необходимо подкрепить альтернативы доказательствами того, почему они помогут достичь цели с учетом наложенных ограничений;

4. ранжирование альтернатив по их значению: после синтеза множества вариантов необходимо упорядочить их по вероятности успешного исхода. Варианты с более высокими шансами на успех также имеют более высокую ценность, в то время как варианты с небольшими доказательствами могут иметь более низкую ценность.

5. необходимо выбрать лучший вариант: определить ценности каждого варианта. Команда должна прийти к консенсусу относительно наилучшего варианта решения, используя собранную информацию. Также на этом этапе прогнозируются все последствия принятого решения.

Подробнее структура рациональной модели принятия решения представлена на рисунке 3.

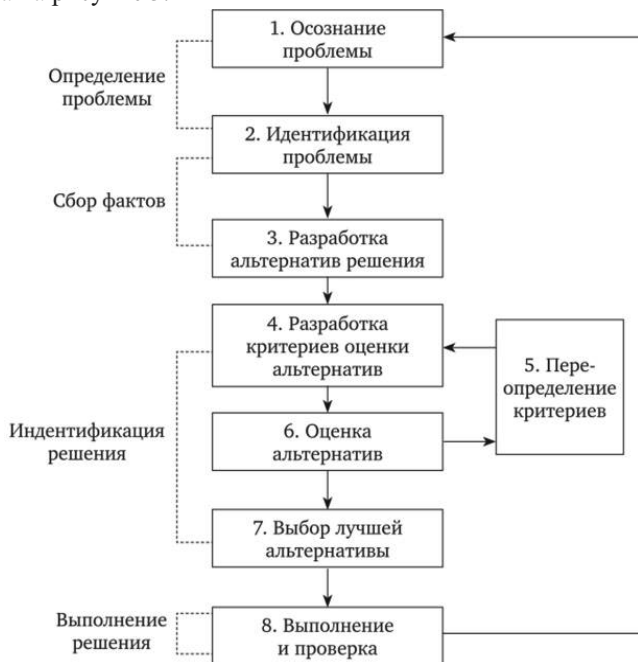


Рисунок 3 – Структура рациональной модели ПР

Интуитивная модель принятия решений.

Вместо логических рассуждений интуитивная модель принятия решений использует интуицию для принятия решений. Часто руководители групп или менеджеры используют эту модель для принятия быстрых решений, когда у них не так много времени на исследования или планирование.

Процесс интуитивного решения менее структурирован и может использовать предыдущие знания о схожих целях и ограничениях для определения оптимального решения.

Компания или организация могут использовать различные варианты систем поддержки принятия решений в зависимости от их целей. Определим некоторые из основных характеристик систем поддержки принятия решений.

Отрасль: организации специализируют свои системы поддержки принятия решений для поддержки своих конкретных стандартов и операций. Для отрасли здравоохранения это может означать доступ к предыдущим медицинским записям для изучения результатов предыдущих решений, а для маркетинга это может означать создание прогнозов на основе текущих рыночных тенденций.

Поддержка: в зависимости от потребностей компании системы поддержки принятия решений могут предлагать различные уровни поддержки. Некоторые системы поддержки принятия решений предлагают прогнозы и траектории, в то время как другие предоставляют конкретные детали для принятия решений, планирования и реализации.

Использование искусственного интеллекта: информация в системе поддержки принятия решений различается по структуре в зависимости от компании, которая ее использует. Называемые интеллектуальными системами поддержки принятия решений (IDSS), ИИ извлекает и обрабатывает большие объемы данных, чтобы получить представление и дать рекомендации для более эффективного принятия решений. Он делает это, анализируя несколько источников данных и выявляя закономерности, тенденции и ассоциации, чтобы имитировать возможности человека по принятию решений. Разработанный, чтобы действовать подобно консультанту-человеку, IDSS собирает и анализирует данные для поддержки лиц, принимающих решения, путем выявления и устранения проблем, а также предоставления и оценки возможных решений. AI-компонент DSS максимально точно имитирует возможности человека, при этом более эффективно обрабатывая и анализируя информацию как

компьютерная система. IDSS может включать расширенные возможности, такие как база знаний, машинное обучение, интеллектуальный анализ данных и пользовательский интерфейс. Примеры реализации IDSS включают гибкие или интеллектуальные производственные системы, интеллектуальные системы поддержки принятия маркетинговых решений и медицинские диагностические системы.

Категория: существует много типов систем поддержки принятия решений, от систем связи до версий, ориентированных на модели. Конкретная структура системы зависит от компании, которая ее использует.

Выделим преимущества использования систем поддержки принятия решений.

Системы поддержки принятия решений могут помочь уменьшить количество ошибок и сделать рабочий процесс более эффективным. Сокращая время, необходимое для принятия важных решений, системы поддержки принятия решений могут упростить процесс и привести к выполнению большего количества задач за определенный период времени. Система также помогает менеджменту избежать ошибок и отрицательных результатов, поддерживая каждое решение прогнозами и сценариями, что еще больше улучшает рабочий процесс.

Система также может улучшить планирование и повысить эффективность управления. Многие системы поддержки принятия решений предлагают точные планы реализации процедур на основе своих данных, что упрощает менеджеру исправление ошибок или запуск новых процессов. Это обеспечивает поддержку руководству и снижает риски, связанные с человеческим фактором.

Системы поддержки принятия решений также могут помочь профессионалам определить эффект, который может иметь их решение. Оценивая прошлые данные и изучая текущие тенденции или информацию, система может синтезировать обоснованные прогнозы о том, как оно повлияет на организацию или клиента. Понимание потенциальных результатов решений позволяет менеджменту определить наилучший сценарий действий.

Эксперты создают системы поддержки принятия решений с учетом конкретных отраслей, организаций и ситуаций. В зависимости от компании и условий система может иметь разные направления и структуру.

Вот некоторые категории, которые можно использовать для маркировки системы:

Коммуникация: направлено на облегчение совместной работы, позволяя различным пользователям получать одновременный доступ к инструментарию и использовать его одновременно. Эта структура ориентирована на предоставление информации группам людей, чтобы они могли сотрудничать при принятии решений.

Основанные на данных: эта категория систем использует как внутренние данные компании, так и внешнюю отраслевую информацию, чтобы помочь пользователям принимать обоснованные решения. Этот метод фокусируется на распознавании, сборе и изучении полезной информации.

На основе документов: система поддержки принятия решений на основе документов использует более неструктурированную информацию для поддержки принятия решений путем хранения, организации и идентификации файлов. Эта категория систем управляет документами, аудиофайлами и видеофайлами.

На экспертных знаниях: это специальная система поддержки принятия решений, которая использует отраслевые факты и процедуры для создания точных рекомендаций для оптимального выбора. Этот тип системы использует прогнозирование, дедукцию и распознавание образов.

Ориентация на модели: системы, ориентированные на модели, используют финансовые данные, моделирование, статистику и процессы оптимизации, чтобы помочь менеджменту при принятии решений. Этот метод меньше полагается на данные, чем другие системы поддержки принятия решений.

Методы поддержки принятия решений в условиях неопределенности и риска

Существует несколько методов принятия решений в условиях неопределенности и риска, в том числе:

Деревья решений. Деревья решений – это визуальное представление процесса принятия решений, позволяющее наметить различные возможные результаты и их вероятности. Они особенно полезны, когда есть несколько точек принятия решений или результатов, которые зависят друг от друга.

Моделирование по *методу Монте-Карло*. Этот метод включает в себя запуск большого количества симуляций с использованием случайных величин для оценки вероятности различных результатов. Этот метод особенно полезен, когда есть много переменных, влияющих на результат, и когда вероятности этих переменных неопределенны.

Анализ чувствительности. Анализ чувствительности включает в себя изучение того, как изменения входных переменных влияют на результат. Это поможет определить, какие переменные оказывают наибольшее влияние на результат, а какие менее важны.

Сценарный анализ. Сценарный анализ включает в себя рассмотрение различных возможных сценариев, которые могут произойти, и анализ того, как они повлияют на результат. Это поможет подготовиться к различным непредвиденным обстоятельствам и принимать более обоснованные решения.

Анализ реальных опционов. Анализ реальных опционов включает в себя рассмотрение потенциальной стоимости различных опционов или решений в будущем, а не только текущей стоимости. Это может помочь принимать решения, учитывающие будущие неопределенность и риск.

Выбор оптимального метода принятия решений в условиях неопределенности и риска зависит от конкретной ситуации и имеющихся данных. Перед принятием решения важно тщательно рассмотреть все доступные варианты и данные, а при необходимости использовать комбинацию методов, чтобы решение было максимально обоснованным и точным.

Наиболее подходящий метод принятия решений в условиях неопределенности и риска будет зависеть от конкретной ситуации и имеющихся данных. Рассмотрим примеров ситуаций, когда другие методы могут быть более подходящими:

Дерева решений. Деревья решений полезны при наличии нескольких точек принятия решений или результатов, которые зависят друг от друга, например, при принятии инвестиционных решений или управлении проектами. Например, если необходимо решить, стоит ли инвестировать в новый продукт, дерево решений поможет спрогнозировать возможные результаты на основе различных рыночных условий и определить наилучший план действий.

Моделирование по методу Монте-Карло. Моделирование по методу Монте-Карло полезно, когда есть много переменных, влияющих на

результат, и когда вероятности этих переменных неопределенны. Его можно использовать в финансовом моделировании, инженерии и других областях. Например, если необходимо спроектировать новый мост, моделирование по методу Монте-Карло может помочь оценить вероятность разрушения при различных условиях и определить уровень риска, связанного с проектом.

Анализ чувствительности. Анализ чувствительности полезен, когда необходимо определить, какие переменные оказывают наибольшее влияние на результат, а какие менее важны. Его можно использовать в финансовом моделировании, управлении цепочками поставок и других областях. Например, если необходимо принять решение о стратегии ценообразования для нового продукта, анализ чувствительности может помочь определить, как изменения в производственных затратах или спросе повлияют на прибыльность.

Сценарный анализ. Анализ сценариев полезен, когда необходимо подготовиться к различным непредвиденным обстоятельствам и принять более обоснованные решения. Его можно использовать в риск-менеджменте, стратегическом планировании и других сферах. Например, если компания является поставщиком медицинских услуг, принимающим решение о плане реагирования на COVID-19, анализ сценариев может помочь подготовиться к различным уровням введения ограничений, увеличения ввиду этого потребностей в услугах компании и определить наилучший план действий.

Анализ реальных опционов. Анализ реальных опционов полезен, когда необходимо принимать решения, учитывающие будущую неопределенность и риск. Его можно использовать в управлении проектами, финансовом моделировании и других областях. Например, если компания является застройщиком и необходимо принять решение о строительстве нового офисного здания, анализ реальных опционов может помочь определить потенциальную будущую стоимость здания на основе различных рыночных условий и определить лучшее время для начала строительства. Выбор наиболее подходящего метода для принятия решений в условиях неопределенности и риска будет зависеть от конкретной ситуации и имеющихся данных.

Ниже приведено более подробное описание каждого метода принятия решений в условиях неопределенности и риска.

Деревья решений – это визуальное представление процесса принятия решений, которое позволяет наметить различные возможные исходы и их вероятности. Процесс создания дерева решений включает в себя следующие этапы:

1. Определить решение, которое необходимо принять, и возможные результаты.

2. Определить различные факторы, которые могут повлиять на результаты. Необходимо присвоить вероятности каждому возможному исходу на основе имеющихся данных.

3. Составить схему дерева решений с каждым возможным исходом и их вероятностями, а также различными точками принятия решений и их потенциальными результатами.

4. Рассчитать математическое ожидание каждого возможного решения на основе вероятностей и исходов. Математическое ожидание дискретной случайной величины X вычисляется как сумма произведений значений x_i , которые принимает случайная величина X , на соответствующие вероятности p_i : $M(X) = \sum_{i=1}^n x_i * p_i$. Для непрерывной случайной величины (заданной плотностью вероятностей $f(x)$), формула вычисления математического ожидания X : $M(X) = \int_{-\infty}^{+\infty} f(x) * x dx$. После создания дерева решений его можно использовать для определения наилучшего плана действий на основе ожидаемого значения каждого возможного решения.

Моделирование по методу Монте-Карло включает в себя запуск большого количества симуляций с использованием случайных величин для оценки вероятности различных исходов. Процесс моделирования по методу Монте-Карло включает в себя следующие этапы:

1. Определить переменные, влияющие на результат, и их распределение вероятностей.

2. Сгенерировать случайные выборки для каждой переменной на основе ее распределения вероятностей.

3. Запустить симуляцию несколько раз, используя разные случайные выборки для каждой переменной.

4. Проанализировать результаты, чтобы оценить вероятность различных исходов.

После того, как моделирование по методу Монте-Карло выполнено, можно использовать его для определения вероятности различных исходов и выявления любых потенциальных рисков или неопределенностей.

Решение задач оптимизации методом Монте-Карло в большинстве случаев основано на многократном случайном выборе вариантов решения и их сравнении. При достаточном количестве испытаний (т. е. вариантов решения, выбранных случайным образом) находится решение, близкое к оптимальному, а в некоторых случаях – оптимальное.

Во многих случаях задачи оптимизации, решаемые методом Монте-Карло, могут быть решены другими (аналитическими) методами, позволяющими найти точное оптимальное решение. Это могут быть методы линейного программирования, динамического программирования, поиска экстремумов функций многих переменных и т. д. Однако применение точных аналитических методов, как правило, возможно только для задач небольшой размерности (с небольшим количеством ограничений, переменных и т. д.). Для реальных задач применение аналитических методов нередко оказывается невозможным или нецелесообразным из-за большого объема сложных вычислений. Поэтому во многих случаях применение метода Монте-Карло может оказаться единственно возможным способом решения задачи. Алгоритм поиска решения на основе метода Монте-Карло полностью зависит от конкретной задачи.

Анализ чувствительности включает в себя изучение того, как изменения входных переменных влияют на результат. Процесс выполнения анализа чувствительности включает в себя следующие этапы:

1. Определение входных переменных и их потенциальные значения.
2. Вычислить результат для каждой комбинации входных переменных.
3. Проанализировать, как изменения каждой входной переменной влияют на результат.

После выполнения анализа чувствительности его можно использовать для определения того, какие переменные оказывают наибольшее влияние на результат, а какие менее важны. Это может помочь принимать более обоснованные решения, основанные на наиболее важных переменных.

Сценарный анализ включает в себя рассмотрение различных возможных сценариев, которые могут произойти, и анализ того, как они повлияют на результат. Процесс анализа сценариев включает в себя следующие шаги:

1. Определение различных сценариев, которые могут возникнуть.
2. Проанализировать, как каждый сценарий повлияет на результат.
3. Назначить вероятности каждому сценарию на основе доступных данных.

После выполнения анализа сценариев его можно использовать для подготовки к различным непредвиденным обстоятельствам и принятия более обоснованных решений на основе вероятности различных сценариев.

Анализ реальных опционов включает в себя рассмотрение потенциальной стоимости различных опционов или решений в будущем, а не только текущей стоимости. Процесс проведения анализа реальных опционов включает в себя следующие этапы:

1. Определение различных вариантов или решения и их потенциальную будущую стоимость.
2. Оценка вероятности различных сценариев будущего и их потенциальное влияние на ценность каждого варианта или решения.
3. Расчет математического ожидания каждого варианта или решения на основе вероятности различных будущих сценариев. После того, как анализ реальных опционов был выполнен, можно использовать его для принятия решений, учитывающих будущую неопределенность и риск.

В целом, каждый метод принятия решений в условиях неопределенности и риска включает в себя различные процессы, но все они направлены на то, чтобы помочь принимать более обоснованные решения на основе имеющихся данных и вероятности различных результатов.

Классические критерии принятия решений. Критерии Вальда, Гурвица и Сэвиджа

Прежде всего отметим принципиальное различие между стохастическими факторами, приводящими к принятию решения в условиях риска, и неопределенными факторами, приводящими к принятию решения в условиях неопределенности. И те, и другие приводят к разбросу возможных исходов результатов управления. Но стохастические факторы полностью описываются известной стохастической информацией, эта информация и позволяет выбрать лучшее в среднем решение [5]. Применительно к неопределенным

факторам подобная информация отсутствует. В общем случае неопределенность может быть вызвана либо противодействием разумного противника, либо недостаточной осведомленностью об условиях, в которых осуществляется выбор решения.

Пусть лицо, принимающее решение, может выбрать один из m возможных вариантов своих решений: x_1, x_2, \dots, x_m и пусть относительно условий, в которых будут реализованы возможные варианты, можно сделать предположений: y_1, y_2, \dots, y_n . Оценки каждого варианта решения в каждом условии (x_i, y_j) известны и заданы в виде матрицы выигрышей лица, принимающего решения: $A = \|a_{ij}\|$.

Теория статистических решений предлагает несколько критериев оптимальности выбора решений. Выбор того или иного критерия неформализуем, он осуществляется человеком, принимающим решения, субъективно, исходя из его опыта, интуиции и т. д. Рассмотрим эти критерии.

Критерий среднего выигрыша. Данный критерий предполагает задание вероятностей состояния обстановки P_i . Эффективность системы оценивается как среднее ожидаемое значение (МОЖ) оценок эффективности по всем состояниям обстановки [2]:

$$K(a_i) = \sum_{j=1}^l P_j k_{ij} \quad i = \overline{1, m}$$

оптимальной системе будет соответствовать эффективность:

$$K^{\text{опт}} = \max_i \sum_{j=1}^l P_j k_{ij} \quad i = \overline{1, m}$$

Критерий минимакса.

$$K(a_i) = \max_j \Delta k_{ij}$$

$$K^{\text{опт}} = \min_i (\max_j \Delta k_{ij})$$

Критерий максимакса.

Этим критерием предписывается оценивать системы по максимальному значению эффективности и выбирать в качестве оптимального решения обследующую эффективность с наибольшим из максимумов:

$$K(a_i) = \max_j k_{ij}, i = \overline{1, m}, j = \overline{1, l}$$

$$K^{\text{опт}} = \max_i (\max_j k_{ij}), i = \overline{1, m}, j = \overline{1, l}$$

Это самое оптимистическое решение. При этом риск максимальный.

Критерий Лапласа. Поскольку вероятности возникновения той или иной ситуации неизвестны, будем их все считать равновероятными. Тогда для каждой строки матрицы выигрышей подсчитывается среднее арифметическое значение оценок. Оптимальному решению будет соответствовать такое решение, которому соответствует максимальное значение этого среднего арифметического, т. е.

$$\bar{F} = F(\bar{X}, Y) = \max_{1 \leq i \leq m} \left(\frac{1}{n} \sum_{j=1}^n a_{ij} \right)$$

Критерий Вальда. В каждой строчке матрицы выбираем минимальную оценку. Оптимальному решению соответствует такое решение, которому соответствует максимум этого минимума, т. е.

$$\bar{F} = F(\bar{X}, Y) = \max_{1 \leq i \leq m} \min_{1 \leq j \leq n} a_{ij}$$

Этот критерий очень осторожен. Он ориентирован на наихудшие условия, только среди которых и отыскивается наилучший и теперь уже гарантированный результат.

Критерий Сэвиджа. В каждом столбце матрицы находится максимальная оценка и составляется новая матрица, элементы которой определяются соотношением

$$r_{ij} = \max_{1 \leq i \leq m} a_{ij} - a_{ij}$$

Величину r_{ij} называют риском [14], под которым понимают разность между максимальным выигрышем, который имел бы место, если бы было достоверно известно, что наступит ситуация u_j , и выигрышем при выборе решения x_i в условиях u_j . Эта новая матрица называется матрицей рисков. Далее из матрицы рисков выбирают такое решение, при котором величина риска принимает наименьшее значение в самой неблагоприятной ситуации, т.е.

$$\bar{F} = F(\bar{X}, Y) = \min_{1 \leq i \leq m} \max_{1 \leq j \leq n} r_{ij} = \min_{1 \leq i \leq m} \max_{1 \leq j \leq n} (\max_{1 \leq i \leq m} a_{ij} - a_{ij})$$

Сущность этого критерия заключается в минимизации риска. Как и критерий Вальда, критерий Сэвиджа очень осторожен. Они различаются разным пониманием худшей ситуации: в первом случае – это минимальный выигрыш, во втором – максимальная потеря выигрыша по сравнению с тем, чего можно было бы достичь в данных условиях.

Принцип гарантированного результата

Рассмотрим задачу, основанную на принципе гарантированного результата. Органы управления некоторой территорией планируют выполнение специальных *работ* (таких, как, например, прокладка путепровода, возведение спортивного комплекса и т.п.) силами подрядчика, выбираемого на основе конкурса. Финансирование *работ* предусмотрено местным бюджетом [14].

Примем, что сроки проведения конкурса не утверждены (например, в силу их зависимости от обстоятельств, определяемых интересами различных групп влияния). Однако известно, что конкретный момент t проведения конкурса после его утверждения будет укладываться в *отрезок* времени, который мы обозначим как $[0,2]$ (например, два ближайших месяца или два квартала). Известно также, что определяющим критерием при выборе победителя конкурса является показатель качества, которое может обеспечить претендент при проведении *работ*. Для количественной оценки этого показателя (в таких единицах, как, например, баллы и доли баллов, пункты и подпункты и т.п.) органами управления утверждена соответствующая методика [14].

Условия конкурса предусматривают ситуацию, когда уровни качества *работ*, заявленные и обоснованные участниками, оказываются одинаковыми (в рамках принятой системы оценки показателей). Для такого случая правила предусматривают согласительную процедуру, допускающую предложение сторонам совместно создать некоторое предприятие, которому и будет дан подряд на выполнение *работ*. Кроме того, конкурсная комиссия может отказать всем участникам, если предлагаемый ими уровень качества оказывается ниже некоторой отметки, также предусмотренной правилами.

Две фирмы, обозначаемые в дальнейшем как P_1 и P_2 , планируют участвовать в конкурсе. Примем, что оценка W_i качества *работ*, которую фирма P_i сможет подтвердить в случае проведения конкурса в момент t , зависит от объема ресурса x_i , вложенного этой фирмой за период $[0,t]$ в освоение более эффективных технологий ведения *работ*. Пусть обуславливаемая зависимость имеет вид

$$W_i(x_i, t) = x_i(t - 1) + 1, 0 \leq x_i \leq 1, i = 1, 2,$$

где максимально доступный объем ресурса принят за единицу.

Если фирма P_i не осуществляет инвестиций в освоение новых технологий (т.е., если P_i выбирает вариант $x_i=0$) и использует имеющиеся средства, например, для немедленного укрепления материальной базы,

обеспечивающей ведение *работ* традиционным способом, то уровень качества $W_i(0,t)$ остается постоянным в течение всего периода $[0,2]$. При этом $W_i(0,t)=1$.

В случае вложения всех ресурсов в освоение новых технологий (т.е. при выборе варианта $x_i=1$, к концу периода $[0,2]$ достигается более высокий уровень качества $W_i(1,2)=2$. Однако при этом в начальный момент $t=0$ уровень качества остается таким же, каким он был до начала подготовки к тендеру. Это *значение* принято за нулевую отметку.

Очевидно, что формирование представления о *лучшей* стратегии x_1^* стороны P_1 предполагает либо возможность определения лучшего варианта для любой пары стратегий x_1', x_1'' , этой стороны, либо возможность установления *равноценности* стратегий, входящих в эту пару. Однако на множестве стратегий стороны P_1 не существует *отношения предпочтения*, позволяющего ответить на эти вопросы для любой пары x_1', x_1'' .

Проиллюстрируем это важное обстоятельство путем сравнения уровней эффективности, обеспечиваемых соответственно стратегиями $x_1=0$ и $x_1=1$.

$$M_1(x_1, x_2, t) = (x_1 - x_2)(t - 1), \text{ откуда следует, что:}$$

$$\Delta_1(x_2, t) = M_1(x_1 = 0, x_2, t) - M_1(x_1 = 1, x_2, t) = 1 - t,$$

следовательно:

$$\Delta_1(x_2, t) \geq 0, 0 \leq t \leq 1; \Delta_1(x_2, t) < 0, 1 < t \leq 2; \Delta_1(x_2, 1) = 0$$

Как следует из предшествующего рассмотрения, для обеспечения сравнимости стратегий принципиально необходимо принять некоторую *гипотезу* о неизвестном *состоянии природы*. В рассматриваемом примере вся имеющаяся у стороны P_1 *информация* о сроке t проведения конкурса сводится к знанию интервала $[0, 2]$, заведомо содержащего этот неизвестный срок. В связи с указанной *неопределенностью* состояния природы, в качестве оценки эффективности любой стратегии можно принять тот уровень эффективности, который *гарантируется* использованием этой стратегии.

Фактически, принятие гарантируемого стратегией уровня эффективности в качестве оценки, на которой будет основано сравнение этой стратегии с другими, означает *ориентацию на худший случай*.

Принятие такой оценки в качестве *прогноза* результатов планируемых действий является рекомендацией, основанной на обширном опыте принятия решений в практической деятельности. К этому "правилу

худшего случая" приходят многочисленные исследователи опыта *принятия решений*, относящегося к самым различным областям человеческой деятельности. Приведем несколько примеров.

Известный американский специалист в области создания больших программных систем Ф.П. Брукс отмечает, что «наши методы оценки весьма несовершенны. Строго говоря, они отражают некоторое неявно высказываемое и в корне неверное допущение, что все будет идти хорошо, выполнение каждого задания займет ровно столько времени, сколько оно «должно» занять». И далее: «Планируйте неудачу: она вас, так или иначе, найдет».

Можно даже говорить о возникновении своего рода «фольклора», вызванного к жизни необходимостью ориентации на худший случай в практике принятия решений. К числу таких новых жанров относятся, например, так называемые «законы Мэрфи»:

1. Все сложнее, чем кажется.
2. Все тянется дольше, чем можно ожидать.
3. Все оказывается дороже, чем планировалось.
4. Если что-то может испортиться, оно обязательно испортится.

По поводу этих законов Каллаген сделал следующее замечание: «Мэрфи был оптимистом». Действительно, например, второй из «законов» неявно предполагает, что планируемая работа, в конце концов, все-таки завершится. Но этого успешного завершения может и не быть.

Вернемся к рассматриваемому примеру и построим оценку эффективности, которую гарантирует стратегия x_1 стороны P_1 при неизвестном сроке t проведения конкурса. Эта оценка худшего случая, очевидно, определяется величиной

$$M_1(x_1, x_2) = \min\{M_1(x_1, x_2, t): 0 \leq t \leq 2\}$$

Подставляя в правую часть выражения функции $M_1(x_1, x_2, t)$, приводим оценку к виду:

$$\min_{0 \leq t \leq 2} (x_1 - x_2)(t - 1) = \min_{0 \leq t \leq 2} \begin{cases} |x_1 - x_2|(t - 1), & x_1 \geq x_2 \\ |x_1 - x_2|(t - 1), & x_1 \leq x_2 \end{cases}$$

Можно сделать вывод, что $M_1(x_1, x_2) = -|x_1 - x_2|$, причем эта гарантированная величина реализуется либо в случае проведения конкурса в момент $t^* = 0$, либо в случае проведения этого конкурса в момент $t^* = 2$.

Первый случай соответствует ситуации, когда $x_1 \geq x_2$, а второй – ситуации, когда $x_1 \leq x_2$, т.е.:

$$M_1(x_1, x_2) = \begin{cases} M_1(x_1, x_2, 0), & x_1 \geq x_2 \\ M_1(x_1, x_2, 2), & x_1 \leq x_2 \end{cases}$$

Теперь проведем аналогичное рассмотрение, руководствуясь интересами второй стороны. Определим уровень эффективности, который может быть обеспечен стороне P_2 выбором стратегии x_2 при некоторой известной стратегии x_1 первой стороны и неизвестном сроке проведения конкурса, т.е. вычислим величину: $M_2(x_1, x_2) = \min\{M_2(x_1, x_2, t): 0 \leq t \leq 2\}$, тогда $M_2(x_1, x_2) = -|x_1 - x_2|$. При этом справедливо следующее соотношение:

$$M_2(x_1, x_2) = \begin{cases} M_2(x_1, x_2, 2), & x_1 \geq x_2 \\ M_2(x_1, x_2, 0), & x_1 \leq x_2 \end{cases}$$

Таким образом: $-1 \leq M_1(x_1, x_2) = M_2(x_1, x_2) \leq 0$. Следовательно, при ориентации обеих сторон на худший случай *противоположность* интересов сторон, характеризуемая нулевой суммой критериев, сменяется ситуацией полного *совпадения интересов*.

ГЛАВА 5. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Современный мир часто характеризуют как "век информации" из-за огромного количества доступной информации, которую мы можем получить из различных источников, и ценным источником информации может быть практически всё что угодно.

Технологии записи и хранения данных непрерывно совершенствуются и приводят к доступности для людей колоссальных потоков информации. В связи с этим скорость роста данных остается одним из ключевых трендов в информационных технологиях. Объемы данных продолжают расти с экспоненциальной скоростью, и это влияет на множество аспектов бизнеса, науки и технологий.

По данным Data Age Report человечество сформировало порядка 51 зеттабайта информации в 2020 году. К 2025 году объем этих данных вырастет до 175 зеттабайтов ежегодно.



Объем данных в мире в зеттабайтах *

Рис.1. Прогноз темпа роста объема данных

Эти данные - это сложные и объемные наборы разной информации, которые, из-за своего размера, сложности и скорости изменений, становятся трудными для обработки с использованием традиционных баз данных и методов обработки данных.

Рассмотрим несколько факторов, объясняющих эту динамику.

- Интернет вещей (IoT) и сенсорные технологии:* Подключение все большего количества устройств и сенсоров приводит к генерации огромных объемов данных от умных устройств, датчиков и промышленного оборудования.
- Медицинские и научные исследования:* Биомедицинские и научные исследования генерируют все больше данных, включая геномные данные,

изображения медицинских сканеров и большие массивы экспериментальных данных.

3. *Цифровая трансформация в компаниях*: Множество компаний активно внедряют цифровые технологии, в частности, цифровые платформы, торговые данные и информацию о клиентах, что способствует росту данных.

4. *Облачные вычисления и хранилища данных*: Многие организации увеличивают использование облачных хранилищ данных, что также увеличивает общий объем данных в цифровой среде.

В современном мире мы окружены большими данными, которые, по прогнозам, но при этом сгенерированные «тонны аморфных данных» оказываются зачастую бесполезными, поскольку полезные данные спрятаны глубоко внутри. Парадокс заключается в том, что мы тонем в данных, но в то же время мы изголодались по знаниям (или полезным данным). Поэтому без мощных инструментов, таких как интеллектуальный анализ данных (Data Mining), мы не сможем добывать ценные данные, и, как следствие, мы не получим от них никакой пользы.

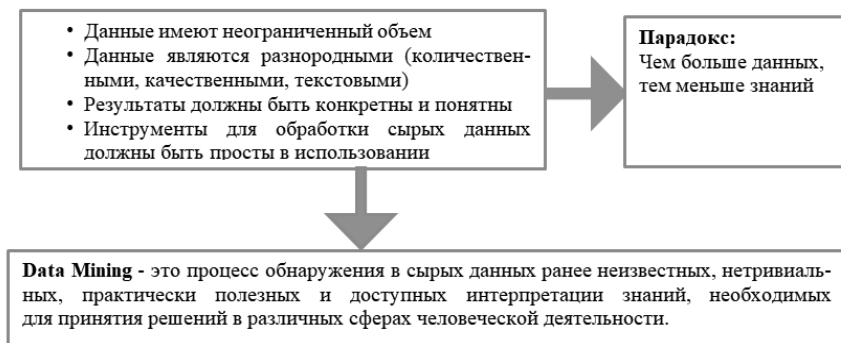


Рис. 2 Взаимосвязь данных и знаний

2. Понятие интеллектуального анализа данных и связь со смежными предметными областями

Интеллектуальный анализ данных — это метод извлечения данных или ранее неизвестных шаблонов данных из огромных наборов данных. Следовательно, как следует из названия, мы «добываем конкретные данные» из большого набора данных.

Термин интеллектуальный анализ данных (Data Mining) получил свое название из двух понятий: поиска ценной информации в большой

базе данных (data) и добычи горной руды (mining). Оба процесса требуют или просеивания огромного количества сырого материала, или разумного исследования и поиска искомым ценностей. Он часто переводится как добыча данных, извлечение информации, раскопка данных, интеллектуальный анализ данных, средства поиска закономерностей, извлечение знаний, анализ шаблонов, "извлечение зерен знаний из гор данных", раскопка знаний в базах данных, информационная проходка данных, "промывание" данных. Поскольку найти шаблон, полученный в результате интеллектуального анализа данных, может быть очень сложно, иногда его сравнивают с добычей золота в реках (рис. 3).



Рис. 3. Добыча золота и интеллектуальный анализ данных

Впервые понятие Data Mining появилось в 1989 году. Изначально оно было связано с автоматизацией и оптимизацией запросов к крупным базам данных. Между тем понятие анализ данных (англ. Data Analysis) существовало намного раньше и означало обработку и интерпретацию данных, полученных в ходе экспериментов, в основном научных. С развитием науки и техники эти понятия расширились и обобщались, стали очень близки друг к другу, и в настоящий момент тесно связаны как с анализом больших объёмов данных (англ. Big Data), так и с понятием машинного обучения (англ. Machine Learning).

История развития интеллектуального анализа данных

История развития интеллектуального анализа данных начинается в 1956 году, когда на семинаре с аналогичным названием в Дартсмутском колледже (США) был предложен термин искусственный интеллект (ИИ) – (AI-artificial intelligence). Семинар был посвящен разработке методов решения логических, а не вычислительных задач. В английском языке данное словосочетание не имеет той слегка фантастической антропоморфной окраски, которую оно приобрело в достаточно не

удачном русском переводе. Слово intelligence означает «умение рассуждать разумно», а вовсе не «интеллект», для которого есть термин intellect.

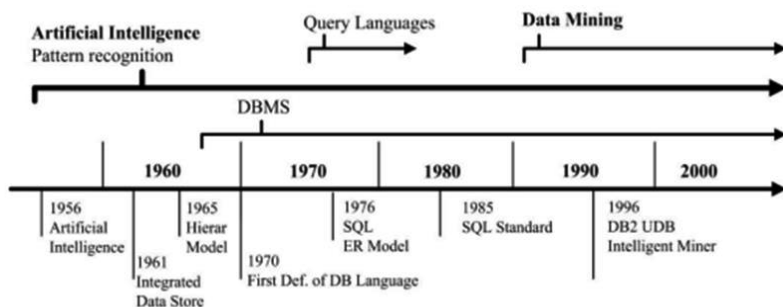


Рис.4. История развития интеллектуального анализа данных

1965 - создание «интегрированной системы хранения» IDS (Integrated DataStore) Чарльзом Бахманом.

Ранние методы выявления закономерностей в данных включают теорему Байеса (1700-е годы) и эволюцию регрессии (1800-е годы). Генерация и растущая мощь информатики способствовали сбору, хранению и обработке данных, поскольку наборы данных имеют большую размерность и уровень сложности. Явное практическое исследование данных постепенно совершенствовалось благодаря косвенной, автоматической обработке данных и другим открытиям в области информатики, таким как нейронные сети, кластеризация, генетические алгоритмы (1950-е годы), деревья решений (1960-е годы) и вспомогательные векторные машины (1990-е годы).

После того, как в 1960-х и 1970-х годах были разработаны примитивные компьютеры, Джон Генри Холланд в 1975 году написал книгу «Адаптации в природе и искусственной науке», в которой изложил генетические алгоритмы. Эти алгоритмы послужили основой для начала интеллектуального анализа данных в том виде, в котором мы его знаем, который был разработан в конце 1980-х годов. В конце 1980-х годов интеллектуальный анализ данных стал «устоявшейся дисциплиной в рамках информатики».

В 1989 году Григорий Пятецкий-Шапиро основал термин «Knowledge Discovery in Databases» (KDD). К началу 1990-х годов, интеллектуальный анализ данных был признан «подпроцессом в рамках

более крупного процесса, называемого обнаружением знаний в базах данных (KDD). KDD относится к обнаружению скрытой информации с широким спектром данных, что относится к различным процессам, которые включают в себя «подготовку данных (хранение, очистку данных, предварительную обработку данных и т. д.) и анализ/визуализацию результатов». Однако, по мнению Коэнена, сегодня для практических целей специалисты по информатике и практики, использующие интеллектуальный анализ данных в своей отрасли, используют KDD и интеллектуальный анализ данных как синонимы. Тем не менее, с технической точки зрения, интеллектуальный анализ данных является подпроцессом KDD.

Истоки интеллектуального анализа данных — это базы данных, статистика. Термин «интеллектуальный анализ данных» появился в сообществе баз данных в 1990-х годах, но в широком смысле интеллектуальный анализ данных — это эволюция предметной области с обширной историей. Таким образом, можно обобщить, что истоки интеллектуального анализа данных восходят к трем родственным линиям: классическая статистика, искусственный интеллект и машинное обучение.

Статистика является основой большинства технологий, на которых строится интеллектуальный анализ данных, таких как регрессионный анализ, стандартное отклонение, стандартное распределение, стандартная дисперсия, дискриминантный анализ, кластерный анализ и доверительные интервалы. Все они используются для анализа данных и соединения данных.

Статистика составляет основную часть интеллектуального анализа данных, которая включает в себя общую процедуру анализа данных. Посмотрим на разницу между интеллектуальным анализом данных и статистикой.

Таблица 1. Разница между ИАД и статистикой

Интеллектуальный анализ данных	Статистика
Интеллектуальный анализ данных — это процесс извлечения полезной информации, закономерностей и тенденций из огромных наборов данных и использования их для принятия решений на основе данных.	Статистика относится к анализу и представлению числовых данных и является основной частью всех алгоритмов интеллектуального анализа данных.
Данные, используемые в интеллектуальном анализе данных, могут быть числовыми или нечисловыми.	Данные, используемые в статистике, являются только числовыми.
В интеллектуальном анализе данных сбор	В статистике важнее сбор данных.

Интеллектуальный анализ данных	Статистика
данных не является более важным.	
К видам интеллектуального анализа данных относятся кластеризация, классификация, ассоциация, нейронная сеть, анализ на основе последовательностей, визуализация и т.д.	Типы статистики: описательная статистическая и выводная статистическая.
Подходит для больших и сверхбольших наборов данных.	Подходит для небольших и больших наборов данных.
Интеллектуальный анализ данных является индуктивным процессом. Это означает генерацию новой теории из данных.	Статистика – это дедуктивный процесс. Он не позволяет себе делать какие-либо прогнозы.
Очистка данных является частью интеллектуального анализа данных.	В статистике для реализации статистического метода используются чистые данные.
Для проверки модели требуется меньше взаимодействия с пользователем, поэтому ее легко автоматизировать.	Для проверки модели требуется взаимодействие с пользователем, поэтому ее сложно автоматизировать.
Приложения для интеллектуального анализа данных включают анализ финансовых данных, розничную торговлю, телекоммуникационную отрасль, анализ биологических данных, некоторые научные приложения и т. д.	Применение статистики включает в себя биостатистику, контроль качества, демографию, операционные исследования и т.д.

Искусственный интеллект (ИИ) основан на эвристике, а не на статистике. Она пытается применить человеческую обработку к статистическим задачам. Искусственный интеллект — это реальный информационный продукт, способный выполнять поставленные задачи и решать проблемы примерно так же, как это делают люди в деловом мире. Функции систем ИИ охватывают обучение, планирование, рассуждение, принятие решений и решение проблем.

Существует два типа искусственного интеллекта:

✓ Слабый ИИ: Узкий ИИ или узкий искусственный интеллект (ANI). Он обучен ИИ и ориентирован на выполнение конкретных задач. Слабый ИИ управляет большей частью ИИ, который окружает нас сегодня. «Узкий» может быть более точным описанием для этого типа ИИ, поскольку он совсем не слаб; Он позволяет использовать некоторые очень надежные приложения, такие как Siri от Apple, Alexa от Amazon, IBM Watson и автономные транспортные средства.

✓ Сильный ИИ: Он состоит из общего искусственного интеллекта (AGI) и искусственного суперинтеллекта (ASI). Общий искусственный интеллект (AGI), или общий ИИ, — это теоретическая

форма ИИ, в которой машина будет иметь интеллект, равный человеческому. У него будет самосознательное сознание, которое может решать проблемы, учиться и планировать будущее.

Искусственный суперинтеллект (ИСИ) превзойдет интеллект и способности человеческого мозга. Несмотря на то, что сильный ИИ по-прежнему является полностью теоретическим и не имеет практических примеров, это не означает, что исследователи ИИ не изучают его развитие.

Несмотря на то, что между ИАД и ИИ существует сильная корреляция и значительное пересечение, они различны и имеют различное применение.

Таблица 2. Различие между ИАД и ИИ

Условия	Искусственный интеллект	Интеллектуальный анализ данных
Понятие	Искусственный интеллект призван облегчить работу программного обеспечения, которое может рассуждать на основе входных данных и объяснять выходные данные. Искусственный интеллект обеспечивает взаимодействие с программным обеспечением, подобное человеческому, и предлагает поддержку принятия решений для конкретных задач, но он не заменяет людей.	Он находит инсайты и строит прогнозы на будущее.
Важность	<ul style="list-style-type: none"> ✓ Возможность работы с большими наборами данных Более высокая скорость ✓ Внедряйте инновации, проектируйте и разрабатывайте высокодоходные продукты и услуги. 	<ul style="list-style-type: none"> ✓ Помогает обнаружить, как различные атрибуты наборов данных коррелируют с помощью шаблонов и методов визуализации данных.
Метод работы	Искусственный интеллект работает, интегрируя огромные объемы данных с быстрой, итеративной обработкой и интеллектуальными алгоритмами.	Углубляется в данные и извлекает из них пол
Использует	<ul style="list-style-type: none"> Производственные роботы Беспилотные автомобили Умные помощники Упреждающее управление здравоохранением Картирование заболеваний Автоматизированное финансовое инвестирование 	<ul style="list-style-type: none"> Веб-майнинг, интеллектуальный анализ текста, обнаружение мошенничества

Условия	Искусственный интеллект	Интеллектуальный анализ данных
	Виртуальный агент по бронированию путешествий Мониторинг социальных сетей Межкомандный чат Бот для разговорного маркетинга Инструменты обработки естественного языка (NLP)	
Вмешательство человека	Машины на основе искусственного интеллекта быстры, точны и логичны, но они не являются врожденными. Им не хватает эмоций, и они не чувствительны к культуре.	Для этого потребовалась ручная техника.

Машинное обучение — это комбинация статистики и искусственного интеллекта. Его можно рассматривать как эволюцию ИИ, потому что он сочетает эвристику ИИ со сложным статистическим анализом. Он использует статистику для основных концепций и добавляет больше эвристик и алгоритмов ИИ для достижения своей цели.

Артур Сэмюэл изобрел термин «машинное обучение», американский пионер в области компьютерных игр и искусственного интеллекта в 1959 году. Он сказал, что «это дает компьютерам возможность обучаться без явного программирования».

Машинное обучение — это метод, который создает сложные алгоритмы для обработки больших данных и предоставляет результаты своим пользователям. Оно использует сложные программы, которые могут учиться на опыте и делать прогнозы.

Алгоритмы усиливаются сами по себе за счет частого ввода обучающих данных. Цель машинного обучения состоит в том, чтобы понимать информацию и строить модели на основе данных, которые могут быть поняты и использованы людьми.

Алгоритмы машинного обучения делятся на два типа:

- ✓ Обучение без учителя
- ✓ Обучение с учителем (контролируемое обучение).

Обучение без учителя не зависит от обученных наборов данных для прогнозирования результатов, но оно использует прямые методы, такие как кластеризация и ассоциация, для прогнозирования результатов. Обученные наборы данных определяются как входные данные, для которых известны выходные данные.

Контролируемое обучение относится к присутствию супервизора в качестве учителя. Контролируемое обучение — это процесс обучения, в котором мы обучаем или тренируем машину, используя данные, которые хорошо выровнены, подразумевая, что некоторые данные уже помечены правильными ответами. После этого машине предоставляются новые наборы данных, чтобы алгоритм обучения с учителем проанализировал обучающие данные и выдал точный результат из размеченных данных.

Основные различия между интеллектуальным анализом данных и машинным обучением:

1. Для внедрения методов интеллектуального анализа данных используется два компонента: первый — база данных, а второй — машинное обучение. База данных предоставляет методы управления данными, в то время как машинное обучение предоставляет методы анализа данных. Но для внедрения методов машинного обучения использовались алгоритмы.

2. Интеллектуальный анализ данных использует больше данных для получения полезной информации, и эти конкретные данные помогут предсказать некоторые будущие результаты. Машинное обучение более сильно зависит от алгоритмом, чем от данных.

3. Интеллектуальный анализ данных не способен к самообучению. Он следует заранее определенным рекомендациям. Он даст ответ на конкретную проблему, но алгоритмы машинного обучения самоопределяются и могут изменять свои правила в зависимости от ситуации, а также находить решение для конкретной проблемы и решать ее по-своему.

4. Основное и самое важное отличие интеллектуального анализа данных от машинного обучения заключается в том, что без участия человека интеллектуальный анализ данных не может работать, но в случае с машинным обучением человеческие усилия вовлекаются только в тот момент, когда алгоритм определен, после чего он все сделает сам собой. После того, как он реализован, мы можем использовать его вечно, но это невозможно в случае интеллектуального анализа данных.

5. Поскольку машинное обучение является автоматизированным процессом, результат, получаемый машинным обучением, будет более точным по сравнению с интеллектуальным анализом данных.

6. Интеллектуальный анализ данных использует базу данных, сервер хранилища данных, механизм интеллектуального анализа данных и

методы оценки шаблонов для получения полезной информации, в то время как машинное обучение использует нейронные сети, прогнозные модели и автоматизированные алгоритмы для принятия решений.

Таблица 3. Различие между ИАД и машинным обучением

Фактор	Интеллектуальный анализ данных	Машинное обучение
Происхождение	Традиционные базы данных с неструктурированными данными.	Есть существующий алгоритм и данные.
Значение	Извлечение информации из огромного количества данных.	Вводите новую информацию из данных, а также из предыдущего опыта.
История	В 1930 году она была известна как обнаружение знаний в базах данных (KDD).	Первая программа, т.е. программа Самуэля по игре в шашки, была создана в 1950 году.
Ответственность	Интеллектуальный анализ данных используется для получения правил из существующих данных.	Машинное обучение учит компьютер тому, как изучать и понимать правила.
Природа	Это больше связано с вмешательством человека в руководство.	Она автоматизирована, после разработки и внедрения нет необходимости в человеческих усилиях.
Техники включают в себя	Интеллектуальный анализ данных — это скорее исследование с использованием такого метода, как машинное обучение.	Это самообучаемая и обучаемая система, позволяющая точно выполнять поставленную задачу.
Размах	Применяется в ограниченных областях.	Его можно использовать в самых разных предметных областях.

Именно поэтому интеллектуальный анализ данных(ИАД) имеет много определений из-за того, что этот термин охватывает широкий спектр методов и подходов к анализу данных. В зависимости от контекста и области применения, определение интеллектуального анализа данных может немного различаться. Приведем несколько из них.

Интеллектуальный анализ данных — это процесс исследования скрытых закономерностей информации с различных точек зрения для категоризации в полезные данные, которые собираются и собираются в определенных областях, таких как хранилища данных, эффективный анализ, алгоритм интеллектуального анализа данных, помощь в принятии

решений и другие требования к данным, чтобы в конечном итоге сократить расходы и получить доход.

Интеллектуальный анализ данных (англ. Data Mining) – это раздел информатики, изучающий процессы обработки данных с целью получения полезной информации и принятия решений.

Интеллектуальный анализ данных — это автоматический поиск больших хранилищ информации для поиска тенденций и закономерностей, выходящих за рамки простых процедур анализа. Интеллектуальный анализ данных использует сложные математические алгоритмы для сегментов данных и оценивает вероятность будущих событий.

Интеллектуальный анализ данных также называют обнаружением знаний о данных (KDD).

На сегодняшний день в англоязычной литературе можно встретить различные термины и их сочетания, описывающие область интеллектуального анализа данных и являющихся достаточно близкими по значению: Data Mining; Statistical Analysis and Data Mining; Machine Learning; Deep Learning; Predictive Analytics and Data Mining; Data Science; Data Science and Data Mining; Discovery Driven Data Mining; Knowledge Discovery in Data bases; Big Data и др.

3. Типы исходных данных в интеллектуальном анализе данных

Интеллектуальный анализ данных может выполняться со следующими типами данных:

- ✓ Реляционная база данных;
- ✓ Хранилища данных;
- ✓ Репозитории данных;
- ✓ Объектно-реляционная база данных;
- ✓ Транзакционная база данных.

Реляционная база данных представляет собой коллекцию из нескольких наборов данных, формально организованных по таблицам, записям и столбцам, из которых можно получить доступ к данным различными способами без необходимости распознавания таблиц базы данных. Таблицы передают информацию и обмениваются ею, что облегчает поиск данных, создание отчетов и организацию.

Хранилище данных — это технология, которая собирает данные из различных источников в организации для предоставления значимых

бизнес-аналитик. Хранилище данных относится к месту, где данные могут храниться для полезного интеллектуального анализа. Это похоже на быструю компьютерную систему с исключительно огромной емкостью хранения данных. Данные из различных систем организации копируются в хранилище, где они могут быть извлечены и согласованы для удаления ошибок. Здесь можно делать расширенные запросы к хранилищу данных.

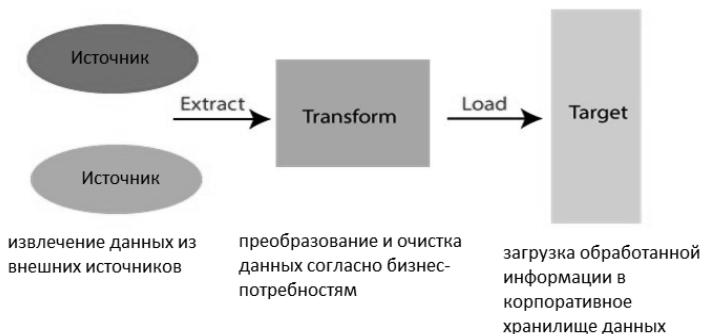


Рис. 5. Понятие ETL (Extract, Transform, Load)

Хранилище данных объединяет данные из многочисленных источников, что обеспечивает качество, точность и согласованность данных. Хранилище данных повышает производительность системы, отделяя аналитическую обработку от транзакционных баз данных. Данные поступают в хранилище данных из разных баз данных. Хранилище данных работает, сортируя данные по шаблону, который отображает формат и типы данных. Инструменты запросов проверяют таблицы данных с помощью шаблонов.

Хранилища данных и базы данных являются относительными информационными системами, но и те, и другие предназначены для разных целей. Хранилище данных предназначено для хранения огромного объема исторических данных и обеспечивает быстрые запросы ко всем данным, как правило, с использованием оперативной аналитической обработки (OLAP). База данных предназначена для хранения текущих транзакций и обеспечения быстрого доступа к конкретным транзакциям для текущих бизнес-процессов, широко известных как оперативная обработка транзакций (OLTP)

Репозиторий данных обычно относится к месту назначения для хранения данных. Тем не менее, многие ИТ-специалисты используют этот термин более четко для обозначения определенного типа настройки в ИТ-

структуре. Например, группа баз данных, в которых организация хранила различного рода информацию.

Комбинация объектно-ориентированной модели базы данных и модели реляционной базы данных называется *объектно-реляционной моделью*. Она поддерживает классы, объекты, наследование и т.д. Одной из основных целей объектно-реляционной модели данных является устранение разрыва между реляционной базой данных и объектно-ориентированной моделью, часто используемой во многих языках программирования.

Транзакционная база данных относится к системе управления базами данных (СУБД), которая может отменить транзакцию базы данных, если она не выполняется должным образом. Несмотря на то, что давным-давно эта возможность была уникальной, сегодня большинство систем реляционных баз данных поддерживают транзакционные базы данных.

4. Методы интеллектуального анализа данных

В целом интеллектуальный анализ данных делится на два класса: описательный анализ и прогнозный анализ с использованием различных методов и функций. В виду того, что дисциплина методы интеллектуального анализа данных находится на стыке многих дисциплин, методы ИАД условно можно разделить на три основные группы: статистические методы, методы машинного обучения и методы искусственного интеллекта.



Рис.6. Методы интеллектуального анализа данных

1. Классификация

Этот метод используется для получения важной и актуальной информации о данных и метаданных. Этот метод ИАД помогает классифицировать данные по различным классам.

Методы интеллектуального анализа данных можно классифицировать по различным критериям, а именно:

✓ Классификация платформ интеллектуального анализа данных в соответствии с типом добываемых источников данных: эта классификация осуществляется в соответствии с типом обрабатываемых данных. Например, мультимедиа, пространственные данные, текстовые данные, данные временных рядов, Всемирная паутина и так далее.

✓ Классификация платформ интеллектуального анализа данных в соответствии с используемой базой данных: классификация, основанная на используемой модели данных. Например. Объектно-ориентированная база данных, транзакционная база данных, реляционная база данных и т. д.

✓ Классификация платформ интеллектуального анализа данных в соответствии с типом обнаруженных знаний. Эта классификация зависит от типов обнаруженных знаний или функциональных возможностей интеллектуального анализа данных. Например, дискриминация, классификация, кластеризация, характеристика и т. д. Некоторые фреймворки, как правило, представляют собой экстенсивные фреймворки, предлагающие несколько функций интеллектуального анализа данных вместе.

✓ Классификация платформ интеллектуального анализа данных в соответствии с используемыми методами интеллектуального анализа данных. Эта классификация основана на используемом подходе к анализу данных, таком как нейронные сети, машинное обучение, генетические алгоритмы, визуализация, статистика, ориентированные на хранилища данных или базы данных и т. д. Классификация также может учитывать уровень взаимодействия с пользователем, участвующего в процедуре интеллектуального анализа данных, например, системы, управляемые запросами, автономные системы, или интерактивные исследовательские системы.

Термин «классификация» обычно используется, когда существует ровно два целевых класса, называемых двоичной классификацией. Когда можно предсказать более двух классов, особенно в задачах распознавания образов, это часто называют мультиномиальной классификацией. Однако

мультиномиальная классификация также используется для категориальных данных ответов, когда нужно предсказать, какая категория из нескольких категорий имеет экземпляры с наибольшей вероятностью.

Классификация является одной из важнейших задач в интеллектуальном анализе данных. Она относится к процессу присвоения предопределенных меток классов экземплярам на основе их атрибутов.

Важные
методы
классификации
в ИАД

Метод логистической регрессии. Метод логистической регрессии используется для прогнозирования переменной отклика.

Метод k-ближайших соседей. Метод K-ближайших соседей используется для классификации наборов данных в так называемые K-наблюдения. Он используется для определения сходства между соседями.

Упрощенный метод Байеса. Метод упрощенного байеса используется для сканирования набора данных и поиска записей, в которых значения предикторов равны..

Метод нейронных сетей. Нейронные сети напоминают структуру нашего мозга, называемую нейроном. Наборы данных проходят через эти сети и, наконец, выходят на выход. Этот метод нейронной сети сравнивает различные классификации. Ошибки, возникающие в классификациях, дополнительно исправляются и передаются в сети. Это повторяющийся процесс.

Метод дискриминантного анализа. В этом методе строится линейная функция и используется для предсказания класса переменных из наблюдений с неизвестным классом.

Рис. 7. Важные методы классификации

2. Кластеризация

В литературе можно встретить разные определения кластера и кластеризации. Приведем несколько из них.

Кластер — это подмножество похожих объектов.

Кластер - это связанная область многомерного пространства со сравнительно высокой плотностью объектов.

Кластеризация — это метод преобразования группы абстрактных объектов в классы похожих объектов.

Кластеризация — это метод секционирования набора данных или объектов на набор значимых подклассов, называемых кластерами.

Чтобы лучше понять сложность выбора того, что формирует группу, на рис. 8 показаны двадцать точек и три различных способа их разделения на кластеры.

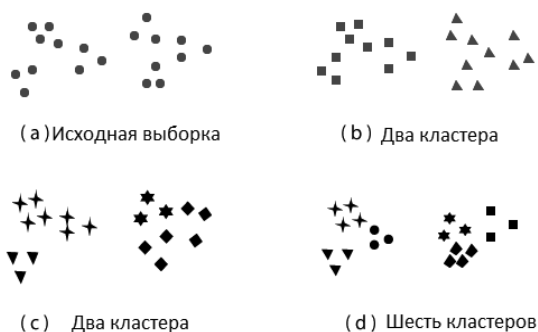


Рис. 8 Различные типы кластеризации

Дизайн маркеров показывает принадлежность к кластеру. На рис. 8 a-d данные разделены на две и шесть секций соответственно. Человеческое зрение позволяет разделить каждый из двух более значимых кластеров на три подкластера. Из рисунка 6 также видно, что значение кластера является неточным. Лучшее определение кластера зависит от характера данных и результатов.

Кластерный анализ - это метод машинного обучения, используемый для выявления некоторой внутренней структуры в наборе данных путем разделения данных на группы (кластеры) таким образом, чтобы точки в одном кластере были более похожи друг на друга, чем на точки из других кластеров.

Кластерный анализ позволяет рассматривать достаточно большой объем информации и резко сокращать, сжимать большие массивы информации, делать их компактными и наглядными.

Ввиду особого положения задачи кластеризации в списке задач интеллектуального анализа данных было разработано множество способов ее решения. Один из них — построение набора характеристических

функций классов, которые показывают, относится ли объект данных к данному классу или нет. Характеристическая функция класса может быть двух типов:

1. Дискретная функция, принимающая одно из двух определенных значений, смысл которых в принадлежности/непринадлежности объекта данных заданному классу.
2. Функция, принимающая вещественные значения, например из интервала 0...1. Чем ближе значение функции к единице, тем больше объект данных принадлежит заданному классу.

К настоящему времени разработано большое количество базовых методов кластеризации. К их числу относят:

- ✓ Методы кластеризации
- ✓ Методы секционирования
- ✓ Иерархическая кластеризация
- ✓ Нечеткая кластеризация
- ✓ Кластеризация на основе плотности
- ✓ Кластеризация на основе моделей

На рис. 9 показаны наиболее часто используемые в ИАД методы кластеризации.

направления
в
кластерном
анализе

Partitioning approach: плоская кластеризация - предполагает разделение объектов на кластеры сразу, причем один объект относится только к одному кластеру
Методы: k-means, k-medoids, CLARANS

Fuzzy approach: Метод нечеткой кластеризации позволяет разбить имеющееся множество объектов p на заданное число нечетких множеств, то есть один и тот же объект может принадлежать разным классам.

Принадлежность характеризуется степенью принадлежности, например вероятностью.

Методы: C-means

Hierarchical approach: Восходящая/нисходящая кластеризация. Иерархическая кластеризация (восходящая) - допускаем наличие подкластеров, осуществляется в несколько приемов, в результате образуется иерархическое дерево (дендрограмма).

Методы: Hierarchical, Diana, Agnes, BIRCH, ROCK, CAMELEON

Density-based approach: Кластеризация на основе плотности. Точки данных в области, разделенные двумя кластерами с низкой плотностью точек, рассматриваются как шум.

Методы: DBSACN, OPTICS, DenClue

Рис. 9. Направления кластерного анализа

Кластерный анализ находит широкое применение в различных областях в рамках интеллектуального анализа данных, включая: маркетинговые исследования и сегментация (используется для исследования поведения потребителей и сегментации рынка, что позволяет компаниям оптимизировать свои продукты и маркетинговые стратегии), медицинские исследования (помогает в выявлении подгрупп пациентов на основе признаков болезни, лечения или реакции на лекарственные средства), финансовая аналитика (применяется для выявления инвестиционных возможностей и выделения групп сходных финансовых инструментов), тренды социальных сетей (используется для выявления групп схожих пользователей или идентификации тематических групп в обсуждениях), анализ геоданных (используется для пространственного анализа и сегментации географических данных, таких как мобильная телефония, геолокационные данные и другие виды геоданных).

При этом есть и оборотная сторона медали - основная проблема алгоритмов кластеризации данных заключается в том, что они не могут быть стандартизированы. Усовершенствованный алгоритм может дать наилучшие результаты с одним типом набора данных, но он может дать сбой или плохо работать с другими типами наборов данных. Несмотря на то, что было предпринято много усилий для стандартизации алгоритмов, которые могут хорошо работать в любых ситуациях, до сих пор не было достигнуто никаких значительных достижений. До сих пор было предложено множество инструментов кластеризации. Однако каждый алгоритм имеет свои преимущества или недостатки и не может работать на всех реальных ситуациях.

Между классификацией и кластеризацией есть сходство, внешне они похожи, но по сути – разные. Основное различие между ними заключается в том, что классификация включает в себя выравнивание элементов в соответствии с их принадлежностью к заранее определенным группам. В табл.4 показана разница между методами.

Таблица 4. Сравнительный анализ классификации и кластеризации в ИАД

Классификация	Кластеризация
Классификация — это контролируемый подход к обучению, при котором машине предоставляется определенная метка для классификации новых наблюдений. В этом случае машина нуждается в	Кластеризация — это подход к обучению без учителя, при котором группировка выполняется на основе сходства.

Классификация	Кластеризация
надлежащем тестировании и обучении для проверки этикетки.	
Контролируемый подход к обучению	Неконтролируемый подход к обучению.
Использует обучающий набор данных	Не использует обучающий набор данных
Использует алгоритмы для категоризации новых данных в соответствии с наблюдениями обучающей выборки	Использует статистические концепции, в которых набор данных разделен на подмножества с одинаковыми признаками.
В классификации есть метки для обучающих данных	В кластеризации нет меток для обучающих данных.
Цель — найти, к какому классу относится новый объект, образуя набор предопределенных классов.	Цель — сгруппировать набор объектов, чтобы выяснить, есть ли между ними какая-либо связь

3. Регрессия

Регрессия относится к методу интеллектуального анализа данных, который используется для прогнозирования числовых значений в заданном наборе данных. Например, регрессия может быть использована для прогнозирования стоимости продукта или услуги или других переменных. Она также используется в различных отраслях для ведения бизнеса и маркетинга, анализа тенденций и финансового прогнозирования.

Регрессия относится к методу контролируемого машинного обучения, который используется для прогнозирования любого атрибута с непрерывным значением.

Регрессия включает в себя технику аппроксимации прямой линии или кривой по многочисленным точкам данных. Это происходит таким образом, что расстояние между точками данных и линией оказывается наименьшим.

Наиболее популярными типами регрессии являются линейная и логистическая регрессии. Кроме того, многие другие типы регрессии могут быть выполнены в зависимости от их производительности на отдельном наборе данных.

Регрессия может предсказывать все зависимые наборы данных, выраженные в выражении независимых переменных, а тренд доступен для конечного периода. Регрессия является хорошим способом прогнозирования переменных, но существуют определенные ограничения и допущения, такие как независимость переменных, присущее нормальным распределениям переменных. Например, предположим, что

мы рассматриваем две переменные, A и B , и их совместное распределение является двумерным распределением, тогда по этой природе. В этом случае эти две переменные могут быть независимыми, но они также коррелируют. Предельные распределения A и B должны быть выведены и использованы. Прежде чем применять регрессионный анализ, необходимо тщательно изучить данные и выполнить определенные предварительные тесты, чтобы убедиться, что регрессия применима. В таких случаях доступны непараметрические тесты.

Типы регрессии

Линейная регрессия. Это тип регрессии, который формирует связь между целевой переменной и одной или несколькими независимыми переменными с помощью прямой

Логистическая регрессия. Когда зависимая переменная является бинарной по своей природе, т.е. 0 и 1, истина или ложь, успех или неудача, возникает метод логистической регрессии. Здесь целевое значение (Y) находится в диапазоне от 0 до 1. В отличие от линейной регрессии, она не нуждается в каких-либо независимых и зависимых переменных, чтобы иметь линейную зависимость

Регрессия Лассо. Термин LASSO расшифровывается как Least Absolute Shrinkage and Selection Operator (оператор наименьшего абсолютного сжатия и выбора). Регрессия лассо — это линейный тип регрессии, использующий сжатие. При регрессии Лассо все точки данных сжимаются к центральной точке, также известной как среднее значение. Процесс лассо лучше всего подходит для простых и разреженных моделей с меньшим количеством параметров, чем другие регрессии. Этот тип регрессии хорошо подходит для моделей, страдающих мультиколлинеарностью

Гребневая регрессия. Ridge-регрессия относится к процессу, который используется для анализа различных регрессионных данных, имеющих проблему мультиколлинеарности. Мультиколлинеарность — это существование линейной корреляции между двумя независимыми переменными.

Ridge-регрессия существует, когда оценки наименьших квадратов наименее смещены с высокой дисперсией, поэтому они сильно отличаются от реального значения.

Полиномиальная регрессия. Если степень независимой переменной в уравнении регрессии больше 1, оно называется полиномиальным уравнением.

Рис. 10. Виды регрессии

Регрессия и классификация очень похожи друг на друга. Классификация и регрессия — это две важные проблемы прогнозирования, которые используются в интеллектуальном анализе данных. Если вы дали обучающий набор входных и выходных данных и изучили функцию, которая связывает их, то, это позволит вам прогнозировать выходные данные на основе новых данных. Единственное отличие состоит в том, что при классификации выходы дискретны, тогда

как в регрессии выходы — нет. Но понятия размыты, как в «логистической регрессии», которую можно интерпретировать либо как классификацию, либо как регрессионный метод. Поэтому, зачастую, сложно понять, когда использовать классификацию, а когда регрессию.

Таблица 5. Сравнительный анализ регрессии и классификации в ИАД

Регрессия	Классификация
Регрессия относится к методу контролируемого машинного обучения, который используется для прогнозирования любого атрибута с непрерывным значением.	Классификация относится к процессу присвоения предопределенных меток классам экземплярам на основе их атрибутов.
В регрессии характер прогнозируемых данных упорядочен.	При классификации характер предикативных данных является неупорядоченным.
Регрессию можно разделить на линейную регрессию и нелинейную регрессию.	Классификация делится на две категории: бинарный классификатор и мультиклассовый классификатор.
В процессе регрессии вычисления в основном выполняются с использованием среднеквадратической ошибки.	В процессе классификации расчеты в основном выполняются путем измерения эффективности.
Примерами регрессий являются дерево регрессии, линейная регрессия и т.д.	Примерами классификаций являются дерево решений.

4. Поиск ассоциативных правил

Одной из наиболее распространенных задач анализа данных является определение часто встречающихся наборов объектов в большом множестве наборов.

Этот метод интеллектуального анализа данных помогает обнаружить связь между двумя или более элементами. Он находит скрытый шаблон в наборе данных.

Ассоциативные правила — это операторы «если-то», которые помогают показать вероятность взаимодействия между элементами данных в больших наборах данных в различных типах баз данных. Интеллектуальный анализ ассоциативных правил имеет несколько применений и обычно используется для корреляции продаж в наборах данных или медицинских данных.

Впервые задача поиска ассоциативных правил (association rule mining) была предложена для нахождения типичных шаблонов покупок,

совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины (market basket analysis).

Анализ потребительской корзины смоделирован на основе интеллектуального анализа ассоциативных правил, т.е. конструкции IF {}, THEN {}. Например, если покупатель покупает хлеб, то он, скорее всего, купит и масло.

Ассоциативные правила обычно представляются в виде: {Хлеб} -> {Масло}

Термины, употребляемые в данном методе:

✓ Предшествующий. Элементы или «наборы элементов», найденные в данных, являются предшествующими. Проще говоря, это компонент IF, написанный в левой части. В приведенном выше примере хлеб является antecedentом.

✓ Последовательный. Консеквент — это элемент или набор элементов, найденных в сочетании с antecedentом. Это компонент THEN, написанный в правой части. В приведенном выше примере последствием является сливочное масло.

Методы анализа потребительской корзины можно классифицировать в зависимости от того, как используются имеющиеся данные.

При анализе потребительской корзины ассоциативные правила используются для прогнозирования вероятности покупки продуктов вместе. Ассоциативные правила подсчитывают частоту элементов, которые встречаются вместе, стремясь найти связи, которые встречаются гораздо чаще, чем ожидалось.

К алгоритмам, использующим ассоциативные правила, относятся AIS, SETM и Apriori. Алгоритм Apriori часто цитируется специалистами по обработке и анализу данных в исследовательских статьях об анализе потребительской корзины. Он определяет часто встречающиеся элементы в базе данных, а затем оценивает их частоту по мере расширения наборов данных до больших размеров.

Виды анализа потребительской корзины

Описательный анализ потребительской корзины. Этот тип извлекает аналитические сведения только из прошлых данных и является наиболее часто используемым подходом. Анализ здесь не делает никаких прогнозов, а оценивает связь между продуктами с помощью статистических методов. Для тех, кто знаком с основами анализа данных, этот тип моделирования известен как обучение без учителя.

Предиктивный анализ потребительской корзины. В этом типе используются модели обучения с учителем, такие как классификация и регрессия. По сути, он направлен на то, чтобы имитировать рынок, чтобы проанализировать, что вызывает то, что происходит. По сути, он рассматривает товары, приобретенные в определенной последовательности, чтобы определить перекрестные продажи. Например, покупка расширенной гарантии с большей вероятностью последует за покупкой iPhone. Несмотря на то, что он не так широко используется, как описательный MBA, он по-прежнему является очень ценным инструментом для маркетологов

Дифференциальный анализ потребительской корзины. Этот вид анализа выгоден для анализа конкурентов. Он сравнивает историю покупок между магазинами, между сезонами, между двумя временными периодами, между разными днями недели и т. д., чтобы найти интересные закономерности в поведении потребителей. Например, это может помочь определить, почему некоторые пользователи предпочитают покупать один и тот же продукт по одной и той же цене на Amazon по сравнению с Flipkart. Ответ может заключаться в том, что у реселлера Amazon больше складов и он может доставлять быстрее, или, может быть, что-то более глубокое, например, пользовательский опыт.

Рис. 11. Виды анализа потребительской корзины

5. Анализ выбросов

Этот тип метода интеллектуального анализа данных относится к наблюдению за элементами данных в наборе данных, которые не соответствуют ожидаемому шаблону или ожидаемому поведению. Этот метод может быть использован в различных областях, таких как вторжение, обнаружение, обнаружение мошенничества и т.д. Он также известен как анализ выбросов или майнинг Outlier. Выброс — это точка данных, которая слишком сильно отличается от остальной части набора данных. Большинство реальных наборов данных имеют выбросы. Обнаружение выбросов играет важную роль в области интеллектуального анализа данных. Обнаружение выбросов полезно во многих областях, таких как идентификация перебоев в сети, обнаружение мошенничества с кредитными или дебетовыми картами, обнаружение искажений в сетевых данных беспроводных датчиков и т.д. Любая нежелательная ошибка возникает в какой-либо ранее измеренной переменной, или есть какое-

либо отклонение в ранее измеренной переменной, называемое шумом. Прежде чем находить выбросы, присутствующие в любом наборе данных, рекомендуется сначала удалить шум.

Типы выбросов

Глобальные выбросы. Глобальные выбросы также называются точечными выбросами. Глобальные выбросы рассматриваются как простейшая форма выбросов. Когда точки данных отклоняются от всех остальных точек данных в данном наборе данных, это называется глобальным выбросом. В большинстве случаев все процедуры обнаружения выбросов направлены на определение глобальных выбросов.

Коллективные выбросы. В заданном наборе данных, когда группа точек данных отклоняется от остальной части набора данных, это называется коллективными выбросами. В этом случае конкретный набор объектов данных не может быть выбросами, но при рассмотрении объектов данных в целом они могут вести себя как выбросы. Чтобы определить типы различных выбросов, необходимо просмотреть справочную информацию о взаимосвязи между поведением выбросов, отображаемых различными объектами данных. Например, в системе обнаружения вторжений перенос DOS-пакета из одной системы в другую воспринимается как нормальное поведение. Поэтому, если это происходит с различными компьютерами одновременно, это считается ненормальным поведением, и в целом они называются коллективными выбросами.

Контекстные выбросы. Как следует из названия, «контекстуальный» означает выброс, введенный в контекст. Например, в технике распознавания речи используется одиночный фоновый шум. Контекстные выбросы также известны как условные выбросы. Эти типы выбросов возникают, если объект данных отклоняется от других точек данных из-за какого-либо конкретного условия в данном наборе данных. Как известно, существует два типа атрибутов объектов данных: контекстуальные атрибуты и поведенческие атрибуты. Контекстный анализ выбросов позволяет пользователям исследовать выбросы в различных контекстах и условиях, что может быть полезно в различных приложениях. Например, значение температуры 45 градусов по Цельсию может вести себя как выброс в сезон дождей. Тем не менее, он будет вести себя как обычная точка данных в контексте летнего сезона.

Рис. 12. Типы выбросов

6. Последовательные паттерны

Последовательный шаблон — это метод интеллектуального анализа данных, предназначенный для оценки последовательных данных с целью обнаружения последовательных закономерностей. Он заключается в нахождении интересных подпоследовательностей в наборе последовательностей, где доля последовательности может быть измерена с точки зрения различных критериев, таких как длина, частота появления и т. д.

Другими словами, этот метод интеллектуального анализа данных помогает обнаруживать или распознавать схожие закономерности в данных транзакций в течение некоторого времени.

7. Прогнозирование

При прогнозировании используется комбинация других методов интеллектуального анализа данных, таких как тренды, кластеризация, классификация и т. д.

Предиктивная аналитика — это анализ, выполняемый для прогнозирования будущего события с использованием предыдущих данных. Это процесс извлечения информации из существующих наборов данных для поиска полезной информации, тенденций и прогнозирования будущих событий. Предиктивная аналитика не говорит точно, что произойдет в будущем. Он предсказывает, что может произойти в будущем.

5. KDD- Обнаружение знаний в базах данных

Термин KDD расшифровывается как Knowledge Discovery in Data bases (обнаружение знаний в базах данных). Он относится к широкой процедуре обнаружения знаний в данных и подчеркивает высокоуровневое применение конкретных методов интеллектуального анализа данных.

Основной целью процесса KDD является извлечение информации из данных в контексте больших баз данных. Это достигается с помощью алгоритмов интеллектуального анализа данных для определения того, что считается знанием.

Обнаружение знаний в базах данных рассматривается как запрограммированный исследовательский анализ и моделирование обширных хранилищ данных. KDD — это организованная процедура распознавания валидных, полезных и понятных шаблонов из огромных и сложных наборов данных. Интеллектуальный анализ данных — это корень процедуры KDD, включающий вывод алгоритмов, которые исследуют данные, разрабатывают модель и находят ранее неизвестные закономерности. Модель используется для извлечения знаний из данных, анализа данных и прогнозирования данных.

Процесс KDD

Процесс обнаружения знаний (рис. 13) является итеративным и интерактивным, состоит из девяти шагов. Процесс является итеративным на каждом этапе, что означает, что может потребоваться возврат к предыдущим действиям. Этот процесс имеет много творческих аспектов в том смысле, что невозможно представить одну формулу или сделать полную научную категоризацию для правильных решений для каждого

шага и типа приложения. Таким образом, необходимо понимать процесс, а также различные требования и возможности на каждом этапе.

Процесс начинается с определения целей КДД и заканчивается внедрением полученных знаний. В этот момент цикл замыкается, и начинается активный интеллектуальный анализ данных. Впоследствии необходимо будет внести изменения в домен приложения. Например, предложение различных функций для пользователей мобильных телефонов, чтобы уменьшить отток клиентов. Это замыкает цикл, и затем изменяется влияние на новые репозитории данных и на процесс КДД.

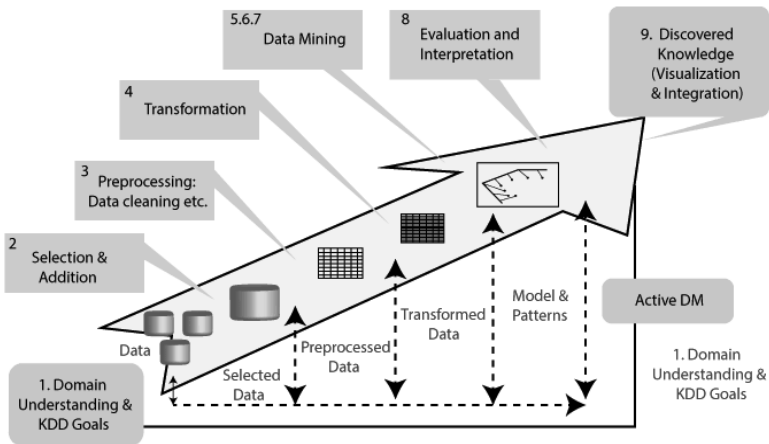


Рис. 13. Процесс КДД

Ниже приведено краткое описание девятиступенчатого процесса КДД, начиная с управленческого этапа:

1. Формирование понимания предметной области приложения. Это начальный предварительный шаг. Он развивает пониманию того, что должно быть сделано с различными решениями, такими как преобразование, алгоритмы, представление и т.д. Лица, отвечающие за процесс КДД, должны понимать и характеризовать цели конечного пользователя и среду, в которой будет происходить процесс обнаружения знаний (включает в себя соответствующие предварительные знания).

2. Выбор и создание набора данных, на котором будет производиться обнаружение. После определения целей необходимо определить данные, которые будут использоваться для процесса обнаружения знаний. Это включает в себя определение доступных данных, получение важных данных, а затем интеграцию всех данных для

обнаружения знаний в один набор включает в себя качества, которые будут учитываться в процессе. Этот процесс важен, так как интеллектуальный анализ данных обучается и обнаруживает на основе доступных данных. Это и есть доказательная база для построения моделей. Если на этом этапе отсутствуют какие-то значимые атрибуты, то все исследование может быть неудачным в этом отношении, чем больше атрибутов рассматривается. С другой стороны, организация, сбор и эксплуатация современных хранилищ данных обходится дорого, и существует договоренность с возможностью наилучшего понимания явлений. Эта схема относится к аспекту, в котором имеет место интерактивный и итеративный аспект KDD. Он начинается с наилучших доступных наборов данных, а затем расширяется и наблюдает за влиянием с точки зрения обнаружения знаний и моделирования.

3. Предварительная обработка и очистка. На этом этапе повышается надежность данных. Он включает в себя очистку данных, например, обработку недостающих величин и удаление шума или выбросов. Он может включать в себя сложные статистические методы или использовать алгоритм интеллектуального анализа данных в этом контексте. Например, если кто-то подозревает, что определенный атрибут недостаточно надежен или имеет много отсутствующих данных, на этом этапе этот атрибут может превратиться в цель алгоритма интеллектуального анализа данных с учителем. Для этих атрибутов будет создана модель прогнозирования, после чего можно будет спрогнозировать отсутствующие данные. Расширение, на которое обращают внимание на этот уровень, зависит от множества факторов.

4. Трансформация данных. На этом этапе подготавливается и разрабатывается создание соответствующих данных для интеллектуального анализа данных. Методы включают в себя уменьшение размерности (например, выбор и извлечение признаков, а также выборку записей), а также преобразование атрибутов (например, дискретизацию числовых атрибутов и функциональное преобразование). Этот шаг может иметь важное значение для успеха всего проекта KDD, и, как правило, он очень специфичен для конкретного проекта. Например, при медицинском обследовании наиболее значимым фактором часто может быть частное атрибутов, а не каждое из них в отдельности. В бизнесе нам, возможно, придется думать о последствиях, находящихся вне нашего контроля, а также об усилиях и временных проблемах. Например, изучение влияния

аккумуляции рекламы. Однако, если мы не используем правильную трансформацию в начале, то мы можем получить удивительный эффект, который даст нам представление о преобразовании, необходимом в следующей итерации. Таким образом, процесс KDD следует сам за собой и подталкивает к пониманию требуемой трансформации.

5. Прогноз и описание. Теперь мы готовы решить, какой тип интеллектуального анализа данных использовать, например, классификацию, регрессию, кластеризацию и т.д. Это в основном зависит от целей KDD, а также от предыдущих шагов. В интеллектуальном анализе данных есть две важные цели: первая — прогнозирование, а вторая — описание. Прогнозирование обычно называют интеллектуальным анализом данных с учителем, в то время как описательный интеллектуальный анализ данных включает в себя аспекты интеллектуального анализа данных без учителя и визуализации. Большинство методов интеллектуального анализа данных основаны на индуктивном обучении, при котором модель строится явно или неявно путем обобщения достаточного числа готовящихся моделей. Фундаментальное допущение индуктивного подхода состоит в том, что подготовленная модель применима к будущим случаям. Методика также учитывает уровень метаобучения для конкретного набора доступных данных.

6. Выбор алгоритма интеллектуального анализа данных. Этот этап включает в себя выбор конкретной техники, которая будет использоваться для поиска паттернов, включающих несколько индукторов. Например, если сравнивать точность и понятность, то первое лучше подходит для нейронных сетей, а второе — для деревьев решений. Для каждой системы метаобучения существует несколько возможностей того, как она может быть успешной. Метаобучение фокусируется на выяснении того, что заставляет алгоритм интеллектуального анализа данных быть эффективным или неэффективным в конкретной проблеме. Таким образом, данная методология пытается понять ситуацию, в которой алгоритм интеллектуального анализа данных является наиболее подходящим. Каждый алгоритм имеет параметры и стратегии наклонения, такие как десятикратная перекрестная проверка или другое деление для обучения и тестирования.

7. Использование алгоритма интеллектуального анализа данных. Наконец-то достигнута реализация алгоритма интеллектуального анализа

данных. На этом этапе нам может потребоваться использовать алгоритм несколько раз, пока не будет получен удовлетворительный результат.

8. Оценка. На этом этапе мы оцениваем и интерпретируем найденные шаблоны, правила и надежность для достижения цели, описанной на первом этапе. Здесь мы рассмотрим этапы предварительной обработки с точки зрения их влияния на результаты алгоритма интеллектуального анализа данных. Например, включить объект в шаг 4 и повторить с этого момента. На этом этапе основное внимание уделяется понятности и полезности индуцированной модели. На этом этапе выявленные знания также записываются для дальнейшего использования. Последним шагом является использование, а также общие отзывы и результаты обнаружения, полученные с помощью интеллектуального анализа данных.

9. Использование полученных знаний. Теперь мы готовы включить эти знания в другую систему для дальнейшей деятельности. Знания становятся эффективными в том смысле, что мы можем вносить изменения в систему и измерять их воздействие. Выполнение этого шага определяет эффективность всего процесса KDD. На этом этапе возникает множество проблем, таких как потеря «лабораторных условий», в которых мы работали. Например, знание было обнаружено из некоего статического изображения, обычно это набор данных, но теперь данные становятся динамическими. Структуры данных могут изменять определенные величины, которые становятся недоступными, а домен данных может быть изменен, например атрибут, который может иметь значение, которое ранее не ожидалось.

6. Развитие методологий анализа данных

Хотя корни сбора данных могут быть прослежены до конца 1980-х, в течение большинства 1990-х, область была все еще в ее младенчестве. Интеллектуальный анализ данных все еще определялся и совершенствовался. Это было, в основном, свободное скопление моделей данных, аналитических алгоритмов и специальной продукции.

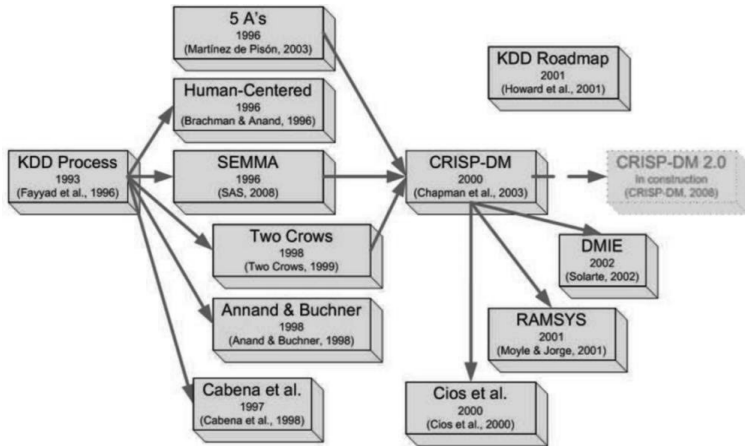


Рис. 14. Эволюция методологий ИАД

В 1999 заинтересованные в стандартизации компании (автомобильная компания Daimler-Benz, страховая компания OHRA, разработчик аппаратного и программного обеспечения NCR Corp, разработчик статистического программного обеспечения SPSS, Inc) начали сотрудничать, чтобы формализовать и стандартизировать подход к сбору данных. Результатом их работы был кросс-индустриальный стандарт глубинного анализа данных (CRISP-DM, the CRoss-Industry Standard Processfor Data Mining).

Хотя у участников создания CRISP-DM, конечно, были имущественные права в определенных инструментах программного и аппаратного обеспечения, процесс был разработан независимым от любого определенного инструмента или вида данных.

Межотраслевой стандартный процесс интеллектуального анализа данных (CRISP-DM) состоит из шести этапов, разработанных в виде циклического метода, как показано на рис. 15.

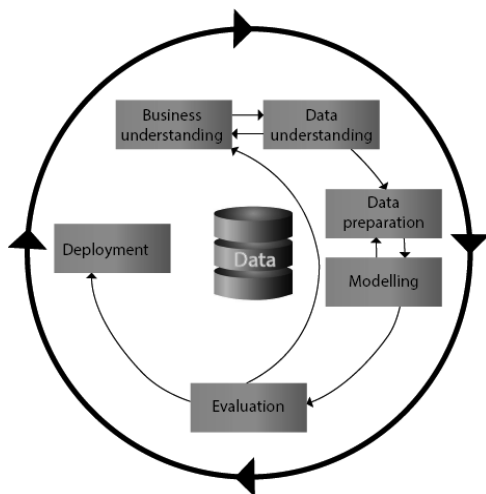


Рис.15. Этапы процесса анализа данных по стандарту CRISP-DM

1. Понимание бизнеса. Он фокусируется на понимании целей и требований проекта с точки зрения бизнеса, а затем преобразовании этой информации в задачу интеллектуального анализа данных, а затем в предварительный план, разработанный для достижения цели.

Задачи:

- ✓ Определение бизнес-целей
- ✓ Ситуация с доступом
- ✓ Определение целей интеллектуального анализа данных
- ✓ Разработка плана проекта

2. Понимание данных. Понимание данных начинается с первоначального сбора данных и продолжается операциями по ознакомлению с данными, проблемам качества данных, поиску лучшего понимания данных или обнаружению интересных подмножеств для гипотезы скрытой информации.

Задачи:

- ✓ Собрать исходные данные
- ✓ Описание данных
- ✓ Изучение данных
- ✓ Проверка качества данных

3. Подготовка данных. Обычно это занимает более 90 процентов времени. Он охватывает все операции по построению окончательного набора данных из исходной необработанной информации.

Подготовка данных, скорее всего, будет производиться несколько раз, а не в каком-либо установленном порядке.

Задачи:

- ✓ Выберите данные
- ✓ Чистые данные
- ✓ Создание данных
- ✓ Интеграция данных
- ✓ Форматирование данных

4. Моделирование. При моделировании подбираются и применяются различные методы моделирования, а их параметры измеряются до оптимальных значений. Некоторые методы предъявляли особые требования к форме данных. Поэтому необходимо вернуться к этапу подготовки данных.

Задачи:

- ✓ Выберите технику моделирования
- ✓ Создание тестового дизайна
- ✓ Модель построения
- ✓ Модель доступа

5. Оценка. На последнем этапе необходимо принять решение об использовании результатов интеллектуального анализа данных.

Основная цель оценки – определить какую-либо существенную бизнес-проблему, которая не была рассмотрена должным образом.

На последнем этапе этого этапа должно быть принято решение об использовании результатов интеллектуального анализа данных.

Задачи:

- ✓ Оценка результатов
- ✓ Процесс рецензирования
- ✓ Определение дальнейших действий.

6. Развертывание. Развертывание относится к тому, как необходимо использовать результаты. Оно включает в себя оценку базы данных, использование результатов в качестве рекомендаций компании, интерактивный интернет-скоринг.

Полученная информация должна быть организована и представлена таким образом, чтобы клиент мог ее использовать.

Задачи:

- ✓ Планирование развертывания
- ✓ Планирование мониторинга и обслуживания

- ✓ Подготовка итогового отчета
- ✓ Обзор проекта

7. Архитектура интеллектуального анализа данных

Важными компонентами систем интеллектуального анализа данных являются источник данных, механизм интеллектуального анализа данных, сервер хранилища данных, модуль оценки шаблонов, графический пользовательский интерфейс и база знаний.

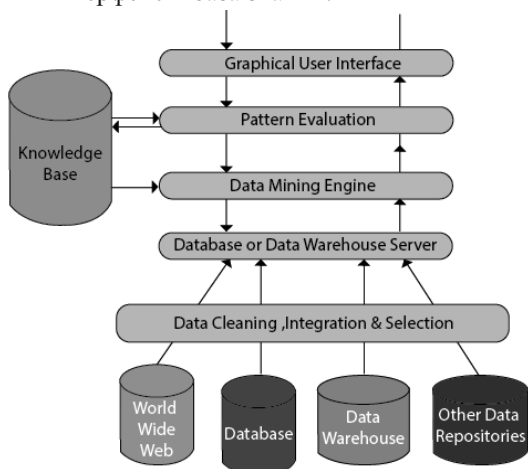


Рис. 16. Архитектура ИАД

Источник данных. Фактическим источником данных является база данных, хранилище данных, Всемирная паутина (WWW), текстовые файлы и другие документы. Для успешного интеллектуального анализа данных требуется огромное количество исторических данных. Организации обычно хранят данные в базах данных или хранилищах данных. Хранилища данных могут состоять из одной или нескольких баз данных, текстовых файлов, электронных таблиц или других хранилищ данных. Иногда информация может содержаться даже в обычных текстовых файлах или электронных таблицах. Другим основным источником данных является Всемирная паутина или Интернет.

Различные процессы. Перед передачей данных в базу данных или на сервер хранилища данных данные должны быть очищены, интегрированы и выбраны. Поскольку информация поступает из разных источников и в разных форматах, она не может быть использована непосредственно для процедуры интеллектуального анализа данных, так как данные могут быть

неполными и неточными. Итак, первые данные необходимо очистить и унифицировать. Из различных источников данных будет собрано больше информации, чем необходимо, и только интересующие данные должны быть отобраны и переданы на сервер. Эти процедуры не так просты, как мы думаем. С данными можно выполнить несколько методов: выбор, интеграцию и очистку.

Сервер базы данных или хранилища данных. База данных или сервер хранилища данных состоит из исходных данных, готовых к обработке. Таким образом, сервер является причиной получения соответствующих данных, основанных на интеллектуальном анализе данных по запросу пользователя.

Механизм интеллектуального анализа данных. Подсистема интеллектуального анализа данных является основным компонентом любой системы интеллектуального анализа данных. Он содержит несколько модулей для выполнения задач интеллектуального анализа данных, включая ассоциацию, характеристику, классификацию, кластеризацию, прогнозирование, анализ временных рядов и т. д.

Другими словами, *можно сказать, что интеллектуальный анализ данных является корнем нашей архитектуры интеллектуального анализа данных.* Он включает в себя инструменты и программное обеспечение, используемые для получения информации и знаний из данных, собранных из различных источников данных и хранящихся в хранилище данных.

Модуль оценки паттернов. Модуль Оценка шаблона в первую очередь отвечает за меру исследования шаблона с помощью порогового значения. Он взаимодействует с механизмом интеллектуального анализа данных, чтобы сосредоточить поиск на интересных закономерностях.

В этом сегменте обычно используются меры ставок, которые взаимодействуют с модулями интеллектуального анализа данных, чтобы сосредоточить поиск на интересных закономерностях. Он может использовать пороговое значение стейкинга для фильтрации обнаруженных закономерностей. С другой стороны, модуль оценки закономерностей может быть скоординирован с модулем интеллектуального анализа данных в зависимости от реализации используемых методов интеллектуального анализа данных.

Графический интерфейс пользователя. Модуль графического пользовательского интерфейса (GUI) обеспечивает связь между системой интеллектуального анализа данных и пользователем. Этот модуль

помогает пользователю легко и эффективно использовать систему, не зная сложности процесса. Этот модуль взаимодействует с системой интеллектуального анализа данных, когда пользователь задает запрос или задачу и отображает результаты.

База знаний. База знаний полезна во всем процессе интеллектуального анализа данных. Это может быть полезно для руководства поиском или оценки доли шаблонов результатов. База знаний может даже содержать пользовательские представления и данные из пользовательского интерфейса, которые могут быть полезны в процессе интеллектуального анализа данных. Подсистема интеллектуального анализа данных может получать входные данные из базы знаний, чтобы сделать результат более точным и надежным. Модуль оценки закономерностей регулярно взаимодействует с базой знаний для получения входных данных, а также для ее обновления.

8. Тенденции в интеллектуальном анализе данных

Интеллектуальный анализ данных является одним из наиболее широко используемых методов извлечения данных из различных источников и их организации для более эффективного использования. Несмотря на наличие различных коммерческих систем интеллектуального анализа данных, при их фактическом внедрении возникает множество проблем. Ожидается, что в связи с быстрым развитием в области интеллектуального анализа данных компании будут идти в ногу со всеми новыми разработками.

Типы последовательности интеллектуального анализа данных в ИАД

1. Временные ряды майнинга. Определенное количество точек данных записывается в определенное время или события, полученные в результате повторных измерений времени во временных рядах интеллектуального анализа. Значения или данные обычно измеряются через равные промежутки времени, например ежечасно, еженедельно или ежедневно. Данные временных рядов также записываются через равные промежутки времени, или характерными компонентами временных рядов являются трендовые, сезонные, циклические или нерегулярные.

Применение временных рядов:

- ✓ Финансы: Анализ фондового рынка
- ✓ Отрасль: Потребляемая мощность
- ✓ Научный: Результат эксперимента

✓ Метеорологические: Осадки

Методы анализа временных рядов:

Анализ тенденций. Категории движений временных рядов:

✓ Долгосрочные или трендовые движения: Общее направление, в котором движется временной ряд на длительном интервале времени.

✓ Циклические движения. Долгосрочное колебание относительно линии тренда или кривой.

✓ Сезонные движения. Временные ряды, по-видимому, следуют по существу идентичным закономерностям в течение соответствующих месяцев последующих лет.

✓ Нерегулярные или случайные движения. Изменения, которые происходят случайным образом из-за незапланированных событий.

Поиск сходства:

✓ Сокращение объема данных

✓ Методы индексирования

✓ Методы поиска сходства

✓ Языки запросов

2. Интеллектуальный анализ символьной последовательности.

Символьная последовательность включает в себя упорядоченный список элементов, которые могут быть записаны с чувством времени или без него. Эта последовательность может быть использована различными способами, включая последовательности покупок потребителей, веб-истории посещений, последовательности выполнения программного обеспечения, биологические последовательности и т. д.

Последовательные шаблоны интеллектуального анализа включают в себя идентификацию подпоследовательностей, которые часто встречаются в одной или нескольких последовательностях. В результате существенных исследований в этой области было разработано множество масштабируемых алгоритмов. В качестве альтернативы мы можем добывать только множество замкнутых последовательных паттернов, где последовательный паттерн s является замкнутым, если правильная подпоследовательность s' и s' имеет ту же поддержку, что и s .

3. Биологическая последовательность. Биологические последовательности состоят из нуклеотидных или аминокислотных последовательностей. Анализ биологических последовательностей

сравнивает, выравнивает, индексирует и анализирует биологические последовательности в биоинформатике и современной биологии. Анализ биологических последовательностей играет важнейшую роль в биоинформатике и современной биологии. Такой анализ может быть разделен на попарное выравнивание последовательностей и множественное выравнивание последовательностей.

Методы биологического секвенирования:

- ✓ Выравнивание биологических последовательностей:
- ✓ Попарное выравнивание
- ✓ Алгоритм локального выравнивания BLAST
- ✓ Несколько методов выравнивания последовательностей

Анализ биологических последовательностей с использованием скрытой марковской модели:

- ✓ Цепь Маркова
- ✓ Скрытая марковская модель
- ✓ Форвардный алгоритм
- ✓ Алгоритм Витерби
- ✓ Алгоритм Баума-Уэлча

Применение интеллектуального анализа данных.

Анализ финансовой информации.

✓ Прогнозирование платежей по кредиту/анализ потребительской кредитной политики

✓ Проектирование и строительство информационного хранилища

✓ Финансовая информация, собранная в территориальных подразделениях банков и денежных учреждений, как правило, является сравнительно полной, надежной и высококачественной.

Розничная торговля:

✓ Многомерный анализ (продажи, клиенты, продукты, время и т.д.)

✓ Анализ кампаний по продажам

✓ Удержание клиентов

✓ Рекомендации по продуктам

✓ Использование средств визуализации для анализа данных

Наука и техника:

✓ Обработка данных и хранилища данных

✓ Типы данных комплекса «Майнинг»

- ✓ Сетевой майнинг
- ✓ Интеллектуальный анализ на основе графов.

Концепции интеллектуального анализа данных все время развиваются, и вот следующие последние тенденции, такие как:

1. Исследование приложений. Интеллектуальный анализ данных все чаще используется для изучения приложений в других областях, таких как финансовый анализ, телекоммуникации, биомедицина, безопасность беспроводных сетей и наука.

2. Интеллектуальный анализ мультимедийных данных. Это один из новейших методов, который набирает обороты из-за растущей способности точно собирать полезные данные. Он включает в себя извлечение данных из различных видов мультимедийных источников, таких как аудио, текст, гипертекст, видео, изображения и т. д. Данные преобразуются в числовое представление в различных форматах. Этот метод можно использовать для кластеризации и классификации, выполнения проверок сходства и выявления ассоциаций.

3. Повсеместный интеллектуальный анализ данных. Этот метод предполагает сбор данных с мобильных устройств для получения информации о физических лицах. Несмотря на наличие ряда проблем такого типа, таких как сложность, конфиденциальность, стоимость и т.д., этот метод имеет много возможностей для использования в различных отраслях, особенно при изучении человеко-компьютерных взаимодействий.

4. Распределенный интеллектуальный анализ данных. Этот тип интеллектуального анализа данных набирает популярность, так как включает в себя добычу огромного количества информации, хранящейся в разных локациях компаний или в разных организациях. Сложные алгоритмы используются для извлечения данных из разных мест и предоставления надлежащих аналитических сведений и отчетов на их основе.

5. Встроенный интеллектуальный анализ данных. Функции интеллектуального анализа данных все чаще находят свое применение во многих сценариях использования корпоративного программного обеспечения, от прогнозирования продаж на платформах CRM SaaS до обнаружения киберугроз в системах обнаружения/предотвращения вторжений. Внедрение интеллектуального анализа данных в программные приложения для вертикального рынка позволяет

прогнозировать результаты для любого количества отраслей и открывает новые возможности для создания уникальной стоимости.

6. Интеллектуальный анализ пространственных и географических данных. Этот новый трендовый тип интеллектуального анализа данных включает в себя извлечение информации из экологических, астрономических и географических данных, включая изображения, сделанные из космоса. Этот тип интеллектуального анализа данных может выявить различные аспекты, такие как расстояние и топология, которые в основном используются в географических информационных системах и других навигационных приложениях.

7. Интеллектуальный анализ данных временных рядов и последовательностей. Основное применение этого типа интеллектуального анализа данных заключается в изучении циклических и сезонных тенденций. Эта практика также полезна при анализе даже случайных событий, которые происходят вне нормальной последовательности событий. Розничные компании в основном используют этот метод для получения доступа к покупательским привычкам и поведению клиентов.

8. Доминирование интеллектуального анализа данных в фармацевтической промышленности и здравоохранении. Фармацевтическая промышленность и здравоохранение уже давно являются новаторами в области интеллектуального анализа данных. Недавняя быстрая разработка вакцин против коронавируса напрямую связана с достижениями в области методов интеллектуального анализа данных фармацевтических испытаний, в частности, обнаружения сигналов в процессе клинических испытаний новых лекарств. В здравоохранении специализированные методы интеллектуального анализа данных используются для анализа последовательностей ДНК для создания индивидуальных методов лечения, постановки более обоснованных диагнозов и многого другого.

9. Повышение автоматизации в интеллектуальном анализе данных. Современные решения для интеллектуального анализа данных, как правило, интегрируют машинное обучение и хранилища больших данных, чтобы обеспечить расширенные функции управления данными наряду со сложными методами анализа данных. Более ранние воплощения интеллектуального анализа данных включали в себя ручное кодирование специалистами с глубоким опытом в области статистики и

программирования. Современные методы в высокой степени автоматизированы, а искусственный интеллект и машинное обучение заменяют большинство этих ранее ручных процессов для разработки алгоритмов обнаружения закономерностей.

10. Консолидация поставщиков интеллектуального анализа данных.

Если судить по истории, то значительная консолидация продуктов в области интеллектуального анализа данных неизбежна, поскольку более крупные поставщики баз данных приобретают стартапы, занимающиеся инструментами интеллектуального анализа данных, чтобы дополнить свои предложения новыми функциями. Текущий фрагментированный рынок и широкий спектр игроков в области интеллектуального анализа данных напоминают смежный ландшафт поставщиков больших данных, который продолжает подвергаться консолидации.

11. Интеллектуальный анализ биологических данных.

Анализ последовательностей ДНК и белков, анализ многомерных данных микрочипов, анализ биологических путей и сетей, анализ связей между гетерогенными биологическими данными и интеграция биологических данных путем интеллектуального анализа данных являются интересными темами для исследований в области интеллектуального анализа биологических данных.

СПИСОК ЛИТЕРАТУРЫ

1. Барсеян А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсеян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с.
2. Белых Д.А. Прикладные методы управления сложными системами и объектами в условиях неопределенности / Д.А. Белых // Мягкие измерения и вычисления. – 2018. – № 9(10). – С. 73-80
3. Вагизов М.Р., Новикова М.А., Степанов С.Ю., Петров Я.А., Елисеев И.В., Бойцов А.К., Новиков Я.А. Информационные технологии и системы : учеб. пособие для студентов высш. учеб. зав. Ч. 1 / ; Минобрнауки России. — СПб., 2020. — 78 с.
4. Вагизов М.Р., Новикова М.А., Петров Я.А., Елисеев И.В., Степанов С.Ю., Бойцов А.К. Информационные технологии и системы : учеб. пособие для студентов высш. учеб. зав. Ч. 2 / ; Минобрнауки России. — СПб., 2022. — 74 с.
5. Власов А.Б., Мещеряков В.И., Перунов А.Г., Салов С.В. Обоснование оптимального варианта принятия решений при управлении силами и средствами, действующих в условиях неопределенности // Современные проблемы науки и образования. – 2015. – № 2-1.
6. Волков В.Ю. Нечеткая когнитивная карта как модель сложной системы управления / В.Ю. Волков, В.В. Волкова // Известия Тульского государственного университета. Технические науки. – 2017. – № 12-1. – С. 17-24
7. Гаврилова Т.А. Базы знаний интеллектуальных систем : учебник / Т.А. Гаврилова, В.Ф. Хорошевский. – Санкт-Петербург: Питер, 2001. – 384 с
8. Замятин А.В. Интеллектуальный анализ данных : учеб.пособие. – Томск: Издательский Дом Томского государственного университета, 2016. – 120 с.
9. Задача принятия решения // Лаборатория системного анализа. – 2018. URL: https://systems-analysis.ru/decision_theory_model.html (дата обращения: 23.01.2024г.).
10. Колченко О. Эволюция процессоров. Официальный сайт: URL <https://www.ferra.ru/review/computers/processor-evolution-part-1.htm> Дата обращения: 14.11.2023
11. Виды популярных архитектур процессоров. Официальный сайт: URL <https://tproger.ru/articles/processors-architectures-review> Дата обращения: 14.11.2023
12. Курзаева Л.В. Конкурентоспособность будущих специалистов по информационным технологиям: феномен, модель и методика развития в процессе профессиональной подготовки в вузе / Л.В. Курзаева. – Магнитогорск: Магнитогорский государственный технический университет им. Г.И. Носова, 2010. – 160 с.
13. Лоран Лелу. Блокчейн от А до Я. Все о технологии десятилетия/ Groupe Eyrolles.- France, Paris – 2017. – с.375
14. Математическая модель задачи выбора решений // Национальный открытый университет ИНТУИТ. – 2023. URL: https://intuit.ru/studies/educational_groups/1487/courses/161/lecture/4435?page=3&ysclid=Irxou544q374570755 (дата обращения: 23.01.2024 г.).

15. Обоснование оптимального варианта принятия решений при управлении силами и средствами, действующих в условиях неопределенности / А.Б. Власов, В.И. Мещеряков, А.Г. Перунов, С.В. Салов // Современные проблемы науки и образования. – 2015. – № 2-1. – С. 764

16. Оценка эффективности и безопасности энерго-ресурсосбережения инвестиционных строительных проектов / В.В. Герасимов, М.П. Сабреков, Е.В. Улитко, Н.В. Светышев // ПОВЫШЕНИЕ качества и эффективности строительных и специальных материалов: Сборник Национальной научно-технической конференции с международным участием, Новосибирск, 18–22 февраля 2019 года. – Новосибирск: ИЦ НГАУ «Золотой колос», 2019. – С. 250-256

17. Проектирование и разработка сложноорганизованной социотехнической системы на основе нечетких интервальных оценок / К.А. Павлова, Н.В. Береза, А.Н. Береза, В.В. Бегляров // Вестник Балтийского федерального университета им. И. Канта. Серия: Физико-математические и технические науки. – 2018. – № 1. – С. 5-17

18. Порядина, В.Л. Оптимизационные модели поддержки принятия стратегических решений в управлении компаниями / В.Л. Порядина, Т.Г. Лихачева, А.В. Зобненко // Управление строительством. – 2018. – № 4(13). – С. 94-101

19. Саниева, А.Р. Особенности выхода микропредприятия на региональный рынок / А.Р. Саниева // Экономика и управление: теория, методология, практика: Сборник материалов XV Международной научно-практической конференции, Уфа, 10 декабря 2020 года. – Уфа: Башкирский государственный университет, 2020. – С. 187-192.

20. Стронгин, Р.Г. Исследование операций и модели экономического поведения: Курс лекций / Р.Г. Стронгин — Москва: Интуит НОУ, 2016. — 245 с.

21. Тапскотт Алекс, Дон Тапскотт. Технология блокчейн - то, что движет финансовой революцией сегодня. — М.: Эксмо, 2017. — 448 с

22. Gregory Piatetsky-Shapiro, Machine Learning and Data Mining.Course Notes, Fall 2003

23. Ray, S. (2019, February). A quick review of machine learning algorithms.In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39).

24. H. Avini1, Z. MirzaeiZavardJani, A. Avini (2022). Improved Recommender Systems Using Data Mining, Transactions on Machine Intelligence, 5(2) 97–114

25. <https://www.sap.com/cis/insights/what-is-blockchain.html>

26. <https://vc.ru/crypto/398036-tehnologiya-blockchain-tipy-sfery-primeneniya-preimushchestva-i-nedostatki>

27. <https://forklog.com/internet-veshhej-i-blokchejn-problemy-preimushchestva-i-sfery-primeneniya>

28. <https://blockchainwiki.ru/tehnologiya-blokchejn-i-internet-veshhi/>

29. https://www.tadviser.ru/index.php/Статья:Блокчейн_в_ИоТ

30. <https://itnan.ru/post.php?c=1&p=674806>

31. <https://merehead.com/ru/blog/how-to-implement-blockchain-in-io/>

32. <https://digiforest.io/blog/20-blockchain-project-examples-part-2>

33. <https://dop-mosreg.ru/mobilnye-operacionnye-sistemy-vidy-i-xarakteristiki>

34. <https://www.compgramotnost.ru/apple/plyusy-i-minusy-apple-ios>

35. <https://techtoday.in.ua/ru/reviews-ru/chto-takoe-kaios-y-mozhet-ly-ona-zamenyt-iphone-y-android-157883.html>
36. <https://japnoj.ru/maynkraft/telefon-na-kaios-2021-obzor-osobennosti-dostoinstva-i-nedostatki>
37. <https://setphone.ru/stati/lineageos-obzor-samoy-populyarnoy-polzovatel'skoy-proshivki-dlya-android/>
38. <https://pc-reanimator.ru/lineage-os-18-1-samsung-obzor-os-ustanovka/>
39. <https://linuxinsider.ru/obzor-lineageos>
40. https://ru.wikipedia.org/wiki/Fire_OS
41. <https://dalabu.ru/obzor-amazon-fire-hd-8-plus-poluchaete-to-za-chto-platite.html>
42. https://ru.wikipedia.org/wiki/Flyme_OS
43. <https://uchet-jkh.ru/i/cto-takoe-flyme-v-telefone-meizu>
44. <https://smages.com/stati/prjeimushhjestvainjedostatkiopjercionnojsistemywindowshone/>
45. <https://www.cy-pr.com>
46. <https://quest999.ru/fakty/symbian-cto-eto-takoe-i-kakim-obrazom-ono-vliyaet-na-mir-mobilnyx-ustroystv>
47. <https://ru.wikipedia.org/wiki/CyanogenMod>
48. <https://xakep.ru/2015/09/17/cyanogenmod/#toc02>
49. https://overclockers.ru/lab/show/75803_3/obzor-kastomnoj-proshivki-dlya-android-ustrojstv-cyanogenmod-13?ysclid=lpngqtwfi627118591
50. https://help.blackberryrussia.ru/page/view/osnovnye_ponyatiya/88-operacionnaya_sistema_blackberry
51. <https://www.iphones.ru/iNotes/user-posts/vse-pro-blackberry-os-10-06-29-2022>
52. <https://www.azone-it.ru/operacionnaya-sistema-avrora>
53. <https://setphone.ru/stati/sailfish-os-osobennosti-i-preimushhestva/>
54. <https://www.javatpoint.com/trends-in-data-mining>
55. https://www.researchgate.net/publication/337034679_Data_Mining_Technology_for_Structural_Control_Systems_Concept_Development_and_Comparison
56. <https://www.javatpoint.com/trends-in-data-mining>
57. https://www.researchgate.net/publication/337034679_Data_Mining_Technology_for_Structural_Control_Systems_Concept_Development_and_Comparison
58. <http://hdl.handle.net/10603/466133>

Учебное издание

**Марсель Равильевич Вагизов
Мария Александровна Новикова
Сергей Александрович Иванов
Наталья Владимировна Меламед**

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ
Часть 3**

*Учебное пособие для студентов
высших учебных заведений*

Подписано в печать 27.12.2023. Формат 60 × 90 ¹/₁₆.
Усл. печ. л. 4,4. Тираж 50 экз. Печать цифровая.
Заказ № 027С.

Отпечатано в соответствии с предоставленным оригинал-макетом
в типографии издательско-полиграфической фирмы «Реноме»,
192007, Санкт-Петербург, наб. Обводного канала, 40.
Тел. (812) 766-05-66. E-mail: book@renomespb.ru
ВКонтакте: https://vk.com/renome_spb
www.renomespb.ru