

Искусственный интеллект и большие данные в управленческой и инженерной деятельности

Судаков Владимир Анатольевич

sudakov@ws-dss.com

2024

Содержание

- Искусственный интеллект
- Экспертные системы
- Машинное обучение
- Системы поддержки принятия решений
- Модели исследования операций
- OLAP
- Большие данные

Литература

- <https://github.com/sudakov/>
- Стюарт Рассел, Питер Норвиг. Искусственный интеллект: современный подход (AIMA-2)
- Ларичев О.И. Теория и методы принятия решений, а также Хроника событий в Волшебных странах
- Афонин К.А., Бомас В.В., Судаков В.А. Поддержка принятия многокритериальных решений по предпочтениям пользователя. СПР DSS/UTES
- Саттон Ричард С., Барто Эндрю Г. Обучение с подкреплением
- Хемди Таха. Исследование операций

Искусственный интеллект (ИИ).

Определение

- Новое захватывающее направление работ по созданию компьютеров, способных думать, ... машин, обладающих разумом, в полном и буквальном смысле этого слова
- [Автоматизация] действий, которые мы ассоциируем с человеческим мышлением, т.е. таких действий, как принятие решений, решение задач, обучение...
- Изучение умственных способностей с помощью вычислительных моделей
- Наука о том, как научить компьютеры делать то, в чем люди в настоящее время их превосходят

Зачем нужен ИИ?

- 1.
- 2.
- 3.

Может быть ИИ не нужен?

А что мешает созданию ИИ?

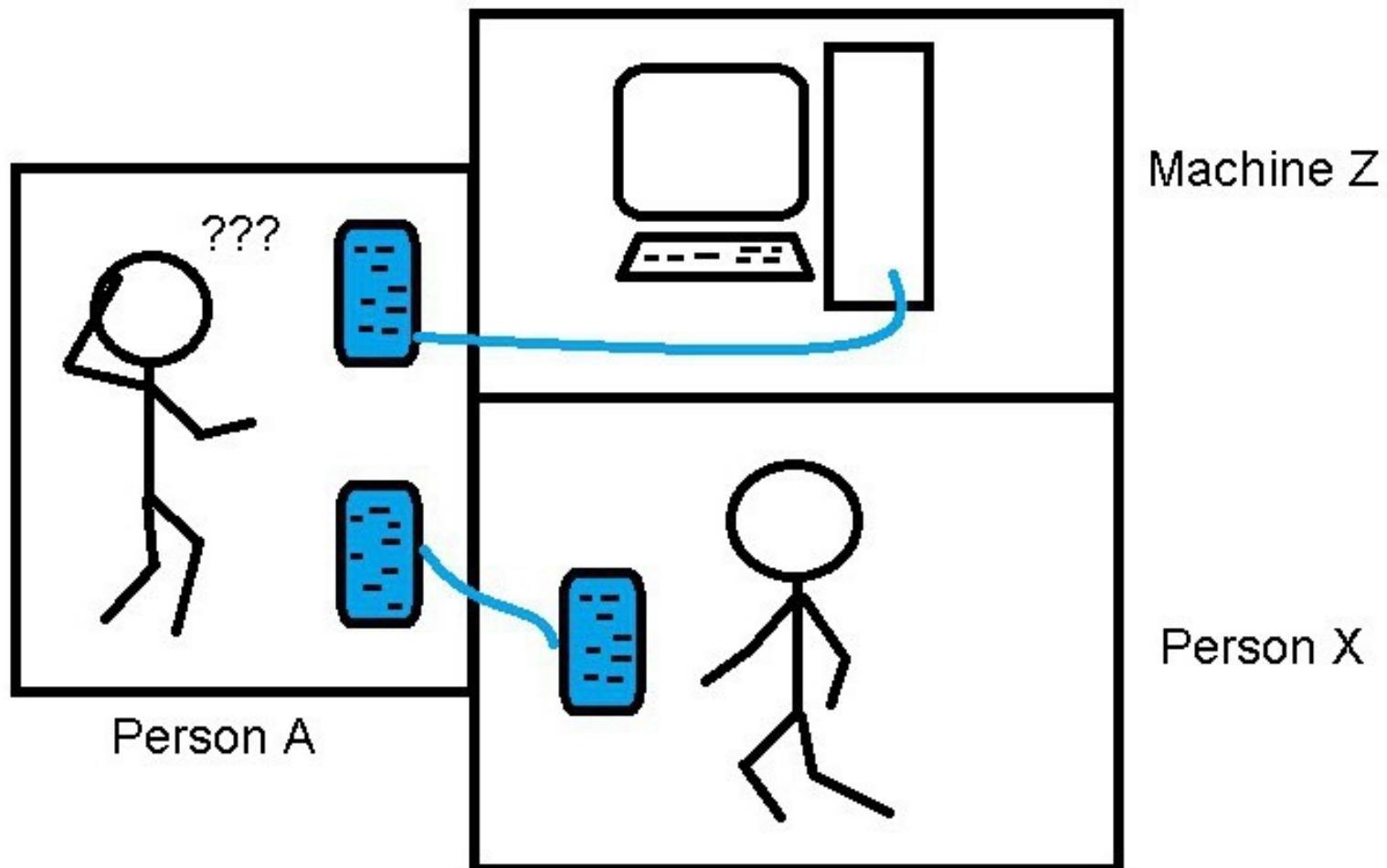
Искусственный интеллект

ГОСТ Р 59277—2020:

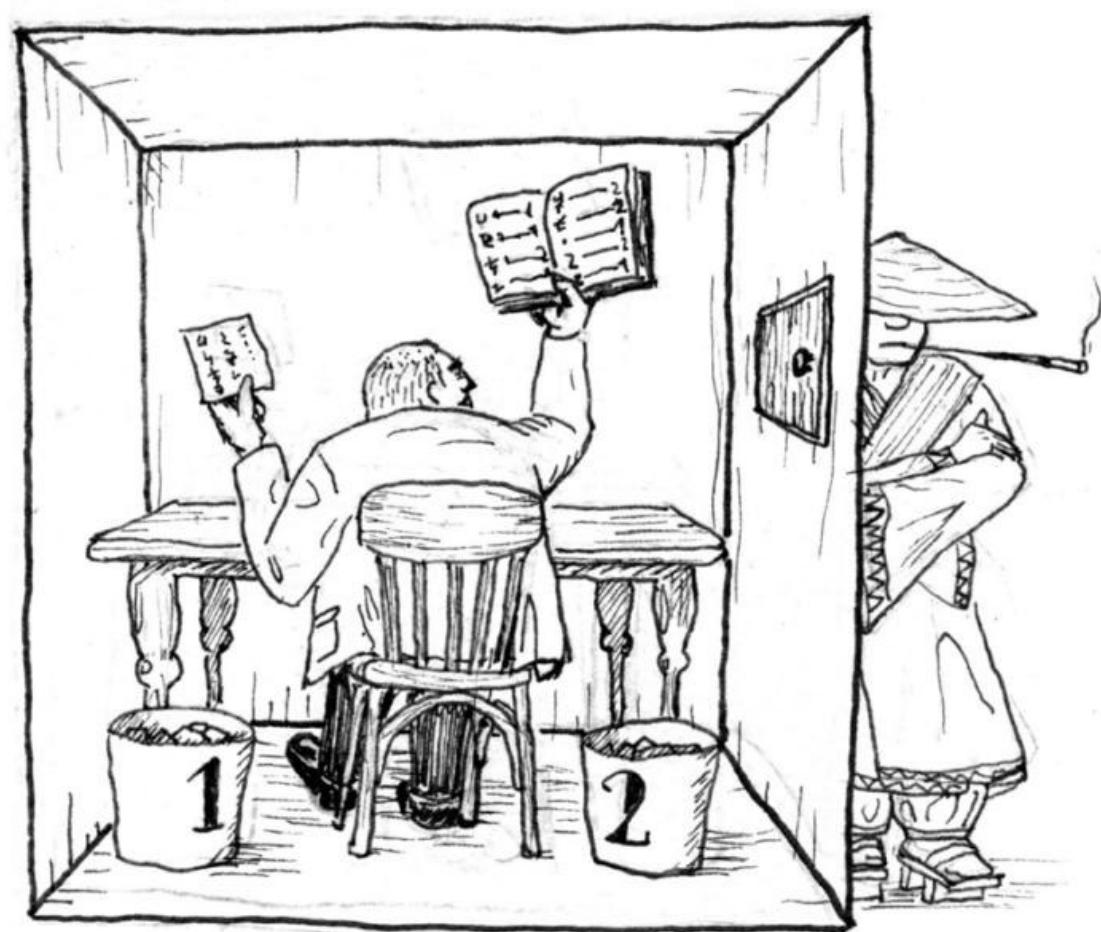
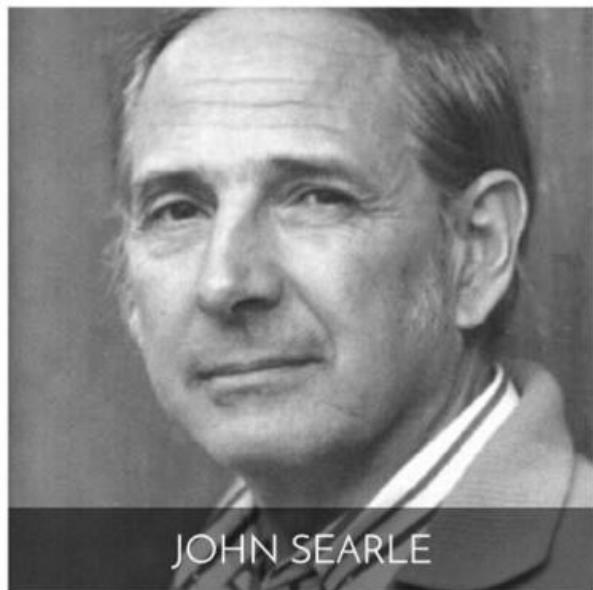
Искусственный интеллект (*artificial intelligence*): Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение, поиск решений без заранее заданного алгоритма и достижение инсайта) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

Примечание — Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных, анализу и синтезу решений.

Тест Тьюринга



Китайская комната



Некоторые популярные на сегодня задачи ИИ

- Обработка естественного языка (NLP), в том числе аннотирование научной-технической информации
- Анализ изображений



Машинное обучение.

Определение

- Машинное обучение (англ. machine learning, ML) — это исследование компьютерных алгоритмов, которые автоматически улучшаются благодаря опыту и использованию данных.
- Алгоритмы машинного обучения создают модель на основе выборочных данных, известных как «обучающие данные», чтобы делать прогнозы или предлагать решения, не будучи явно запрограммированными на это.

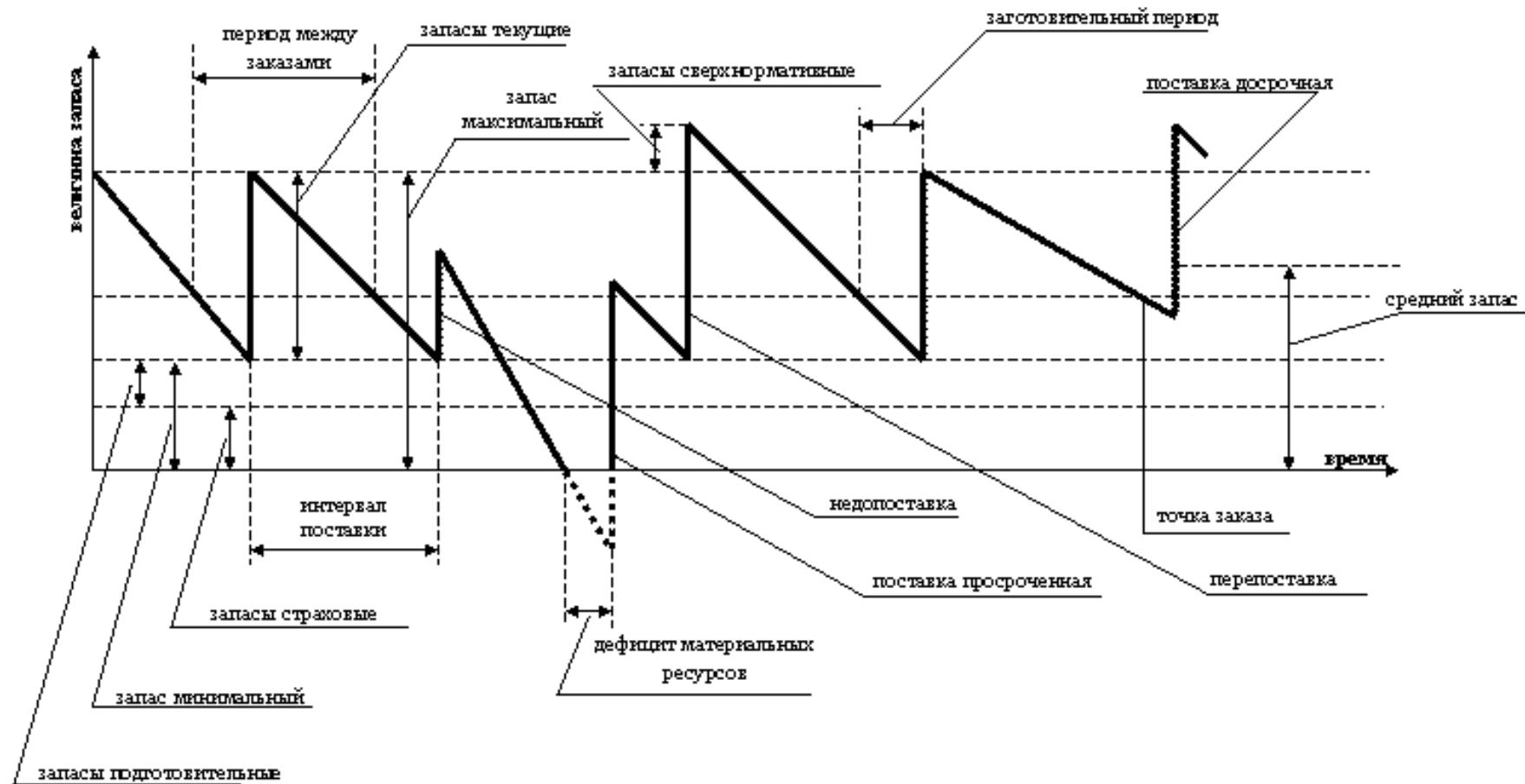
Про цели

- Зачем мы учим машины? Например, прогноз продаж
- Какие у нас цели?
- А какие цели у заказчика?

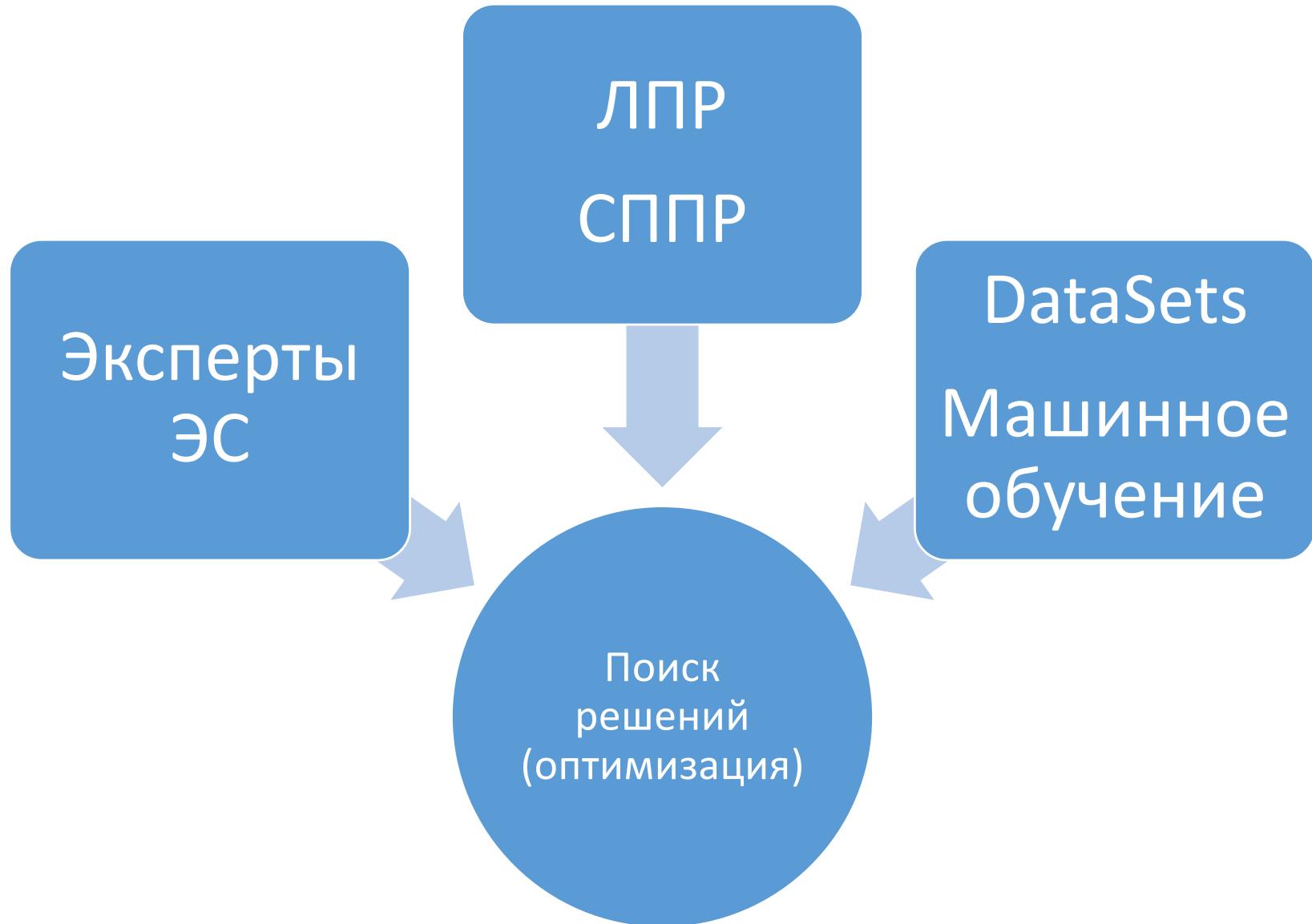
Цели

- Прибыль (важность не велика в краткосрочной перспективе)
- Доля рынка (важность велика)
- Сделать людей счастливыми
- Прославится
- Обеспечить долгую, стабильную жизнь компании
- Не знают....

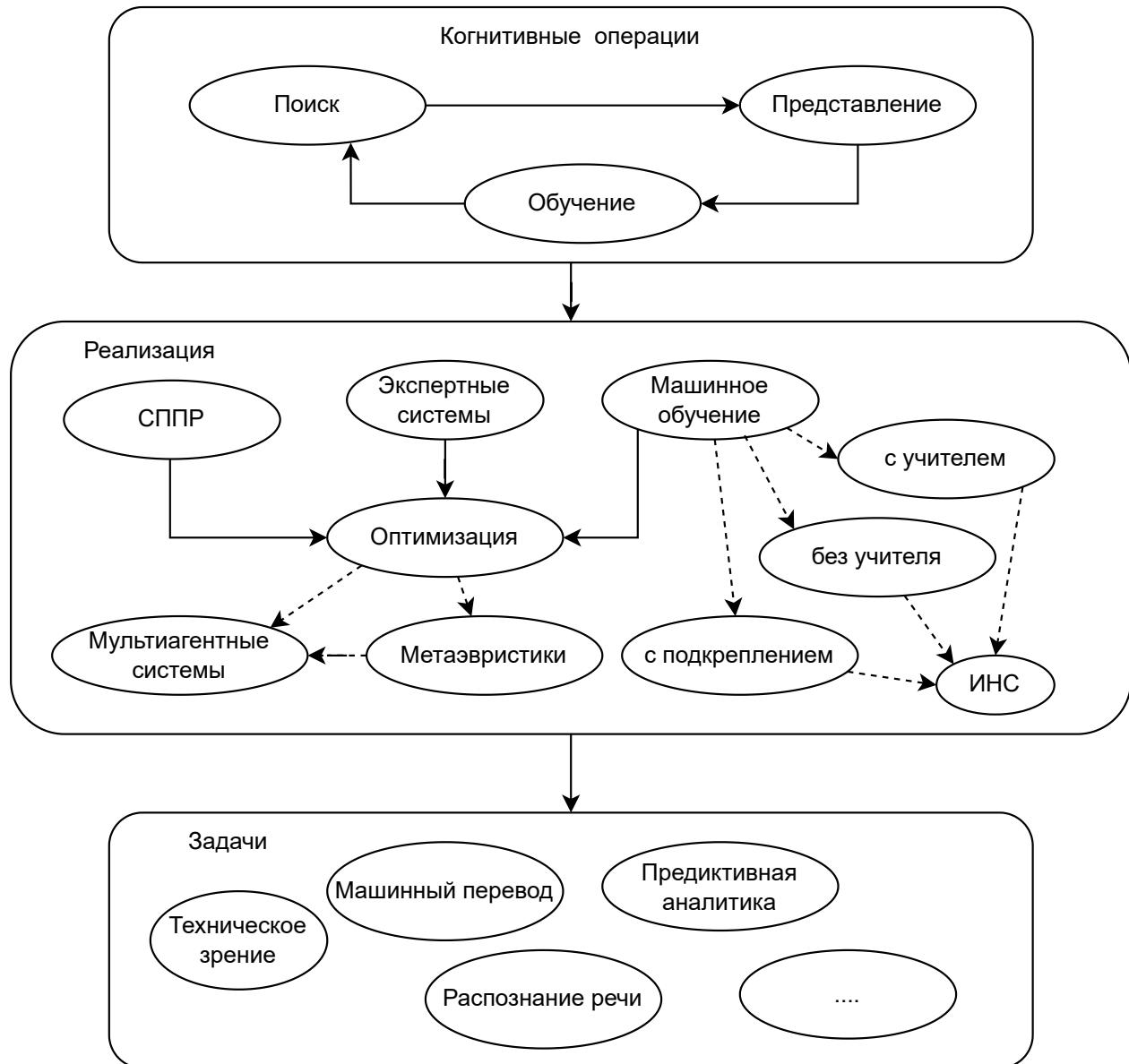
Реальные цели прогноза



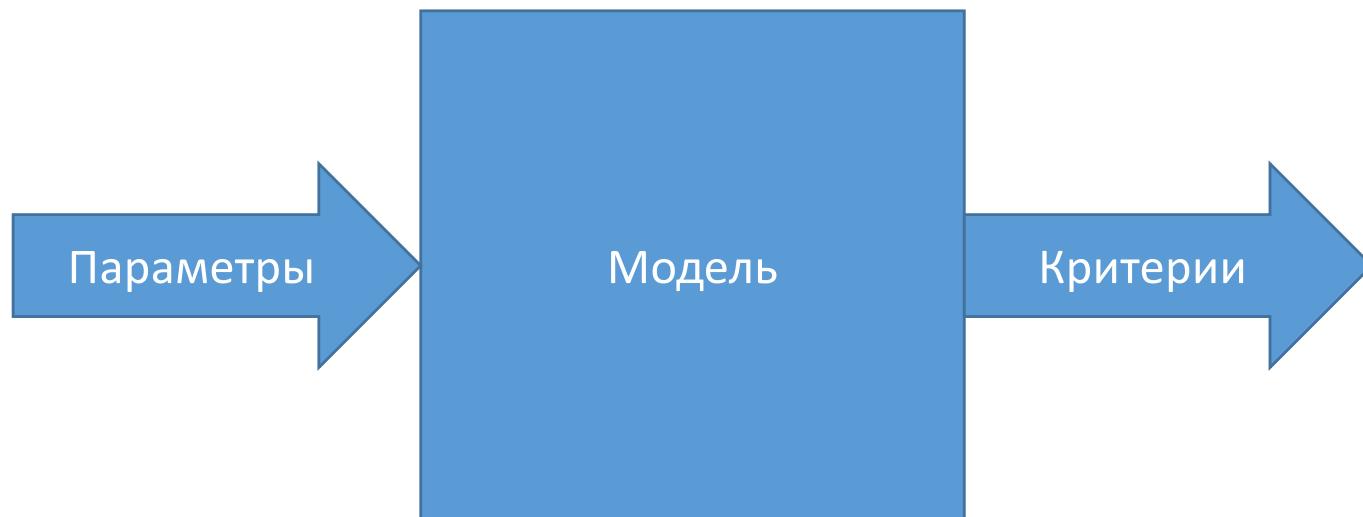
Подходы ИИ



Направления ИИ и области применения



Модель



Дискуссия

- Но можно ли описать автомобиль одной моделью?
- А есть ли польза от одной модели?
- Что такое система?

Анализ систем

Система –

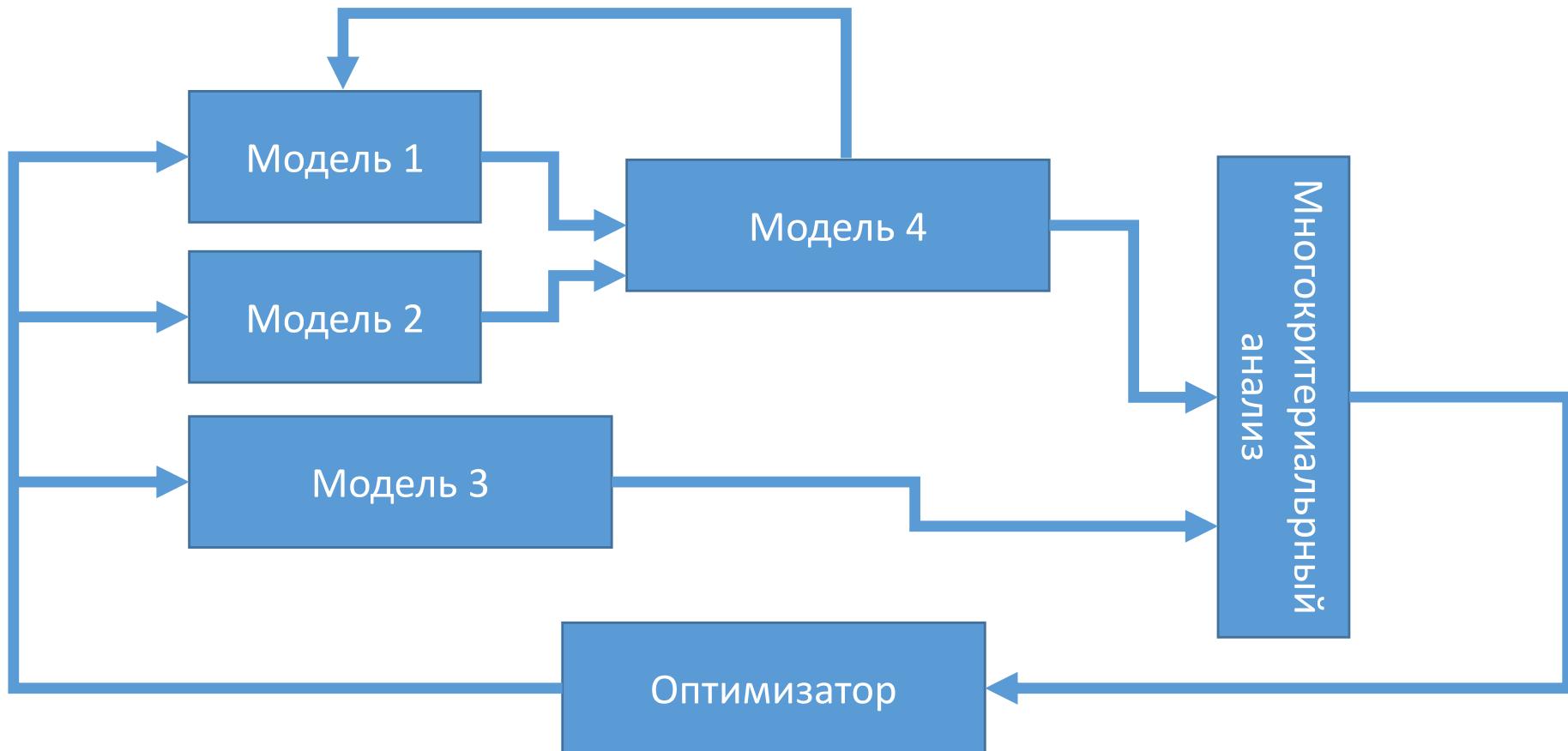
- множество элементов и
 - отношений между ними,
 - объединенных для достижения цели
-
- В качестве элементов могут быть другие системы
 - Система взаимодействует со средой как единое целое

Моделей много, их нужно
объединять в систему

Свойства:

- Интегративность
- Синергичность
- Эмерджентность
- Ингерентность

Использование моделей



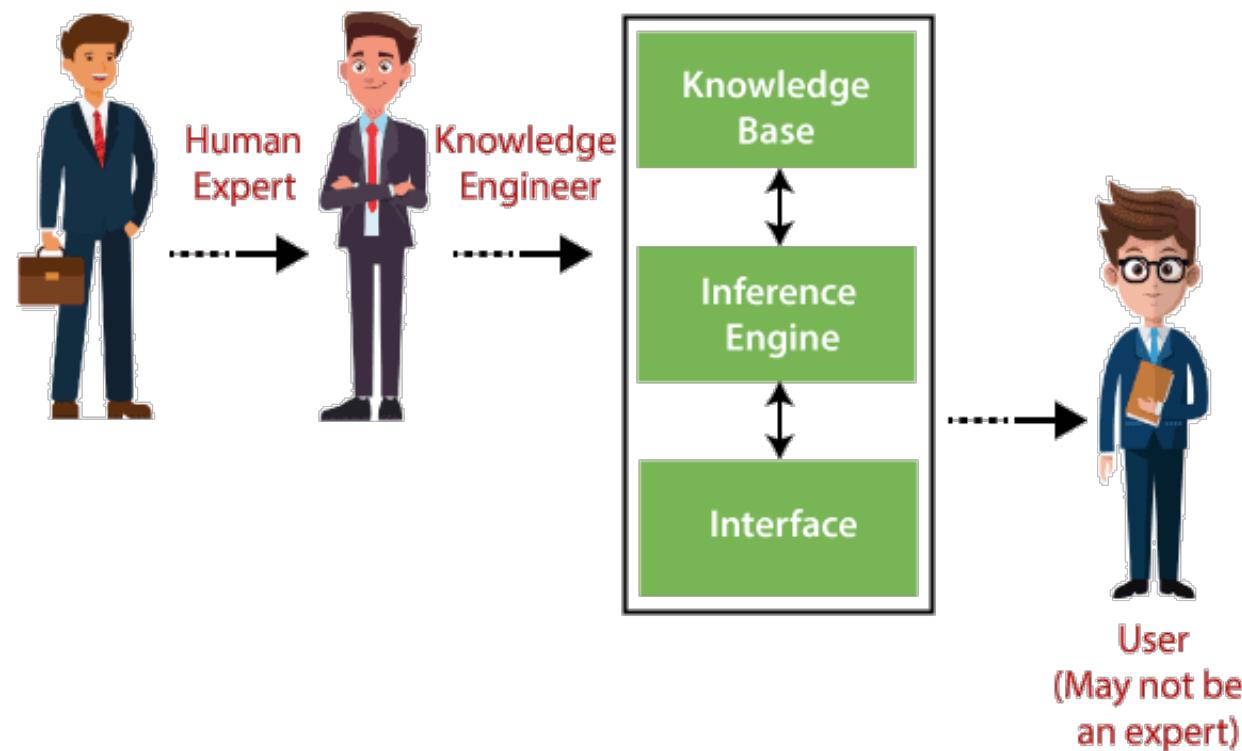
Экспертные системы

Экспертная система — компьютерная система, способная частично заменить специалиста-эксперта.

База знаний

- База знаний — база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области
- Важнейшей частью экспертной системы являются базы знаний как модели поведения экспертов в определённой области знаний с использованием процедур логического вывода и принятия решений, иными словами, базы знаний — совокупность фактов и правил логического вывода в выбранной предметной области деятельности.

Схема взаимодействия с экспертной системой



Элементы ЭС

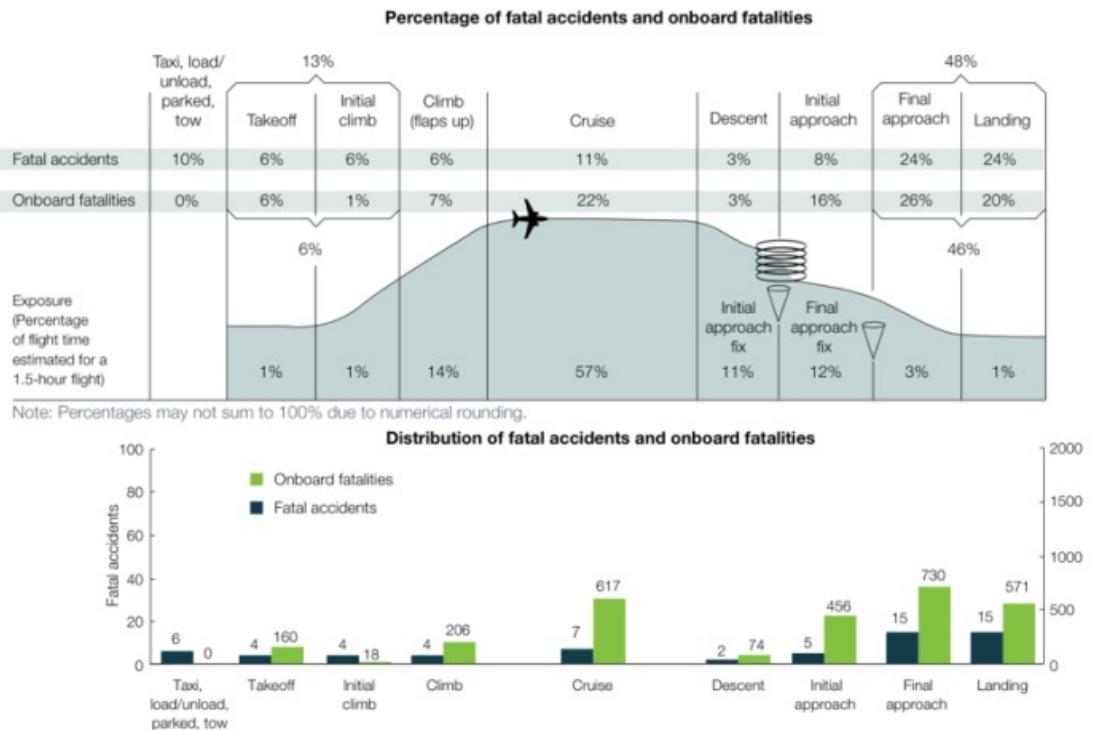
- Пользователь
- Интеллектуальный редактор базы знаний
- Эксперт
- Инженер по знаниям
- Рабочая (оперативная) память
- База знаний
- Решатель (механизм логического вывода (МЛВ))
- Подсистема объяснений

Этапы разработки ЭС

- Этап идентификации проблем — определяются задачи, которые подлежат решению, выявляются цели разработки, определяются эксперты и типы пользователей.
- Этап извлечения знаний — проводится содержательный анализ проблемной области, выявляются используемые понятия и их взаимосвязи, определяются методы решения задач.
- Этап структурирования знаний — выбираются ИС и определяются способы представления всех видов знаний, формализуются основные понятия, определяются способы интерпретации знаний, моделируется работа системы, оценивается адекватность целям системы зафиксированных понятий, методов решений, средств представления и манипулирования знаниями.
- Этап формализации — осуществляется наполнение экспертом базы знаний. В связи с тем, что основой ЭС являются знания, данный этап является наиболее важным и наиболее трудоёмким этапом разработки ЭС. Процесс приобретения знаний разделяют на извлечение знаний из эксперта, организацию знаний, обеспечивающую эффективную работу системы, и представление знаний в виде, понятном ЭС. Процесс приобретения знаний осуществляется инженером по знаниям на основе анализа деятельности эксперта по решению реальных задач.
- Реализация ЭС — создаётся один или несколько прототипов ЭС, решающие требуемые задачи.
- Этап тестирования — производится оценка выбранного способа представления знаний в ЭС в целом.

Fatal Accidents and Onboard Fatalities by Phase of Flight

Fatal Accidents | Worldwide Commercial Jet Fleet | 2007 through 2016



37% произошедших
с 2007 по 2016
авиационных инцидентов
и катастроф
приходится на этапы
взлета и посадки *

* Согласно Boeing "Statistical Summary of Commercial Jet Airplane Accidents -2016 | Worldwide Operations"

Введение

Расчет ВПХ автоматизирован для иностранных ВС

Для расчетов ВПХ на ВС иностранного производства (Boeing, Airbus) широкое распространение получили специализированные программные приложения, работающие на базе электронного планшета летчика

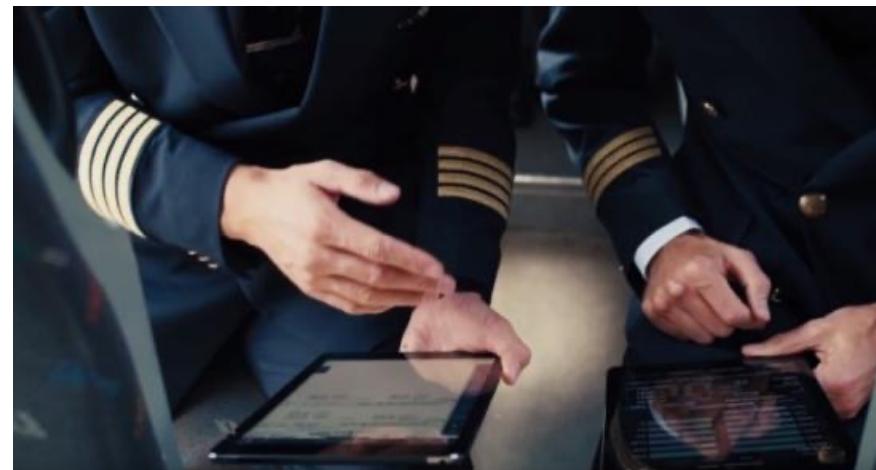


Аппаратная платформа

Классификация EFB



Встроенные



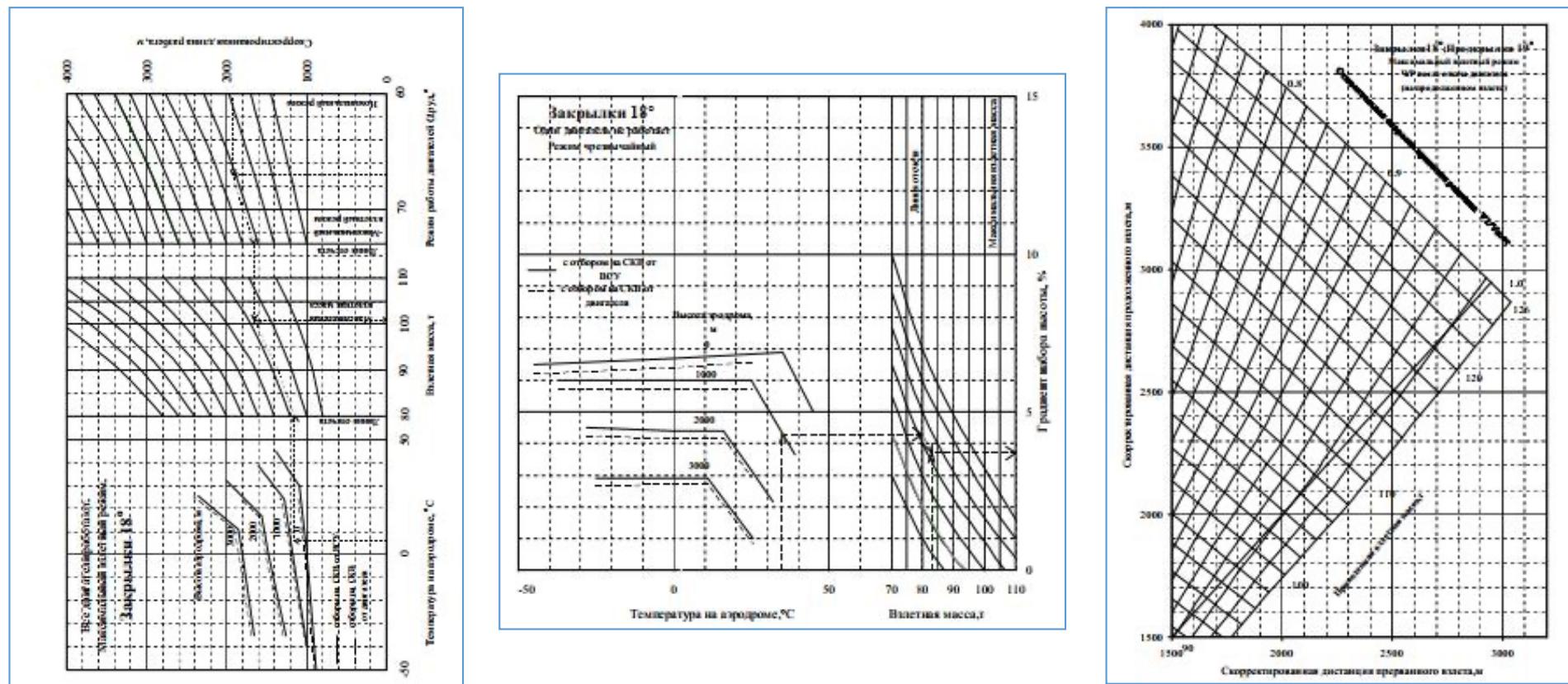
Переносные

Наиболее перспективными являются EFB переносного типа, аппаратно представляющие собой планшетный компьютер.

Расчет ВПХ для отечественных ВС

Пилоты отечественных ВС перед каждым полетом производят определение ВПХ вручную в соответствии с номограммами зависимостей ВПХ, приведенных в руководстве по лётной эксплуатации (РЛЭ) воздушного судна.

Штатное бортовое оборудование не позволяет выполнять расчет ВПХ.



Реализация

Интерфейс ПО

Выберите ВС Выберите Аэропорт и ВПП Обновить БД аэропортов

Выберите ВС Данные по аэродрому

| | | | | |
|-------------|-------|---|---------|---|
| Ту-204-100B | TODA | 0 | ASDA | 0 |
| МС-21 | Время | 0 | Heading | 0 |

Sukhoi Superjet 100 Фактические условия Загрузить фактическую погоду

| | | | | | |
|------------|-------------------------|--------------------|-----|----------|-------|
| Ил-96-300 | Снег 3-12мм Мокрый снег | Снег 10-50мм сухой | | | |
| Як-42Д | OAT | 20 | QNH | 1016 | |
| Скор.ветра | 3 | Направление ветра | 280 | Масса ВС | 99116 |

Расчет взлета Расчет посадки

Ту-204-100B UUWW01 Обновить БД аэропортов

Данные по аэродрому

| | | | | | |
|-----------|--------|-------|--------|---------|--------|
| TORA | 3060.0 | TODA | 3210.0 | ASDA | 3060.0 |
| Elevation | 192.9 | Slope | 0.01 | Heading | 13.0 |

Фактические эксплуатационные условия Загрузить фактическую погоду

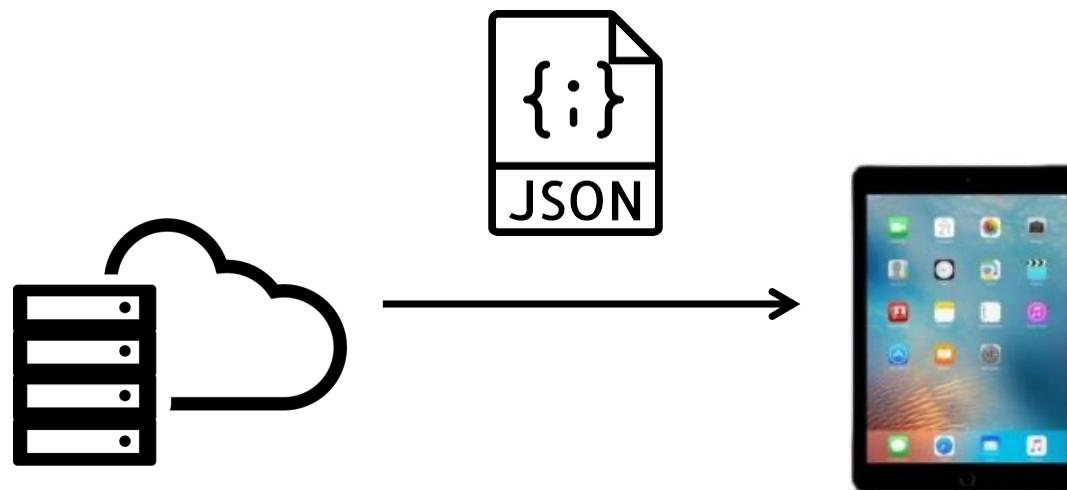
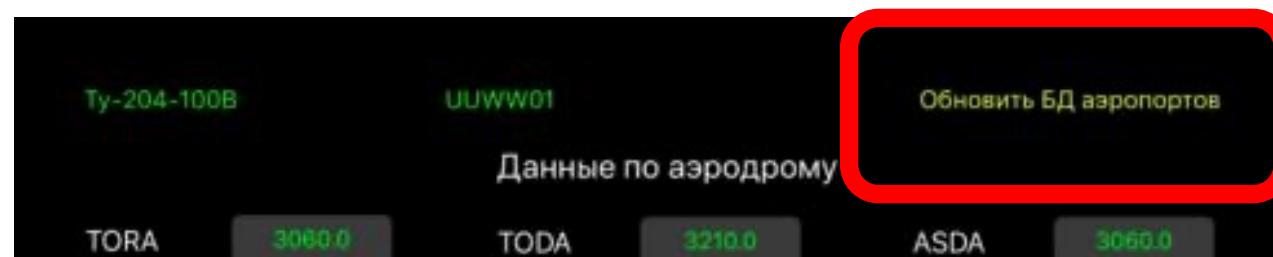
| | | | | | |
|--------------------------|-------------------------|--------------------|-----|----------|-------|
| Слой осадков отсутствует | Снег 3-12мм Мокрый снег | Снег 10-50мм сухой | | | |
| Коэф.сцеп. | 0.6 | OAT | 20 | QNH | 1016 |
| Скор.ветра | 3 | Направление ветра | 280 | Масса ВС | 99116 |

Рассчитать

Расчет взлета Расчет посадки

Реализация

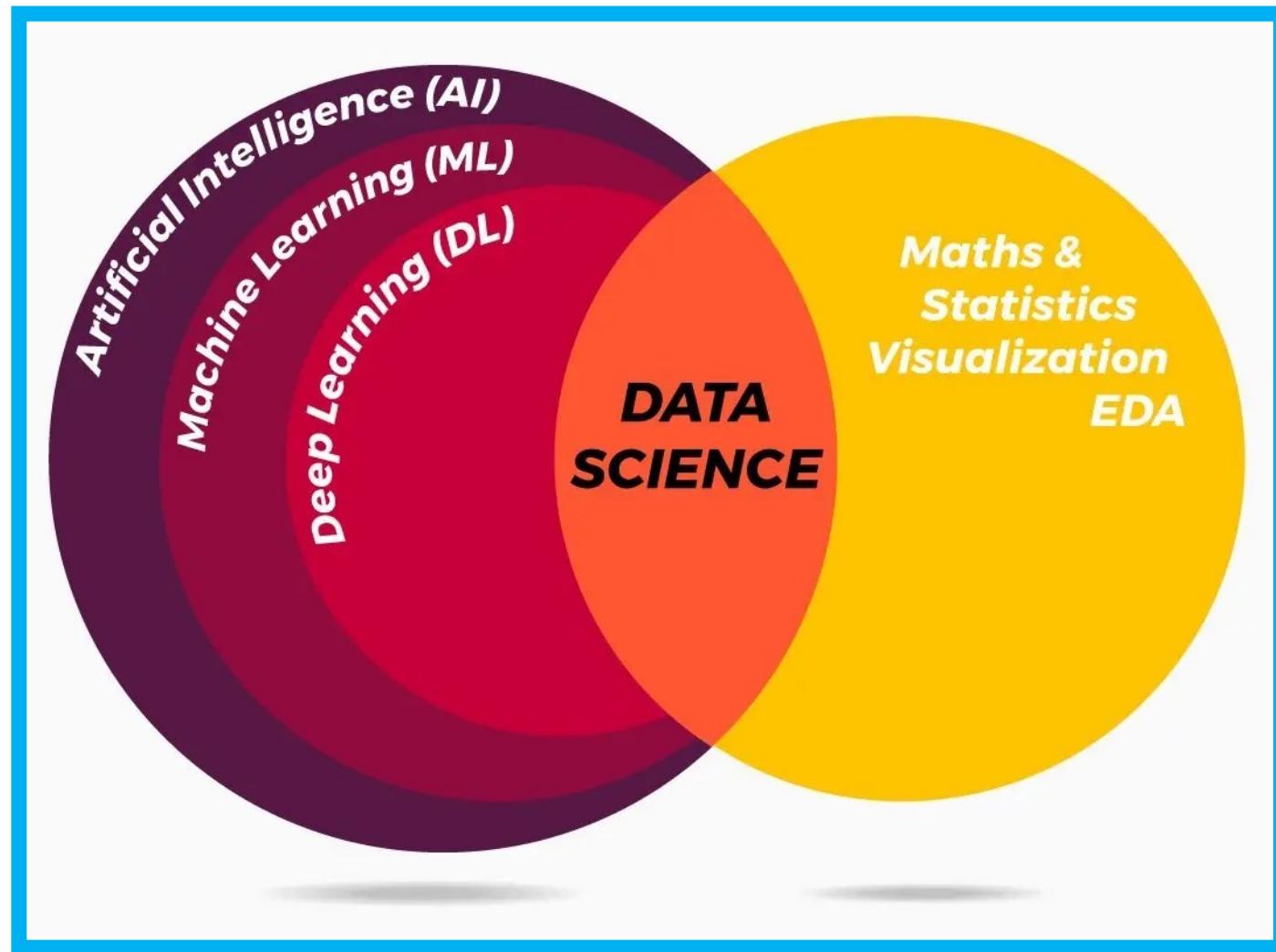
Характеристики аэродромов



Машинное обучение



Взаимосвязи между науками



Вопрос для обсуждения

Многие методы Data Science и Machine Learning появились достаточно давно - 50-70 года прошлого века, но активно использоваться в бизнесе начали только сейчас

С чем это связано?

Что такого случилось?

Что есть сейчас и чего не было тогда?

Задачи машинного обучения

- Регрессия
- Классификация
- Ранжирование
- Кластеризация
- Понижение размерности

Задача для самостоятельной работы

- Найдите в близкой Вам среде задачу, где модель была бы полезна
- Чем определяется эффективность бизнеса в этой среде? Какие критерии и цели?
- Как измерить полезность для бизнеса модели?
- Постараетесь сформулировать числовые показатели

Методы машинного обучения с учителем

Дано

Множество объектов X .

Множество допустимых ответов Y .

Целевая функция (target function) $y^* : X \rightarrow Y$, значения которой $y_i = y^*(x_i)$ известны только на конечном подмножестве объектов $\{x_1, \dots, x_l\} \subset X$.

Пары «объект– ответ» (x_i, y_i) называются прецедентами.

Совокупность пар $X^l = (x_i, y_i)_{i=1}^l$ называется обучающей выборкой (training sample).

Требуется найти

зависимость y^* по выборке X' , то есть построить решающую функцию (decision function)

$$a: X \rightarrow Y,$$

которая приближала бы целевую функцию $y^*(x)$, причём не только на объектах обучающей выборки, но и на всём множестве X .

Решающая функция a должна допускать эффективную компьютерную реализацию.

Типы задач

Задача регрессии – прогноз на основе выборки объектов с различными признаками. На выходе - вещественное число (2, 35, 76.454 и др.). Например, цена квартиры, стоимость ценной бумаги через неделю, ожидаемый доход магазина на следующий месяц, качество вина. Множество Y – бесконечное.

Задача классификации – получение категориального ответа на основе набора признаков. Имеет конечное количество ответов (часто, в формате «да» или «нет»): является ли изображение человеческим лицом, давать ли клиенту кредит, к какой категории отнести товар. Множество Y – конечное.

Пример задачи

Вход X :

4, 8, 9, 26

Выход Y :

39.2, 46.4, 48.2, 78.8

Задача:

Если $x = 256$,
то чему равен y ?

Подготовка к решению

jupyter ml1 Last Checkpoint: 2 минуты назад (autosaved)

Файл Редактировать Вид Вставка Ячейка Ядро Widgets Справка

File Cell Kernel Help

In [1]:

```
1 from sklearn import linear_model
2 import numpy as np
3 import matplotlib.pyplot as plt
```

In [2]:

```
1 x = np.array([4, 8, 9, 26])
2 y = np.array([39.2, 46.4, 48.2, 78.8])
```

In [3]:

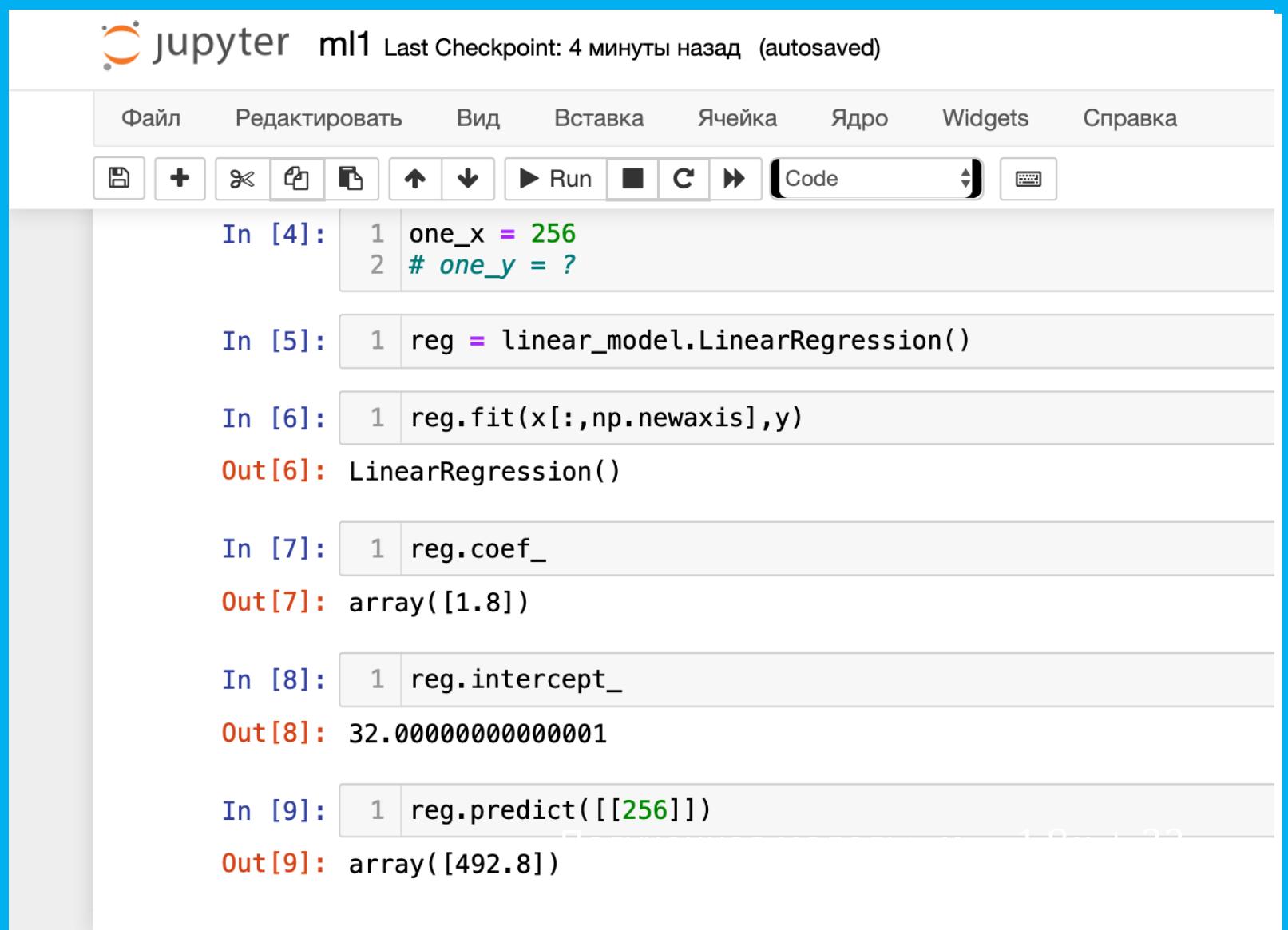
```
1 plt.plot(x, y)
```

Out[3]: [<matplotlib.lines.Line2D at 0x7fbe6b2144f0>]

In [4]:

```
1 one_x = 256
2 # one_y = ?
```

Решение в scikit-learn



The screenshot shows a Jupyter Notebook interface with a blue border. At the top, it displays the logo and the notebook name "jupyter ml1 Last Checkpoint: 4 минуты назад (autosaved)". Below the title bar is a menu bar with Russian labels: Файл, Редактировать, Вид, Вставка, Ячейка, Ядро, Widgets, Справка. Underneath the menu bar is a toolbar with various icons for file operations like save, new, cut, copy, paste, and run. The main area contains the following code snippets:

```
In [4]: 1 one_x = 256
         2 # one_y = ?

In [5]: 1 reg = linear_model.LinearRegression()

In [6]: 1 reg.fit(x[:,np.newaxis],y)

Out[6]: LinearRegression()

In [7]: 1 reg.coef_

Out[7]: array([1.8])

In [8]: 1 reg.intercept_

Out[8]: 32.00000000000001

In [9]: 1 reg.predict([[256]])

Out[9]: array([492.8])
```

Решение в многомерном случае

| x1 | x2 | x3 | x4 | y |
|----------------|----|--------------------|--------------|------------|
| 18 | 71 | 69 | 60 | 77 |
| 27 | 74 | 97 | 27 | 337 |
| 5 | 42 | 33 | 41 | -27 |
| 49 | 38 | 44 | 56 | 123 |
| 83 | 76 | 42 | 37 | 196 |
| 91 | 81 | 34 | 42 | 155 |
| 78 | 93 | 35 | 49 | 76 |
| 97 | 41 | 89 | 93 | |
| | | | | |
| | | | | |
| Коэффициенты | | Стандартная ошибка | t-статистика | P-Значение |
| Y-пересечение | | 2,2737E-13 | 1,4927E-13 | 1,52328045 |
| Переменная X 1 | | 3 | 1,45E-15 | 2,0689E+15 |
| Переменная X 2 | | -2 | 3,6581E-15 | -5,467E+14 |
| Переменная X 3 | | 5 | 1,6099E-15 | 3,1058E+15 |
| Переменная X 4 | | -3 | 3,2317E-15 | -9,283E+14 |

Математическая постановка

Дана обучающая выборка: $X^l = \{(x_i, y_i)\}_{i=1}^l$

x_i – вектор размерности n .

Необходимо построить такую функцию:

$$a(x, w) = \sum_{j=1}^n w_j f_j(x_j)$$

Для поиска w используем метод наименьших квадратов:

$$\min_w \sum_{i=1}^l (a(x_i, w) - y_i)^2$$

Вопросы для контроля

Мы начали с простейшей модели:

$$y = 1.8x + 32$$

А как в формуле общего вида

$$y = \sum_{j=1}^n w_j x_j$$

Отразить свободный член 32?

А можно ли используя линейную регрессию получить квадратичную модель вида:

$$y = ax^2 + bx + c ?$$

Как оценить качество решения?

Перед началом обучения выборку разбивают на обучающую и тестовую.

Качество следует оценить на тестовой выборке.

Меры качества:

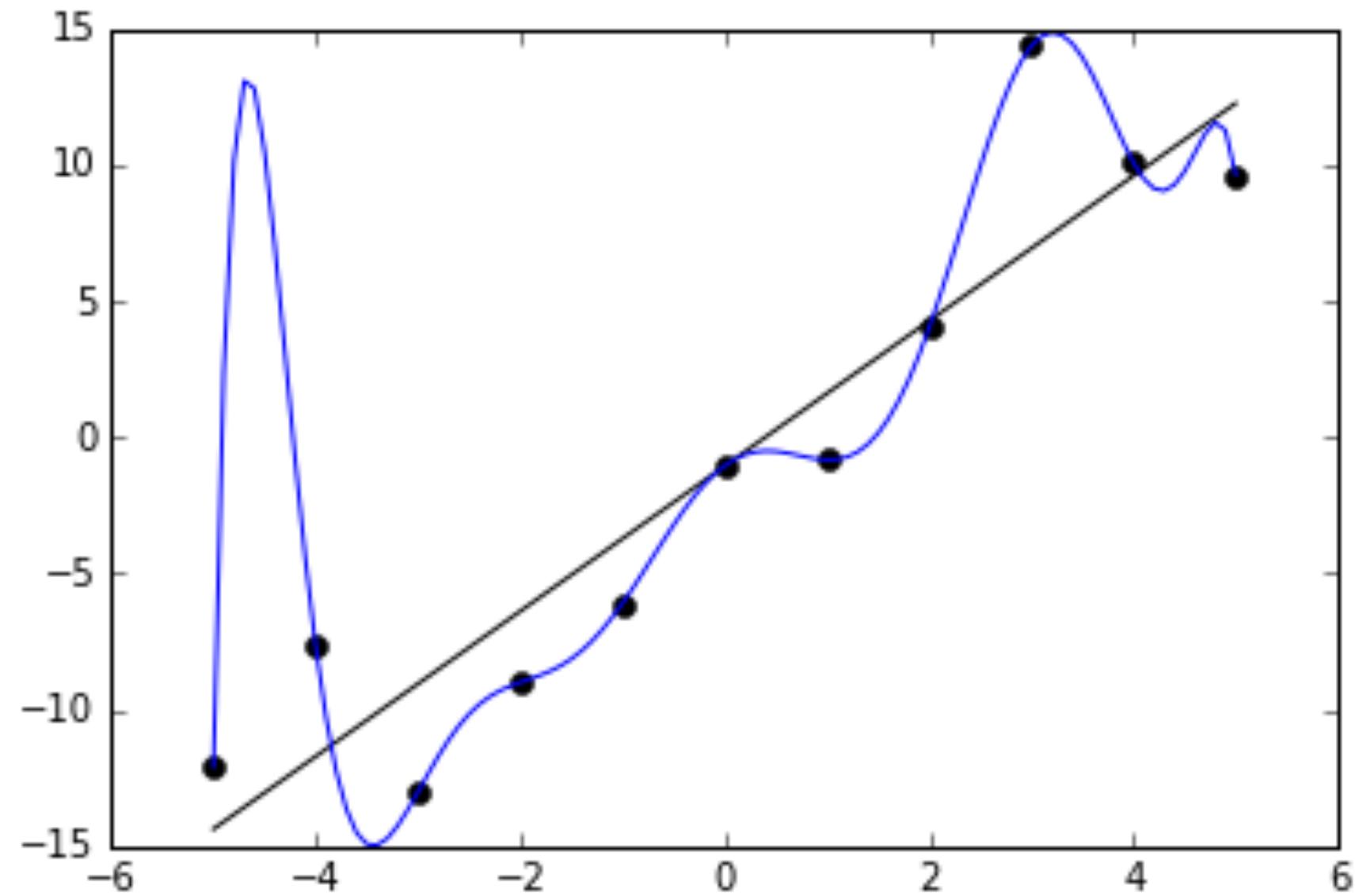
Средняя квадратичная ошибка (англ. Mean Squared Error, MSE).

Средняя абсолютная ошибка (англ. Mean Absolute Error, MAE).

Коэффициент детерминации (R^2). Коэффициент детерминации измеряет долю дисперсии, объясненную моделью, в общей дисперсии целевой переменной. Если она близка к единице, то модель хорошо объясняет данные, если же она близка к нулю, то прогнозы сопоставимы по качеству с константным предсказанием.

$$R^2 = 1 - \frac{\sum_{i=1}^n (a(x_i) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Переобучение



Вопросы для контроля

Чему равен y для последней строки в примере с многомерной регрессией?

| x1 | x2 | x3 | x4 | y |
|----|----|----|----|-----|
| 97 | 41 | 89 | 93 | ??? |

Постройте линейную регрессию для предсказания качества вина по набору данных:

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

Оцените качество решения

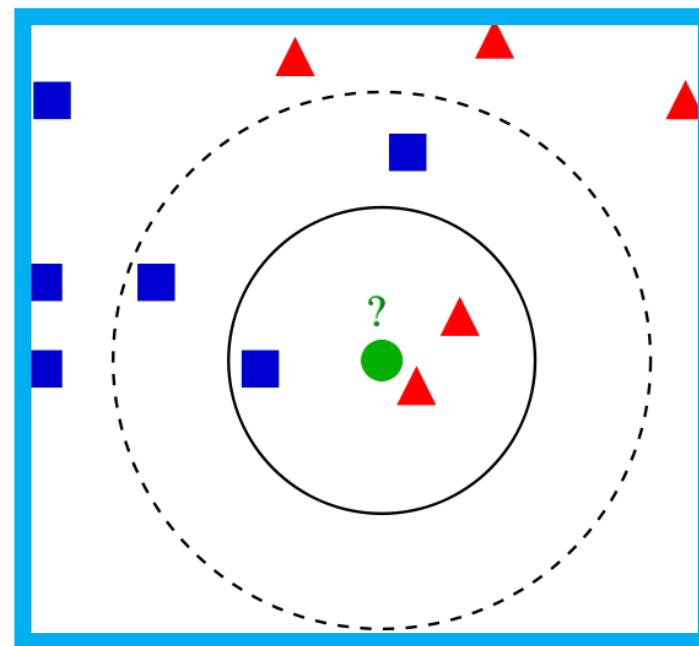
Метод ближайших соседей

Правило классификации: объект принадлежит тому же классу что и его k-ближайший соседей.

Близость определяется в пространстве признаков.

Для применения метода необходимо решить задачи:

- нормализация признаков,
- выбор метрики,
- выбор k.



Оценка качества классификации

Пусть есть два класса $Y=\{0,1\}$. Пусть банк использует систему классификации заёмщиков на кредитоспособных и некредитоспособных. Обнаружение некредитоспособного заёмщика ($y=1$) можно рассматривать как "сигнал тревоги", сообщающий о возможных рисках.

Возможны следующие исходы классификации:

- Некредитоспособный заёмщик классифицирован как некредитоспособный, т.е. положительный класс распознан как положительный (True Positive — TP).
- Кредитоспособный заёмщик классифицирован как кредитоспособный, т.е. отрицательный класс распознан как отрицательный. (True Negative — TN).
- Кредитоспособный заёмщик классифицирован как некредитоспособный, т.е. имела место ошибка, в результате которой отрицательный класс был распознан как положительный (False Positive — FP) – это ошибка I рода (ложная тревога).
- Некредитоспособный заёмщик распознан как кредитоспособный, т.е. имела место ошибка, в результате которой положительный класс был распознан как отрицательный (False Negative — FN) – это ошибка II рода (пропуск цели).

Вопрос для самоконтроля

Где ошибка первого рода и где ошибка второго рода?



Метрики качества классификации

Аккуратность (англ. Accuracy) – доля правильных ответов.
Бесполезна в задачах с неравными классами.

Точность (англ. Precision) - доля правильных ответов модели в пределах класса:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Полнота (англ. Recall) - это доля истинно положительных классификаций:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F-мера (англ. F-score) – гармоническое среднее между точностью и полнотой.

Задача

- Давайте познакомимся:
 - Что Вы пьете по утрам? Чай или Кофе? Научите ИИ прогнозировать утренний напиток методом k ближайших соседей.
- Разбиваемся на команды 3-4 человека:
 - Распределение ролей
 - Парное программирование
 - Подготовка исходных данных
 - Тестирование
 - Анализ – какое k лучше?
 - Показ решения
 - Code review чужой бригадой.
- Обсуждение
 - Какое решение лучше и почему?

Итоги

- Машинное обучение с учителем решает задачи регрессии и классификации
- Нужно сформировать обучающую и проверочную выборки
- Визуализируйте данные, выбирайте признаки
- Оценивайте метрики результата
- Совершенствуйте модели

Методы машинного обучения без учителя

Отличия

Обучение с учителем (supervised) vs

Обучение без учителя (unsupervised)



Задачи и приложения

- Задачи кластеризации
- Задачи обобщения
- Задачи обнаружения аномалий
- Задачи поиска правил ассоциации
- Задачи сокращения размерности
- Маркетинговые исследования: разбиение множества всех клиентов на кластеры для выявления типичных предпочтений.
- Анализ рыночных корзин: выявление сочетаний товаров, часто встречающихся вместе в покупках клиентов.

Метод k-средних

Задача: разбить выборку на кластеры

$$S = \{S_1, S_2, \dots, S_k\}$$

Метод стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров:

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2$$

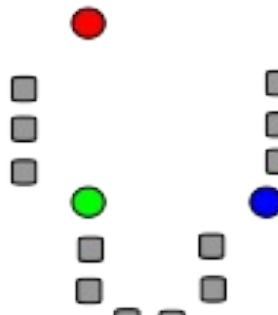
μ_i - центр кластера S_i .

Алгоритм k-средних

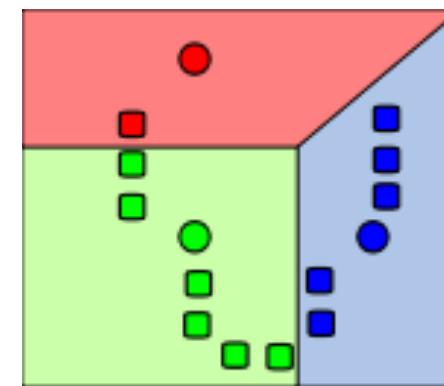
1. Начальные μ_i выбираются случайно.
2. Относим наблюдения к тем кластерам, чье μ_i к ним ближе всего.
3. Затем μ_i перевычисляется.
4. Если μ_i существенно изменились, то возврат к шагу 2.

Визуализация алгоритма k-средних

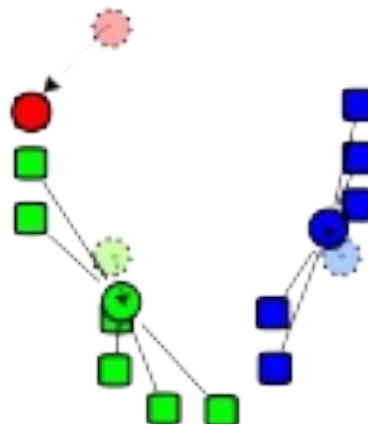
1



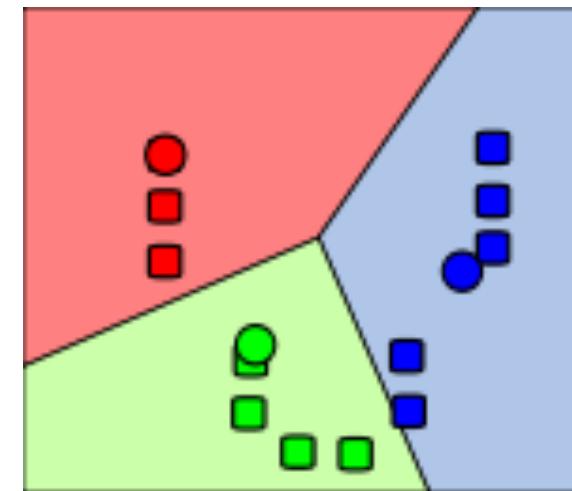
2



3



4



Пример задачи

Опасное вождение



Есть данные по трекам, скорости и ускорению

Нет (и не будет) данных о тому что такое
опасное вождение

А что делать?

Кластеризация вождения

Вход модели:

- скорости и ускорения конкретных ТС по конкретным водителям
- число кластеров
- уровень иерархии объектов для прогноза

Выход модели:

- принадлежность водителя определённому кластеру

По результатам работы модели нужно:

- выбрать несколько водителей из каждого кластера – отследить насколько хорошо они водят.
- присвоить кластерам категории опасности.
- на регулярной основе информировать о попадании водителя в «опасный» кластер.

Оценка качества кластеризации

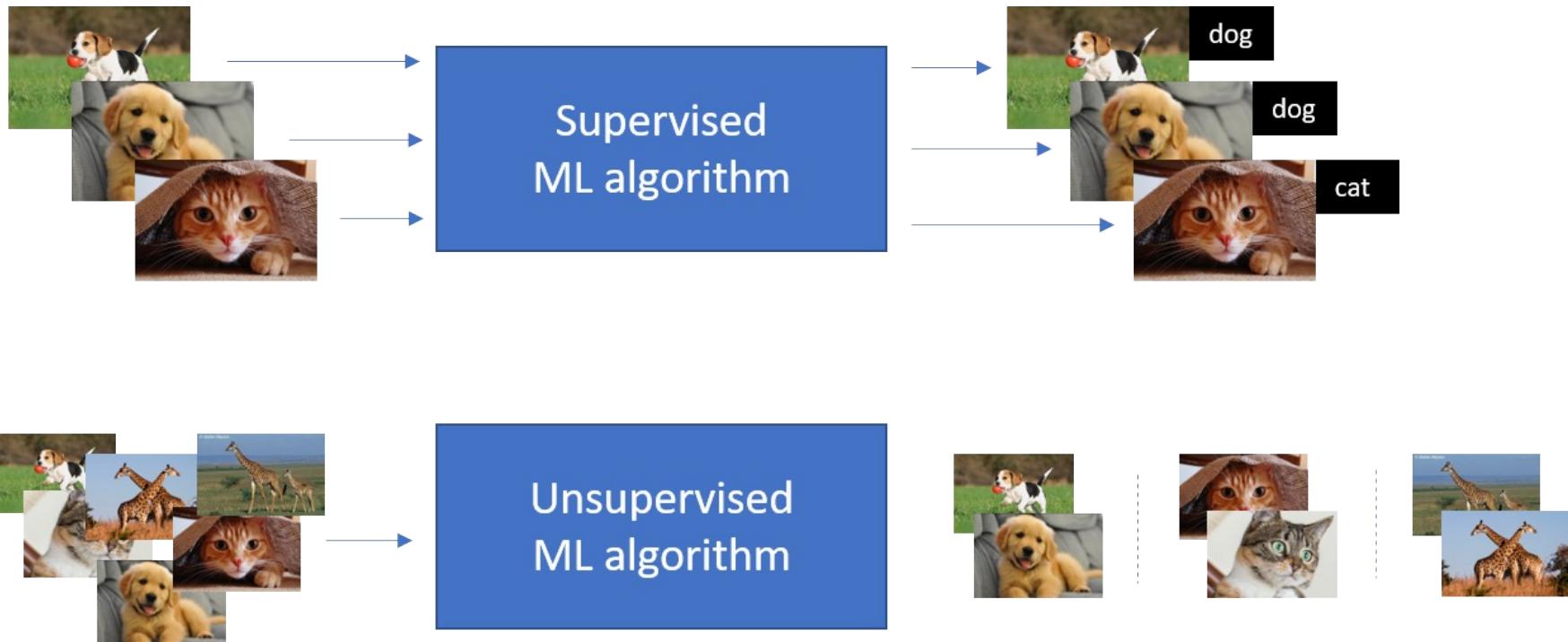
Внешние (англ. External) меры основаны на сравнении результата кластеризации с априори известным разделением на классы.

Внутренние (англ. Internal) меры отображают качество кластеризации только по информации в данных.

Из внутренних мер часто применяют:

Силуэт (англ. Silhouette) – он показывает, насколько объект похож на свой кластер по сравнению с другими кластерами.

Цели машинного обучения



А может быть нам не важно кто на фото?

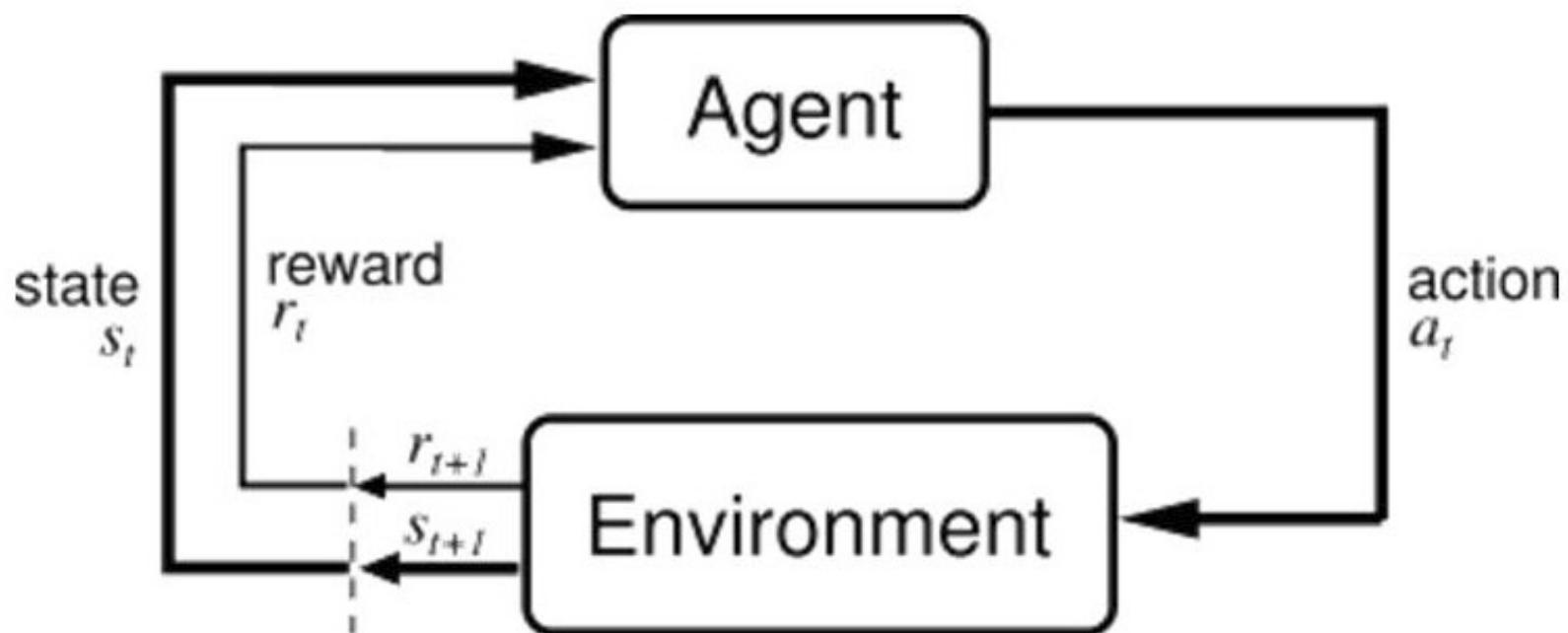
Нам важно достигать тех целей которые преследует бизнес

Обучение с подкреплением

- Обучение тому, что делать
- Как отобразить ситуации на действия, чтобы максимизировать вознаграждение
- Обучаемому не говорят, что делать
- Он должен сам понять какие действия приносят максимальное вознаграждение

Характеристики

- Поиск методом проб и ошибок
- Отложенное вознаграждение



Основные понятия

- Стратегия – отображение множества состояний на действия. Могут быть стохастическими
- Сигнал вознаграждения – вознаграждение на каждом временном шаге. Цель агента – максимизировать полное вознаграждение
- Функция ценности – полное вознаграждение в будущем, если агент начнет работу в данном состоянии. Показывает, что хорошо в длительной перспективе.
- Модель – прогноз поведения среды

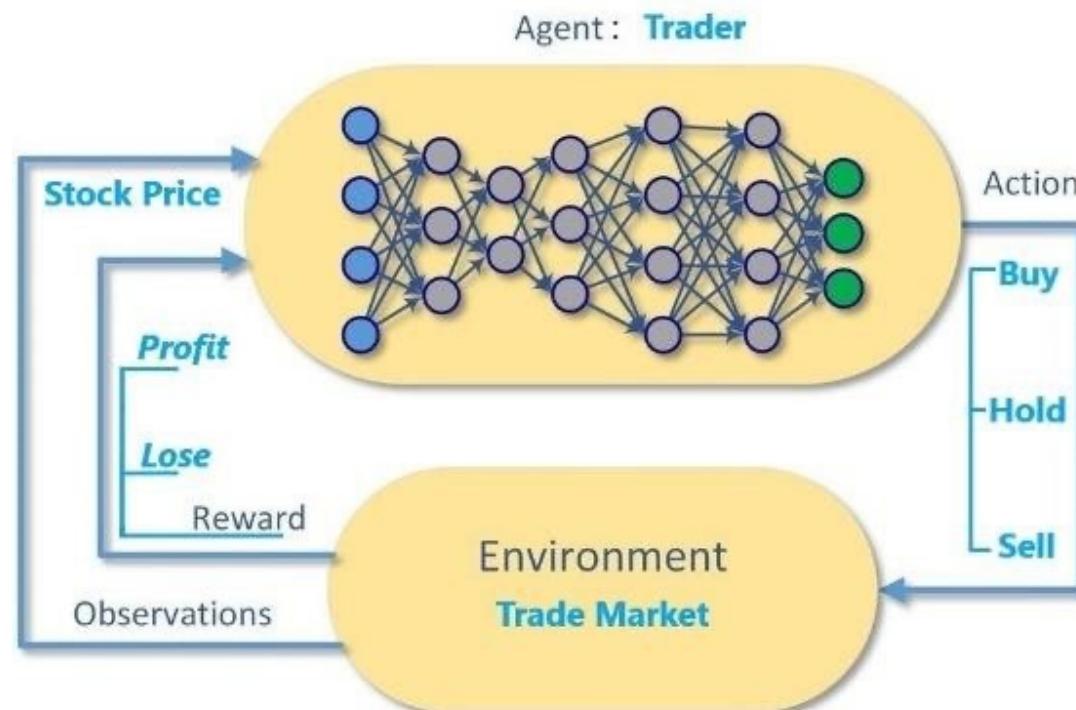
Обучение с подкреплением для управления вертолетами



Coates A., Abbeel P., Ng A.Y. (2017) Autonomous Helicopter Flight Using Reinforcement Learning. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_16

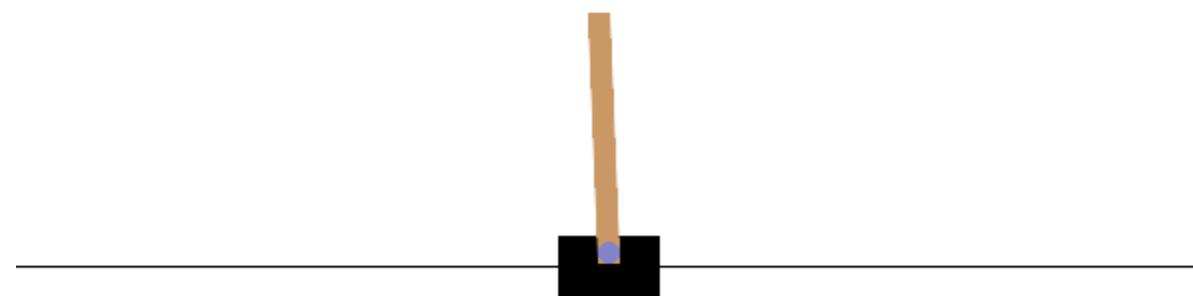
Это не только про роботов

- Задача коммивояжёра
- Портфель инвестиций



<https://gymnasium.farama.org>

CartPole



А как это происходит в реальности

Чем определяется состояние среды?

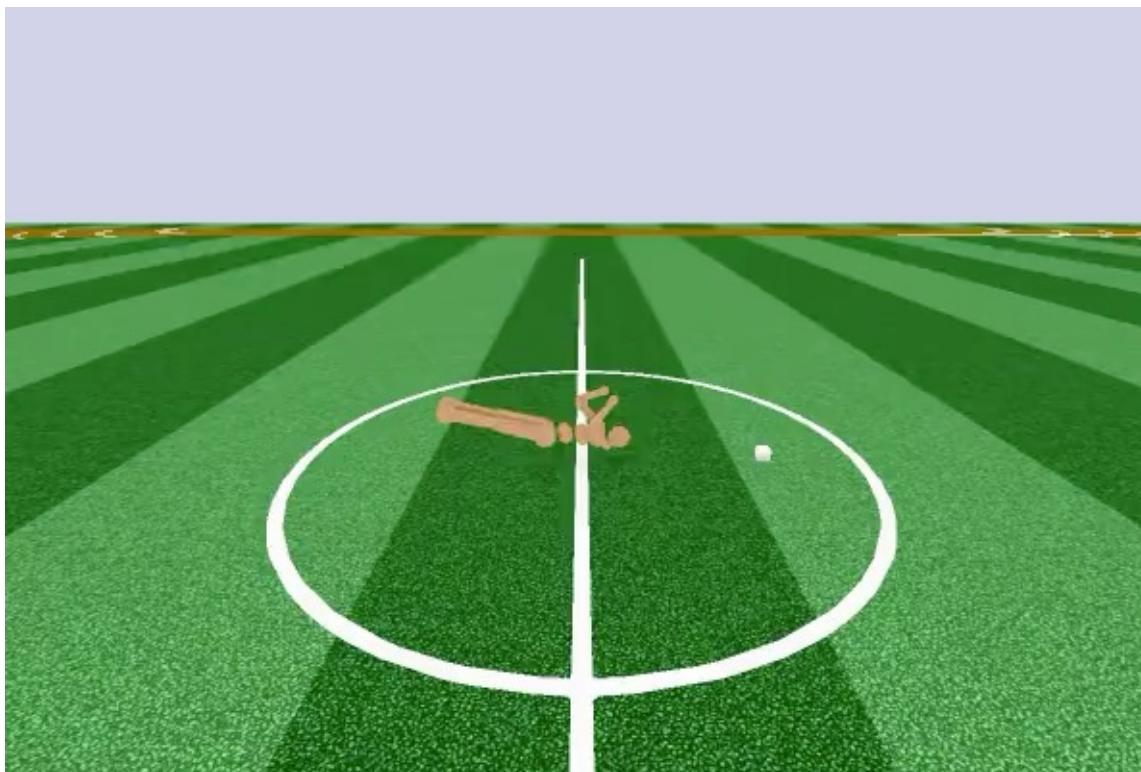
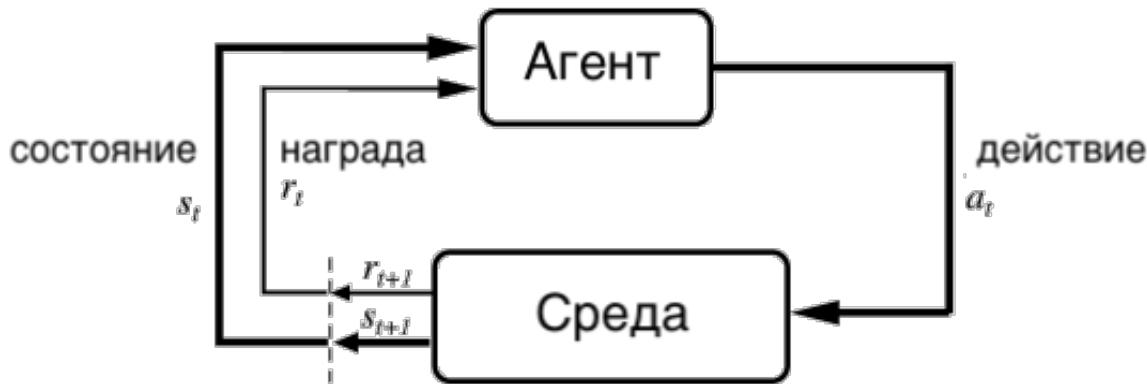
Каково множество действий агента?

Что в этой задаче сигнал вознаграждения?

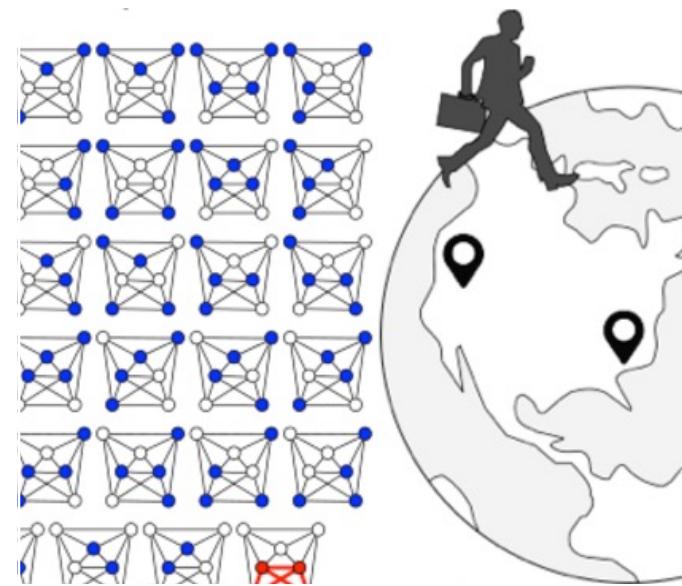
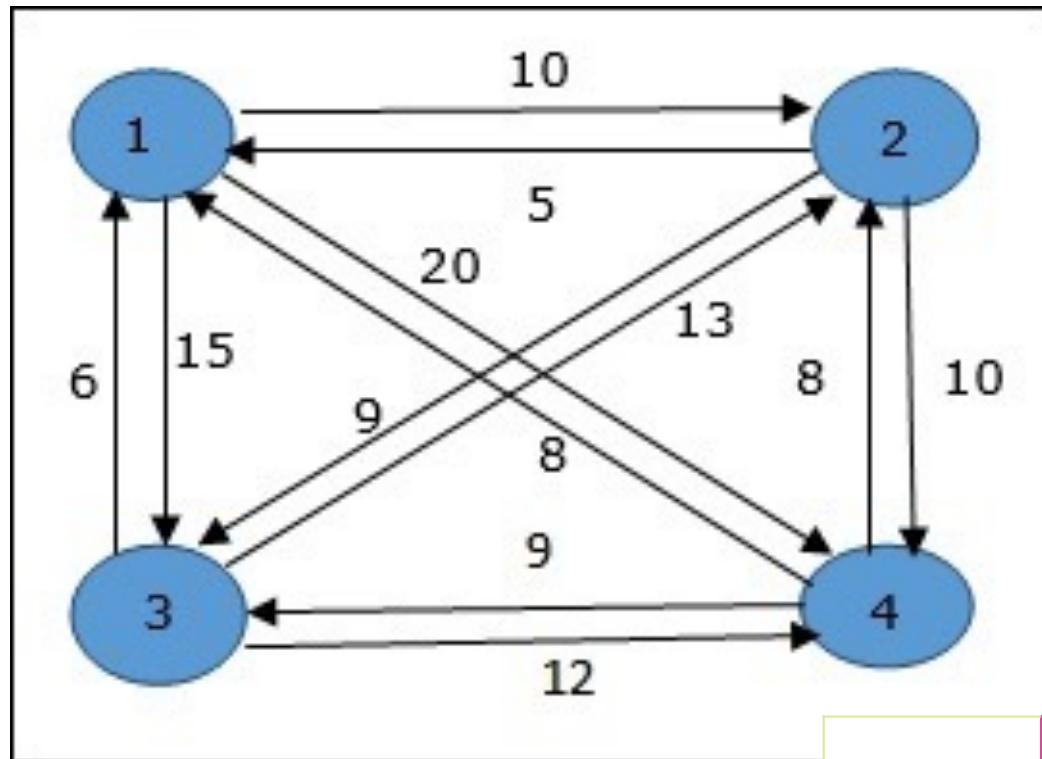
**Cart-Pole
Reinforcement
Learning**

Temporal Difference

Обучение с подкреплением



Задача коммивояжёра



Сколько решений для N вершин?

| | 1 | 2 | 3 | 4 |
|---|---|----|----|----|
| 1 | 0 | 10 | 15 | 20 |
| 2 | 5 | 0 | 9 | 10 |
| 3 | 6 | 13 | 0 | 12 |
| 4 | 8 | 8 | 9 | 0 |

Подходы к оптимизации транспорта

```
print("checkpoint saved at", checkpoint)

agent_timesteps_total: 4000
custom_metrics: {}
date: 2021-11-30_22-49-41
done: false
episode_len_mean: 21.11111111111111
episode_media: {}
episode_reward_max: 68.0
episode_reward_mean: 21.11111111111111
episode_reward_min: 8.0
episodes_this_iter: 189
episodes_total: 189
experiment_id: da8a46474b184f019acc264a210af66a
hostname: MacBook-Pro-Vladimir.local
info:
  learner:
    default_policy:
      custom_metrics: {}
    learner_stats:
      allreduce_latency: 0.0
```

TensorSpace Airplane .
A PEN BY Vladimir Sudakov

HTML

```
src="https://tensorspace.org/assets/jslib/tensorspace.min.js">
```

JS

```
filters: 6, strides: 1 } );  
model.add( new TSP.layers.Conv2d( { shape : [ 18, 18, 20 ] } ) );  
model.add( new TSP.layers.Activation2d( { shape : [ 18, 18 ] } ) );
```

Выбрана задача Оценка грузового ЛА

DSS Задачи Модели Экспертные оценки Об авторе Студенты

Список критерии

| ID | Лист дерева? | Имя |
|-----|--------------|---------------------------------|
| 196 | Нет | Integral assessment of aircraft |
| 197 | Нет | Profitability |
| 198 | Нет | Specifications |
| 199 | Да | Стоимость владения руб/мес |
| 200 | Да | Стоимость перевозки 1 кг |
| 201 | Да | Грузоподъемность |

Last Tour | Nodes: 6 | Tour Distance: 537.123

longitude

latitude

Best Tour | Nodes: 31 | Tour Distance: 3704.945

longitude

latitude

Преимущества решения



Масштабируемость на вычислительном кластере



В большинстве случаев задачу можно описать без использования сложных аналитических математических моделей, а сразу на языке программирования

Определение СППР

- СППР (DSS) – автоматизированная система, которая *помогает лицам, принимающим решения (ЛПР)*, использовать данные и модели, чтобы решать *слабоструктурированные и неструктурированные* проблемы.
- Слабоструктурированные проблемы – это проблемы, которые содержат как количественные так и качественные переменные. Причем качественные аспекты проблемы имеют тенденцию доминировать.
- СППР и экспертная система не одно и тоже!

Классификация на уровне пользователя

- Пассивная СППР - система, которая помогает процессу принятия решения, но не может вынести предложение, какое решение принять.
- Активная СППР - может сделать предложение, какое решение следует выбрать.
- Кооперативная СППР - позволяет ЛПР изменять, пополнять или улучшать решения, предлагаемые системой, посылая затем эти изменения в систему для проверки. Система изменяет, пополняет или улучшает эти решения и посыпает их опять пользователю. Процесс продолжается до получения согласованного решения.

Предварительные шаги

- Проанализировать текущую ситуацию (предпроектное исследование, мониторинг, онтология предметной области)
- Определить цели
- Как определить степень достижения цели?
- Критерий (др.-греч. κριτήριον — способность различения, средство суждения, мерило) — признак, основание, правило принятия решения по оценке чего-либо на соответствие предъявленным требованиям.

Критерии

- Критерий – позволяет оценить степень достижения цели/целей

Отношение критерий-цели – многие-ко-многим

Различают задачи:

- Однокритериальные – решать проще
- Многокритериальные – решать сложнее

Требования к критериям

- Полнота – набор критериев должен охватывать все существенные аспекты решаемой задачи.
- Действенность – набор критериев может быть с пользой применен при анализе задачи.
- Разложимость – набор критериев можно разбить на части, чтобы упростить решаемую задачу.
- Не избыточность – критерии не должны дублировать одни и те же аспекты решаемой задачи.
- Минимальность – векторный критерий должен иметь по возможности минимальную размерность.

Атрибуты критерия

- имя критерия,
- тип (лексический или численный),
- шкала,
- единицы измерения,
- неформальное текстовое описание,
- минимальное и максимальное значение,
- направление улучшения (на минимум или на максимум) или идеальное значение.

Формальная постановка

$$W = \Phi(U, S, C)$$

- Φ - некоторый функционал
- U - вектор управления
- S - вектор, характеризующий внешнюю среду
- C - вектор, характеризующий процесс (систему)

В общем случае W - векторный критерий

$$W = (w_1, w_2, \dots, w_n)$$

Альтернатива

- Альтернатива - каждое из несовместных решений отображаемое точкой критериального пространства.
- Совокупность всех точек представляет собой полное множество альтернатив. Оно содержит как реализуемые, так и не реализуемые решения. Альтернативами являются как решения по выбору управления, так и решения по выбору структуры или параметров управляемой системы.
- Основная задача принятия решений - задача оптимизации или ранжирования альтернатив.

Пусть $A = (a_1, a_2, \dots, a_n)$ – множество реализуемых альтернатив.

Тогда оптимальная альтернатива запишется в виде

$$a^0 = \arg \max_{a} F(a)$$

где $F(a)$ - значение критерия при альтернативе a .

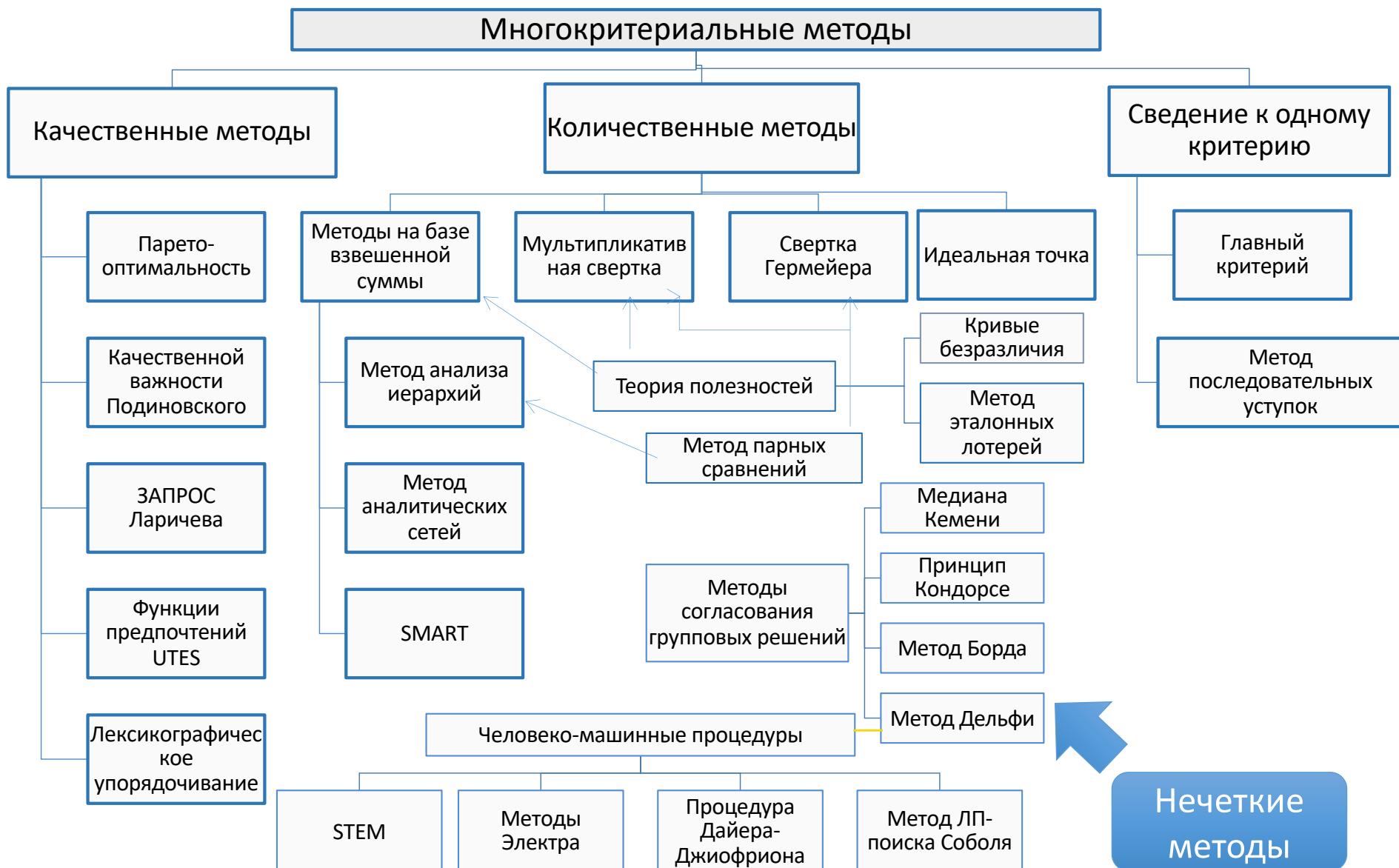
Оптимальность по Парето

- Альтернатива A1 **доминирует** над альтернативой A2, если по всем показателям (локальным критериям) A1 не уступает A2, а хотя бы по одному из них лучше.
- Решение следует искать среди недоминируемых альтернатив

Проблемы:

- Недоминируемых альтернатив может быть много
- Нет способа ранжировать внутри множеств недоминируемых и доминируемых альтернатив

Методы поддержки решений



Взвешенная сумма

$$W = \sum_{i=1}^n \alpha_i u_i , \text{ где}$$

$u_i = (w_i - w_{i \min}) / (w_{i \max} - w_{i \min})$ если $w_i \rightarrow \max$

$u_i = (w_{i \max} - w_i) / (w_{i \max} - w_{i \min})$ если $w_i \rightarrow \min$

$$\sum \alpha_i = 1$$

непосредственное назначение для $n < 10$

Нахождение весов

- Матрица парных сравнений
- Элементы матрицы a_{ij} показывают во сколько раз i -й показатель важнее j -го

$$x_i = \sqrt[n]{\prod_{j=1}^n a_{ij}} \quad \alpha_i = \frac{x_i}{\sum_n x_i}$$

$$L_{\max} = \sum_{i=1}^n \left(\alpha_i \left(\sum_{j=1}^n a_{ij} \right) \right) \quad IC = \frac{L_{\max} - n}{n - 1}$$

Прочие количественные методы

- Свертка Гермейера $W = \min \alpha_i u_i$
- Мультипликативная свертка $W = \prod u_i^{\alpha_i}$
- Идеальная точка

$$d(w_1, w_2, \dots, w_n) = \sqrt{(w_1 - w_1^u)^2 + (w_2 - w_2^u)^2 + \dots + (w_n - w_n^u)^2}$$

Медиана Кемени

$$d(A, B) = \frac{1}{2} \sum_{i,j=1}^N |a_{ij} - b_{ij}|, \text{ где } N - \text{число альтернатив}$$

$$a_{ij} = \begin{cases} 1 & \text{если } i - \text{я альтернатива предпочтительней } j - \text{й} \\ 0 & \text{если } i - \text{я альтернатива эквивалентна } j - \text{й} \\ -1 & \text{если } j - \text{я альтернатива предпочтительней } i - \text{й} \end{cases}$$

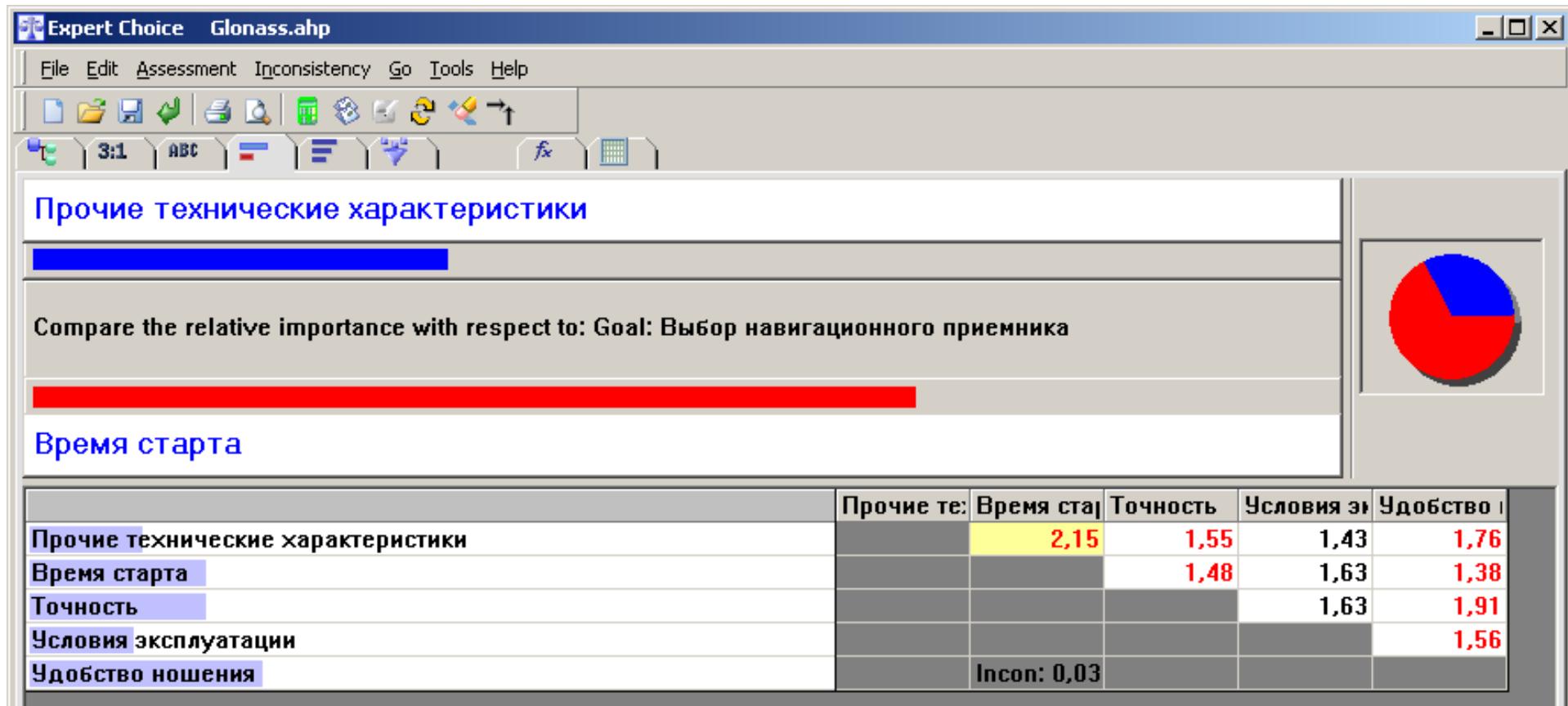
$$\min_K \sum_{i=1}^n d(K, R_i)$$

где K – среднее ранжирование(медиана Кемени),
R_i – ранжирование полученное экспертом

Некоторые другие методы

- Главный критерий
- Метод последовательных уступок
- Критерий «эффективность - стоимость»
- Качественный учет важностей Подиновского
- Функция полезности, метод эталонных лотерей
- Методы группы Электра
- Нечеткие модификации....

Пример парных сравнений



ФУНКЦИЯ ПРЕДПОЧТЕНИЙ

ФП

Значение [Очень хорошо]

Фиксируемые показатели

| Наименование | Значение |
|--------------|----------|
| Показатель 3 | Отлично |

Изменить значение фиксированного показателя

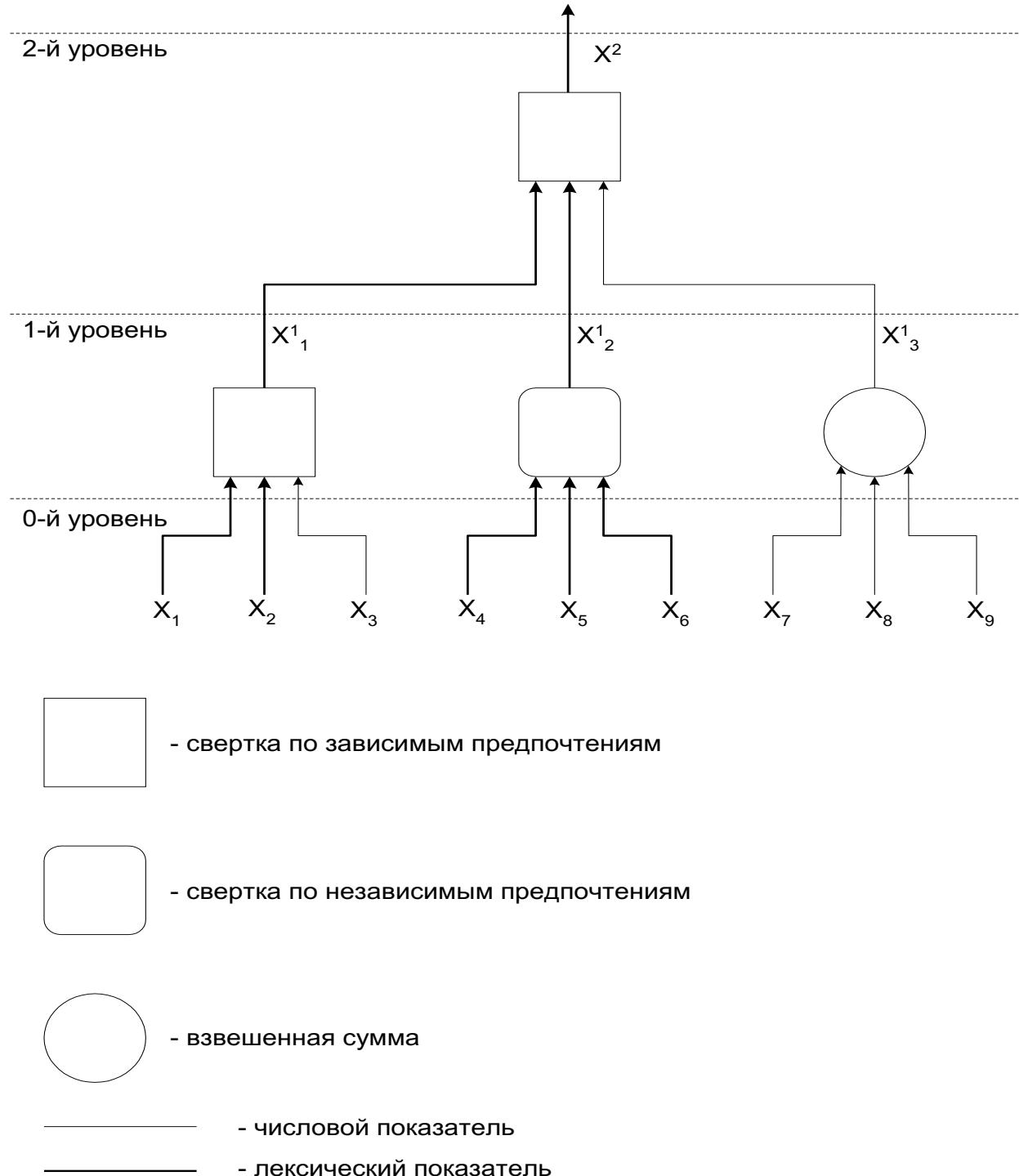
| | |
|--------------|---------|
| Показатель 3 | Отлично |
|--------------|---------|

Ось X [Показатель 1]

Ось Y [Показатель 2]

| | Недопуск | Плохо | Удовлетворительно | Хорошо | Отлично |
|-------------------|----------|-------|-------------------|--------|---------|
| Недопустимо | ■ | ■ | ■ | ■ | ■ |
| Плохо | ■ | ■ | ■ | ■ | ■ |
| Удовлетворительно | ■ | ■ | ■ | ■ | ■ |
| Хорошо | ■ | ■ | ■ | ■ | ■ |
| Отлично | ■ | ■ | ■ | ■ | |

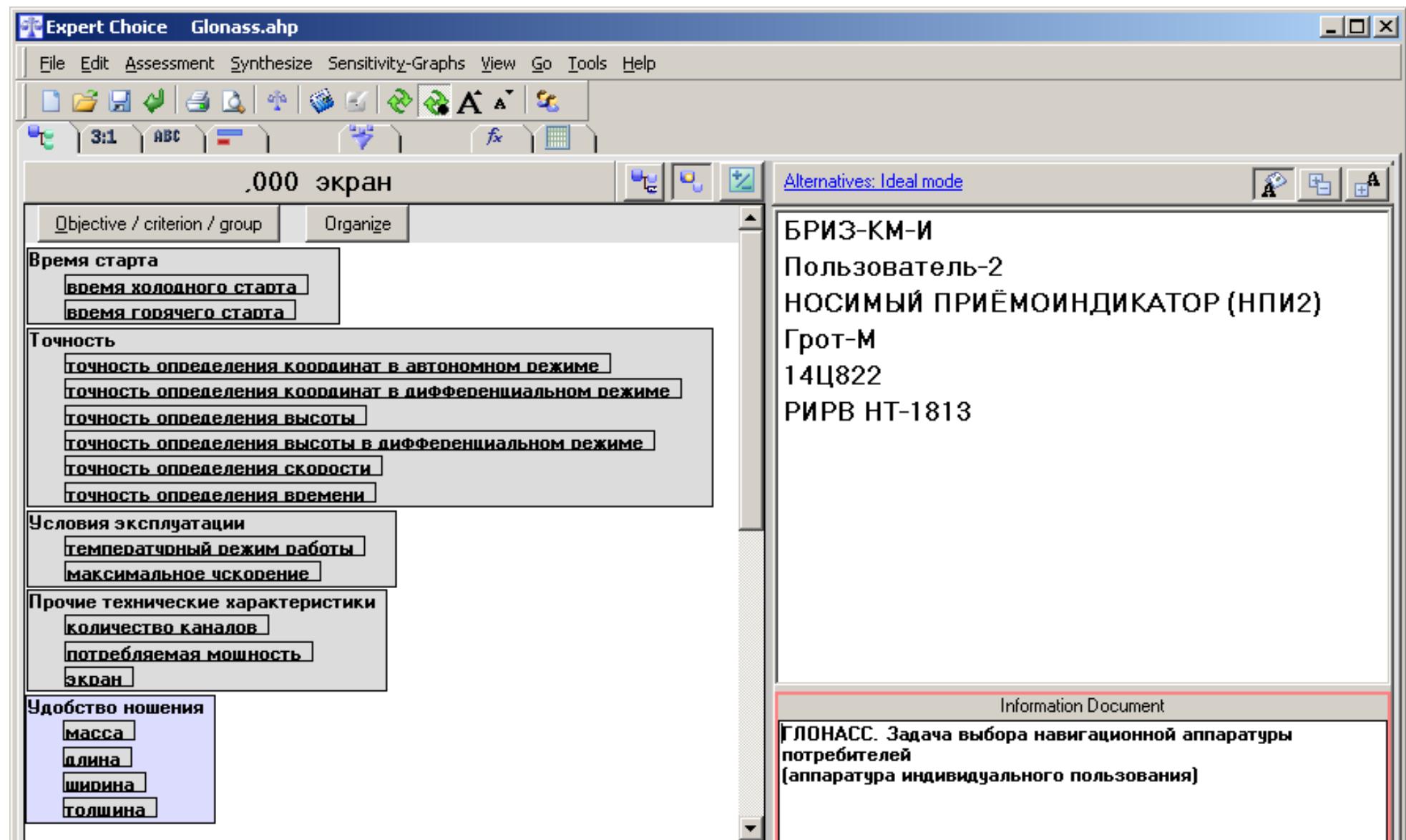
Агрегирование критериев



СППР оболочки

- Expert Choice
- Super Decisions
- Criterium Decision Plus
- Decision Lens
- ИАС “ОЦЕНКА и ВЫБОР”
- «Быстрый прототип»
- DASS
- FEASIBLE GOALS
- ПК MOVI
- Web-HIPRE

Фрагменты интерфейсов



Фрагменты интерфейсов

Screenshot of the Criterium DecisionPlus software interface showing a hierarchy of prioritization criteria.

The main window title is "Hierarchy - Q:\CDP Marketing\Marketing & Mailing Materials\CDP 3.0 release...".

The menu bar includes: File, Edit, View, Block, Level, Model, Results, Analysis, Window, Help.

The toolbar includes: New, Open, Save, Print, Prevw, Snap, Undo, Navig, Options, Rate, Scores, Help.

The left sidebar shows a "Navigator" pane with a tree view and a red rectangular selection box.

The central pane displays a hierarchical structure:

- "prioritize" (highlighted with a red box) branches into:
 - "Status"
 - "market necessity"
 - "desirability"
 - "costs"
 - "difficulty"
- "market necessity" branches into:
 - "User Demand"
 - "design priority"
 - "bug priority"
- "desirability" branches into:
 - "user path proximity"
 - "frequency"
 - "repro hours"
- "costs" branches into:
 - "testing hours"
 - "develop hours"
 - "scope"
- "difficulty" branches into:
 - "complexity"
 - "Alters Database"
 - "New DA Algorithms"

A vertical list on the right lists "Record # 94" through "Record # 109" corresponding to the items in the hierarchy.

At the bottom, status indicators include: Kelley Bevans InfoHarvest, Inc., Hierarchy - CDP Roll Out Prioritization, SMART, WEIGHTS, Connected, Rated, and a page number "1 96".

Фрагменты интерфейсов

Глонасс - СППР DSS-UTES 2.0

Файл Показатель Вид Помощь

Структура критерия

- Оценка приемника ГЛОНАСС
 - Точность
 - точность определения координат в автономном режиме
 - точность определения координат в дифференциальном режиме
 - точность определения высоты
 - точность определения высоты в дифференциальном режиме
 - точность определения скорости
 - точность определения времени
 - Прочие технические характеристики
 - количество каналов
 - потребляемая мощность
 - дисплей
 - температурный режим работы
 - максимальное ускорение
 - Удобство эксплуатации
 - масса
 - Время старта
 - холодный старт
 - горячий старт
 - Габариты
 - длина
 - ширина
 - высота

Показатели (кол-во:23)

| Имя | 0... | Тип | Исп. | Ч.град. | Метод |
|--|------|-----|------|---------|-------------------------------|
| количество каналов | Ч... | Да | | | |
| холодный старт | Ч... | Да | | | |
| горячий старт | Ч... | Да | | | |
| точность определения координат в автономном режиме | Ч... | Да | | | |
| точность определения координат в дифференциальном режиме | Ч... | Да | | | |
| точность определения высоты | Ч... | Да | | | |
| точность определения высоты в дифференциальном режиме | Ч... | Да | | | |
| точность определения скорости | Ч... | Да | | | |
| точность определения времени | Ч... | Да | | | |
| потребляемая мощность | Ч... | Да | | | |
| температурный режим работы | Л... | Да | | | |
| максимальное ускорение | Ч... | Да | | | |
| длина | Ч... | Да | | | |
| ширина | Ч... | Да | | | |
| высота | Ч... | Да | | | |
| масса | Л... | Да | | | |
| Время старта | 0... | Да | | | Свертка по зависимым предпочт |
| Точность | 0... | Да | | | Взвешенная сумма (непосредств |
| Прочие технические характеристики | 0... | Да | | | Взвешенная сумма (непосредств |
| дисплей | Л... | Да | | | |
| Удобство эксплуатации | 0... | Да | | | Свертка по зависимым предпочт |
| Габариты | 0... | Да | | | Взвешенная сумма (непосредств |
| Оценка приемника ГЛОНАСС | 0... | | | | Свертка по зависимым предпочт |

Фрагменты интерфейсов

| | Priority | Value | |
|--|----------|--------|---|
| точность определения координат в автономном режиме | 0,125 | 0,1250 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| точность определения координат в дифференциальном режиме | 0,125 | 0,1250 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| точность определения высоты | 0,25 | 0,2500 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| точность определения высоты в дифференциальном режиме | 0,25 | 0,2500 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| точность определения скорости | 0,125 | 0,1250 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| точность определения времени | 0,125 | 0,1250 | <div style="width: 100%; background-color: #00008B; height: 10px;"></div> |
| Total | 1 | 1,0000 | |

Фрагменты интерфейсов

Expert Choice Glonass.ahp

File Edit Assessment Inconsistency Go Tools Help

3:1 ABC F E Y f fx

точность определения координат в автономном режиме

точность определения координат в дифференциальном режиме

Compare the relative preference with respect to: Точность

| | точность о | точность о | точность о | точность о | точность о | точность о |
|--|------------|-------------|------------|------------|------------|------------|
| точность определения координат в автономном режиме | | 1,0 | | | | |
| точность определения координат в дифференциальном режиме | | | 2,0 | | | |
| точность определения высоты | | | | 1,0 | | |
| точность определения высоты в дифференциальном режиме | | | | | 2,0 | |
| точность определения скорости | | | | | | 1,0 |
| точность определения времени | | Incon: 0,00 | | | | |

Фрагменты интерфейсов

Expert Choice Glonass.ahp

File Edit Assessment View Go Plot Set Tools Formula Type Totals Help

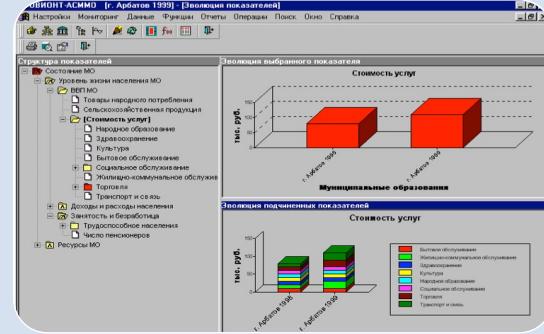
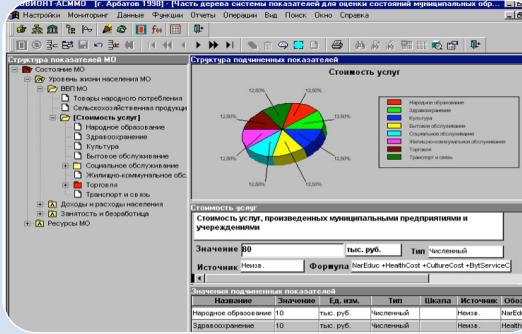
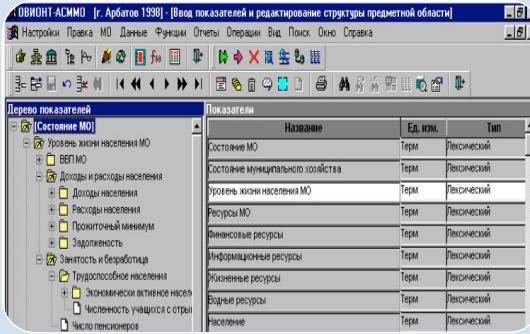
Plot Covering Objective Priorities

Move ← → ↑ ↓

| Ideal mode | | PAIRWISE | PAIRWISE | INCR | DIRECT |
|----------------|-------|------------------------------|---------------------------------|-----------------|----------------------------------|
| Alternative | Total | количество каналов (L: ,105) | потребляемая мощность (L: ,291) | экран (L: ,605) | время холодного старта (L: ,373) |
| БРИЗ-КМ-И | ,179 | ,987 | ,200 | 1 | 1 |
| Пользователь-2 | ,125 | ,537 | ,600 | 1 | 0,2 |
| НОСИМЫЙ | ,149 | ,272 | ,800 | 23 | 0,3 |
| Грот-М | ,136 | ,575 | ,600 | 2 | 0,2 |
| 14Ц822 | ,173 | 1,000 | 1,000 | 1 | 0,5 |
| РИРВ НТ-1813 | ,141 | ,541 | ,400 | 1 | 0,5 |

[Needs re-extraction]

Автоматизированная система мониторинга муниципальных образований (АСМ МО)

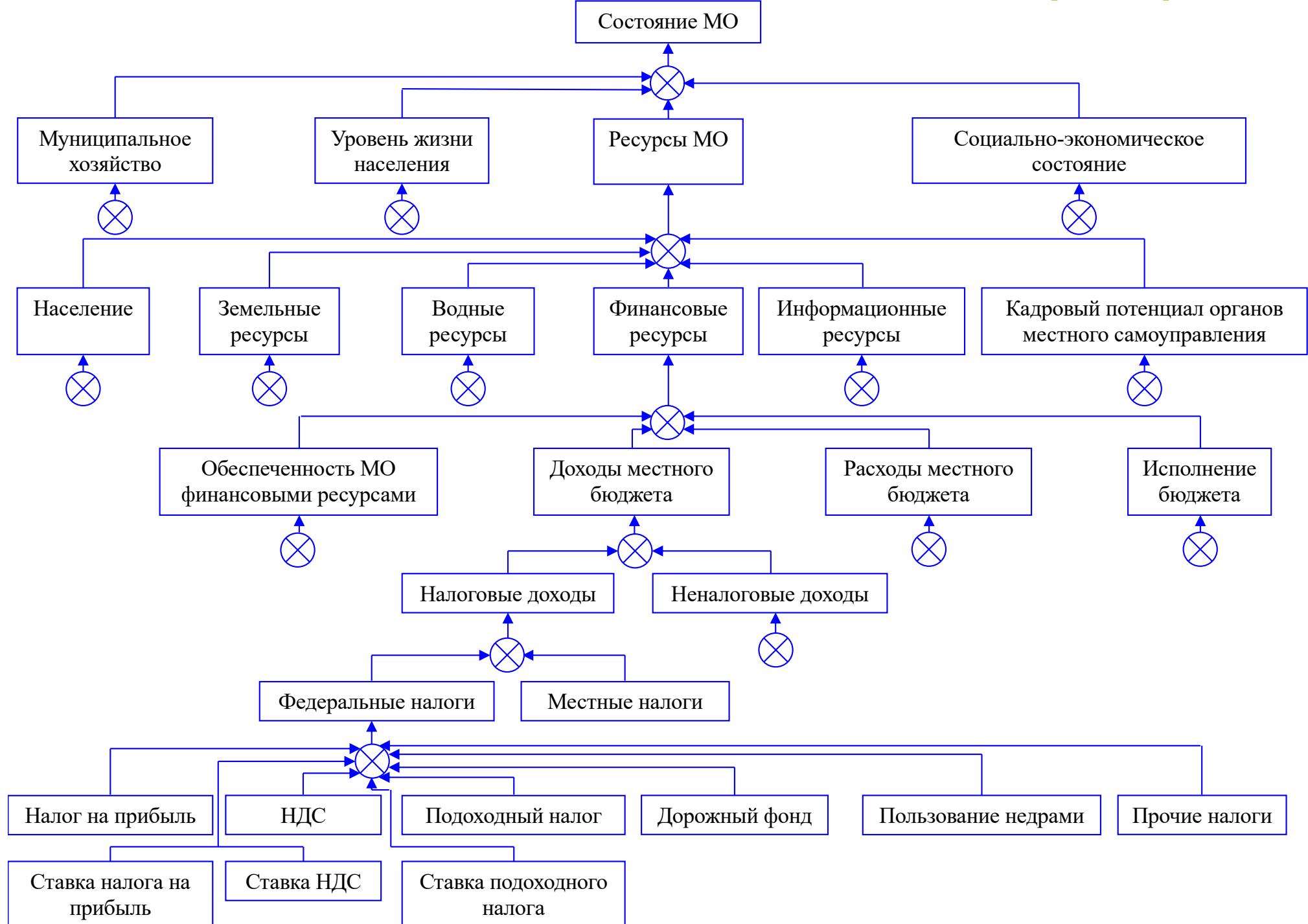


Позволяет осуществлять мониторинг муниципальных образований по 200-м показателям

АСМ МО создавалась по заказу МИННАЦа и эксплуатировалась в ВЦМИ и в ряде городов: Иваново, Псков, Печоры, Дзержинский

Средства разработки PowerBuilder, MS SQL
Разработка БД заняла около 2-х месяцев

Система основных показателей (СОП)



Автоматизированная система мониторинга муниципальных образований (АСМ МО)

ОВИОНТ-АСММО [г. Пермь от 12.12.00]

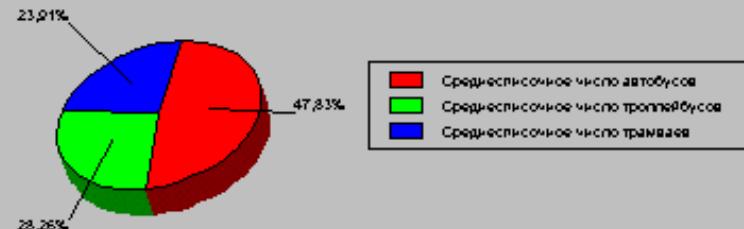
Настройки Мониторинг Данные Функции Помощник Отчеты Операции Вид Поиск Окно Справка

Структура показателей МО

- Состояние МО
 - Состояние муниципального хозяйства
 - Общественный порядок и безопасность личности
 - Состояние муниципальных учреждений
 - Обеспеченность услугами
 - Состояние банно-прачечного и гостиничного хозяйства
 - Состояние муниципального транспорта
 - Обеспеченность жителей городским пасс. транс
 - [Среднесписочное число машин гор. пасс]
 - Среднесписочное число автобусов
 - Среднесписочное число троллейбусов
 - Среднесписочное число трамваев
 - Соотношение доходов и расходов транспортных предприятий
 - Число пассажиров, перевезенных городским транспортом
 - Число платных пасс., перевез. городским транспортом
 - Число бесплатных пасс., перевез. городским транспортом
 - Сост. муниципального дор. строительства и благоустройства
 - Качество условий жилья
 - Состояние жилищного фонда
 - Уровень озеленения и окружающая среда
 - Уровень жизни населения МО
 - ВВП МО
 - Доходы и расходы населения
 - Занятость и безработица
 - Ресурсы МО
 - Финансовые ресурсы
 - Информационные ресурсы

Структура подчиненных показателей

Среднесписочное число машин гор. пасс. транспорта



Среднесписочное число машин гор. пасс. транспорта

Среднесписочное число машин городского пассажирского транспорта

Значение 230000 ед. Тип Численный

Источник Расчетный | Формула AvNumBus +AvNumTrolley +AvN

Значения подчиненных показателей

| Название | Значение | Ед. изм. | Тип | Обозначение |
|------------------------------------|----------|----------|-----------|-------------|
| Среднесписочное число автобусов | 110000 | ед. | Численный | AvNumBus |
| Среднесписочное число троллейбусов | 65000 | ед. | Численный | AvNumTroll |
| Среднесписочное число трамваев | 55000 | ед. | Численный | AvNumTram |

Ready

Автоматизированная Система Мониторинга МО

Влияние изменения исходных показателей (Число автобусов, троллейбусов, трамваев)

| Наименование показателя | 12.12.1999 | 12.12.2000 |
|--|------------|---------------|
| Состояние МО | Неважно | Посредственно |
| Ресурсы МО | Терпимо | Терпимо |
| Уровень жизни населения МО | Удовл. | Удовл. |
| Муниципальное хозяйство | Плохо | Удовл. |
| Состояние муниципальных учреждений | Хорошо | Хорошо |
| Качество условий жилья | Неважно | Неважно |
| Обеспеченность услугами | Удовл. | Хорошо |
| Состояние банно-прачечного и гостиничного хозяйства | Хорошо | Хорошо |
| Состояние муниципального дорожного строительства и благоустройства | Удовл. | Удовл. |
| Состояние муниципального транспорта | 0,69 | 0,73 |
| Число пассажиров, перевезенных городским транспортом, млн. чел. | 561,8 | 572,8 |
| Соотношение доходов и расходов транспортных предприятий, % | 157,6 | 160,3 |
| Обеспеченность жителей городским пассажирским транспортом, шт./тыс. чел. | 1,24 | 1,317 |
| Среднегодовая численность населения, тыс. чел. | 1034,9 | 1031,4 |
| Среднесписочное число машин городского транспорта, шт. | 1283 | 1358 |
| Среднесписочное число автобусов, шт. | 911 | 989 |
| Среднесписочное число троллейбусов, шт. | 120 | 123 |
| Среднесписочное число трамваев, шт. | 252 | 246 |

Автоматизированная система мониторинга муниципальных образований (АСМ МО)

ОВИОНТ-АСММО [г. Пермь от 12.12.99] - [Эволюция показателей]

Настройки Мониторинг Данные Функции Помощник Отчеты Операции Поиск Окно Справка

Структура показателей

- Состояние МО
 - Состояние муниципального х
 - Общественный порядок и
 - Состояние муниципальн
 - Обеспеченность услугами
 - Состояние банно-прач
 - Состояние муниципал
 - Обеспеченность ж
 - [Среднесписочное число машин гор. пасс. транспорта]
 - Среднесписочное число машин гор. пасс. транспорта
 - Среднесписочное число автобусов
 - Среднесписочное число трамваев
 - Среднесписочное число троллейбусов
 - Соотношение дохс
 - Число пассажиров
 - Сост. муниципальног
 - Качество условий жилья
 - Уровень жизни населения МС
 - Ресурсы МО

Эволюция выбранного показателя

| | г. Пермь от 12.12.99 | г. Пермь от 12.12.00 |
|---|----------------------|----------------------|
| Среднесписочное число машин гор. пасс. транспорта | 210000 | 230000 |

Эволюция подчиненных показателей

Среднесписочное число машин гор. пасс. транспорта

Год

Муниципальные о

Среднесписочное число автобусов

Среднесписочное число трамваев

Среднесписочное число троллейбусов

Ready

Система Поддержки Принятия Решений DSS-UTES

Ввод предпочтений методом парных сравнений

Ввод предпочтений

Настройки Справка

| предпочтительнее критерий: | на | чем критерий: |
|----------------------------|------|-------------------|
| Зеленые насаждения | 3.00 | Болота и водоемы |
| Зеленые насаждения | 6.00 | Свободные участки |
| Сельхозугодья | 3.00 | Свободные участки |
| Сельхозугодья | 4.00 | Кустарники |
| Леса | 4.00 | Кустарники |

Укажите на сколько критерий "A" предпочтительнее критерия "B":

Критерий "A" предпочтительнее на: **Критерий "B"**

Зеленые насаждени = 3.00 Болота и водоемы

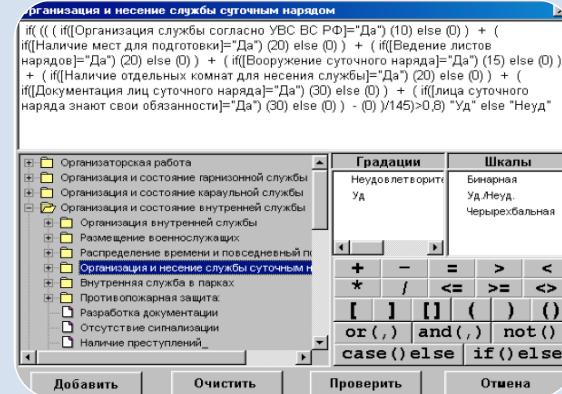
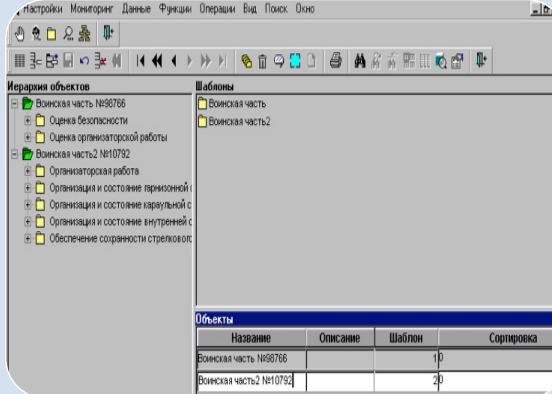
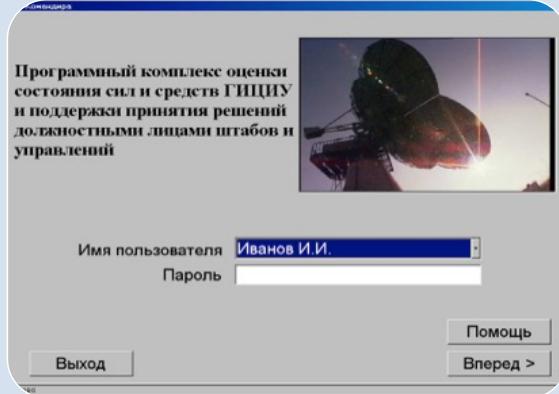
0 10.00

A <--> B

<< Предыдущая пара Следующая пара >>

OK

Автоматизированная система контроля и управления (АСКУ)



Создана в интересах
ГИЦИУ КС для
решения задач:

- Оценка боевой подготовки
- Оценка безопасности военной службы
- Оценка службы войск
- Оценка полевых районов применения

Предназначена для
мониторинга
состояния иерархии
структурных
подразделений
космических войск
(количество исходных
показателей порядка
100)

Для вычисления
обобщенных
показателей
использовались
продукционные
правила ЕСЛИ...ТО.

Средства разработки
PowerBuilder, Oracle
Разработка ИО около
1,5 мес.

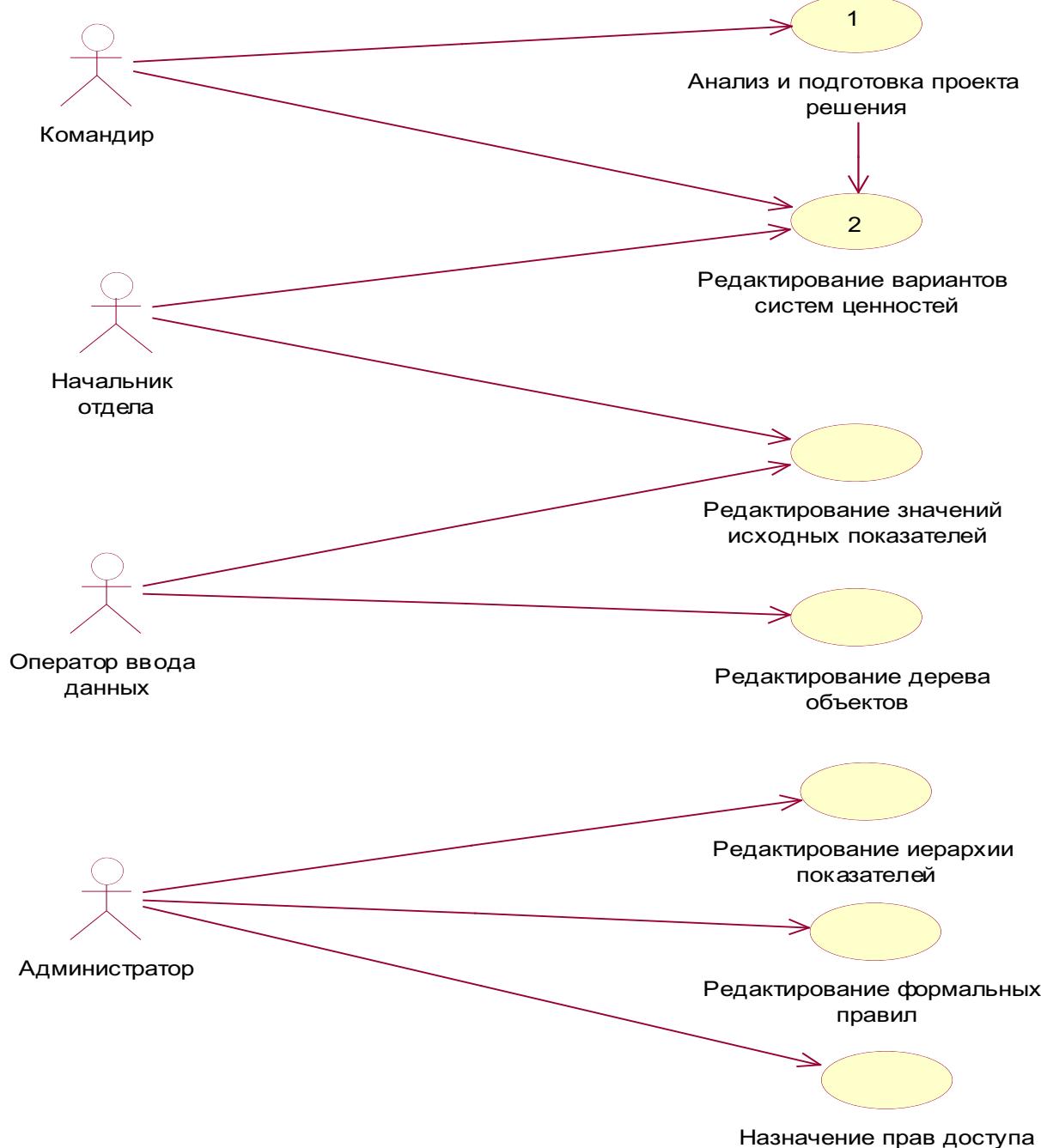
Задачи

- Оценка боевой подготовки
- Оценка безопасности военной службы
- Оценка службы войск
- Оценка полевых районов применения

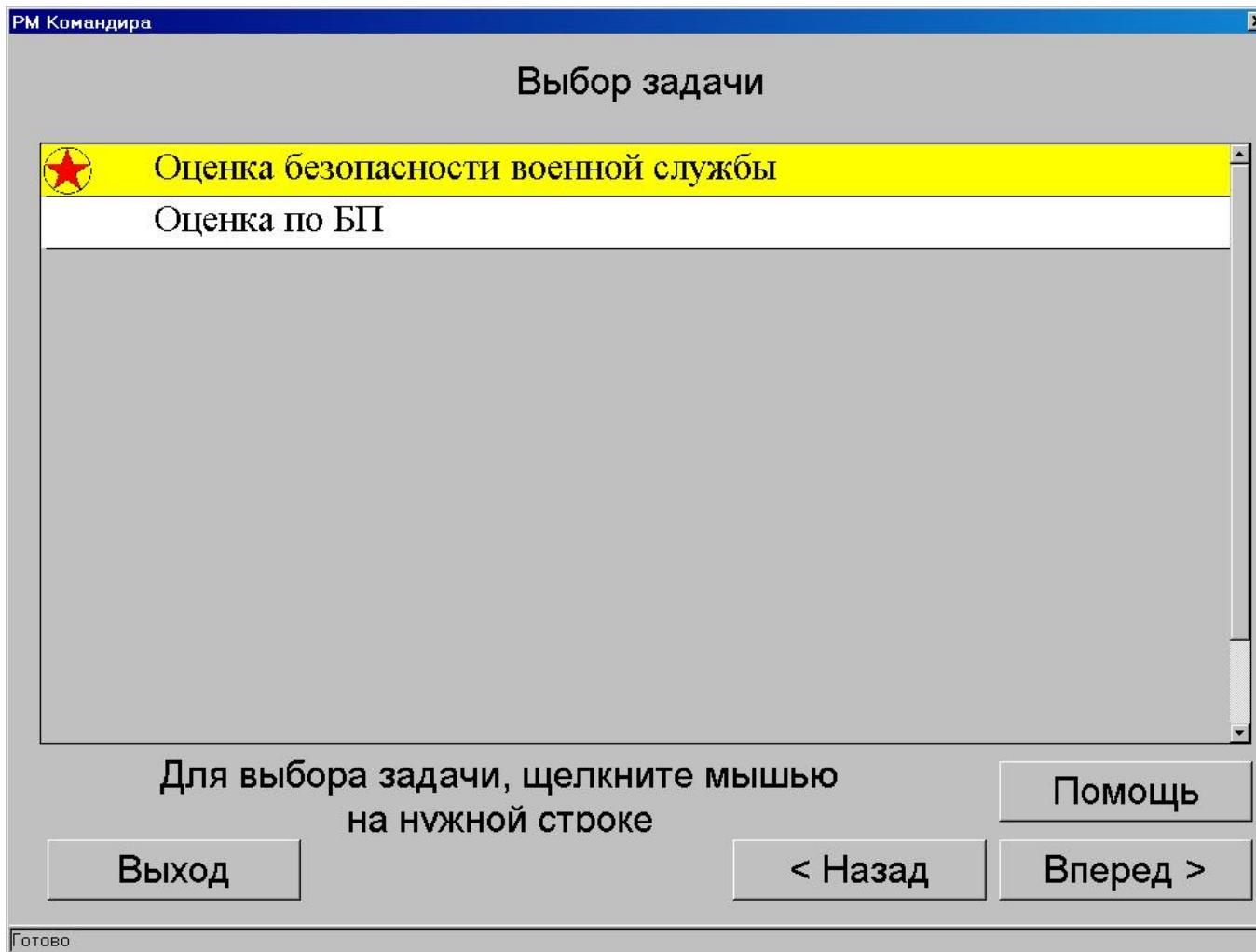
Возможности системы

- Оценка объектов иерархических структур по показателям
- Определение состояния сил и средств по методике и предпочтениям пользователя
- Анализ ситуации
- Подготовка предложений в проект решения

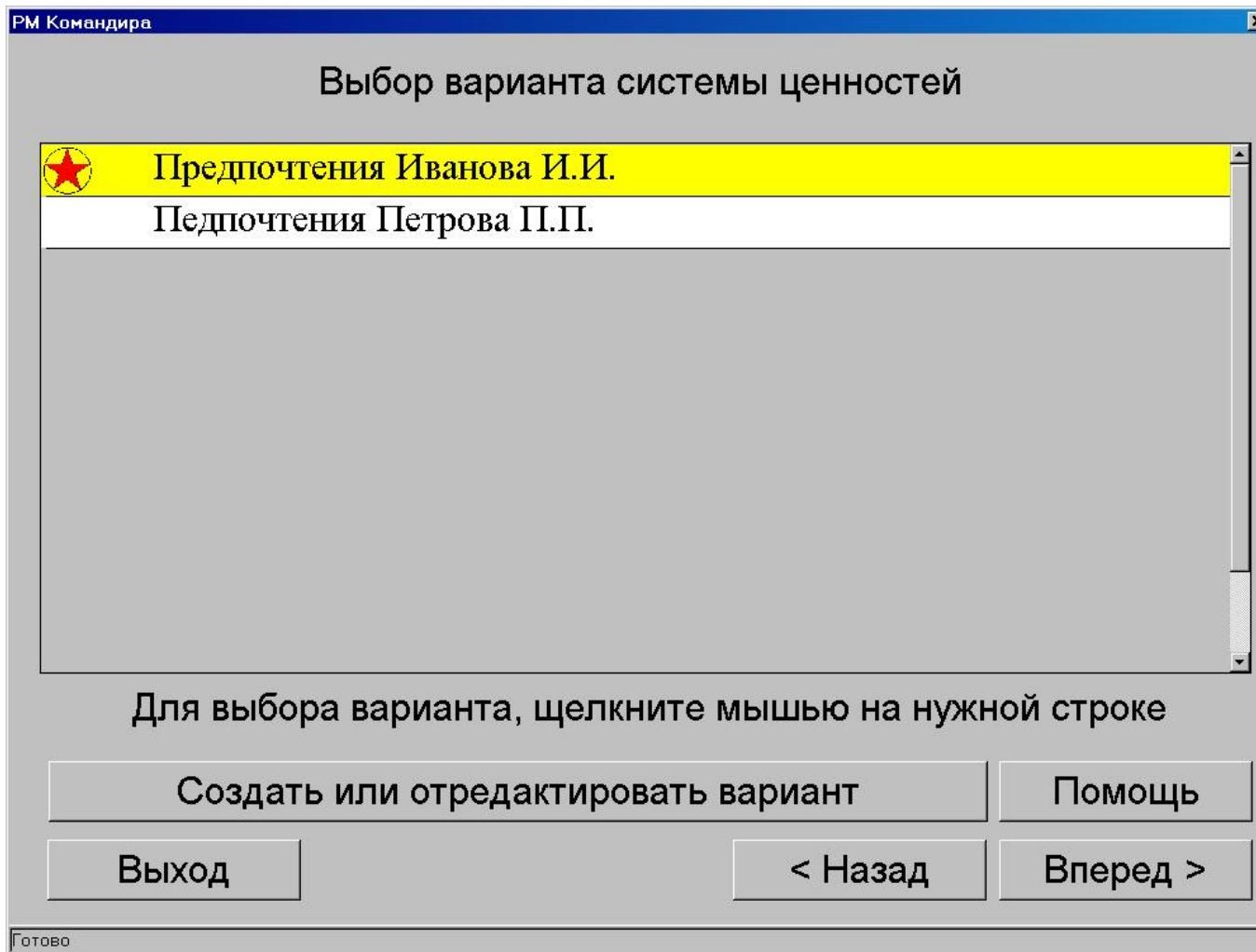
Диаграмма вариантов использования



АСКУ



АСКУ



АСКУ

PM Командира

Редактирование варианта

Имя варианта: Предпочтения Иванова И.И. Удалить вариант

| Метод свертки | |
|--|--|
| Оценка безопасности военной службы | Предпочтения. Иванов И.И.. 11.02.2003 |
| УПРАВЛЕНИЕ ОБЕСПЕЧЕНИЕМ БЕЗОПАСНОСТИ ВОЕННОЙ СЛУЖБЫ | Предпочтения. Иванов И.И.. 16.04.2003 |
| организация подготовки безопасности | Предпочтения. Иванов И.И.. 16.04.2003 |
| состояние условий безопасности военной службы | |

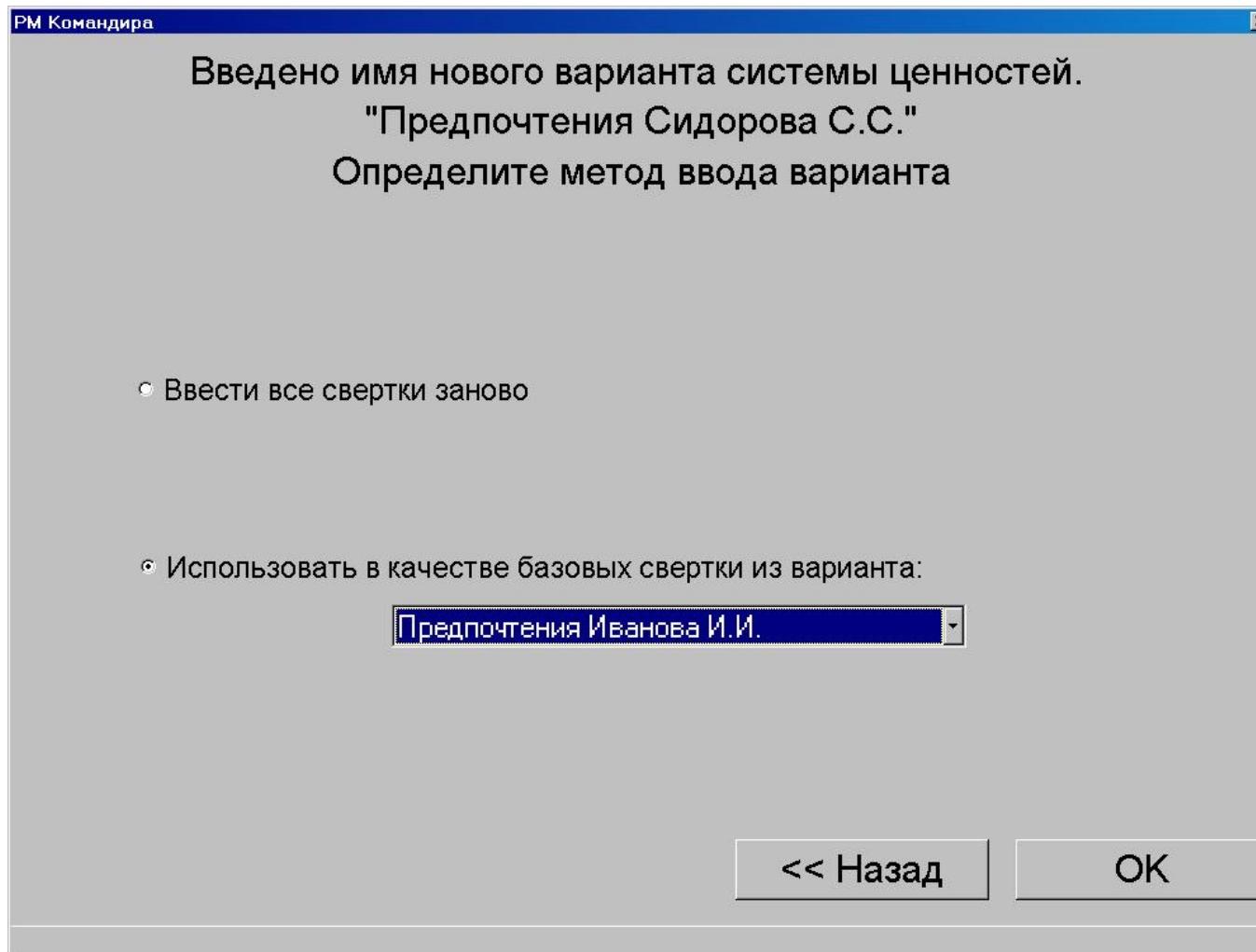
Для выбора варианта, щелкните мышью на нужной строке

Редактирование выбранного показателя Помощь

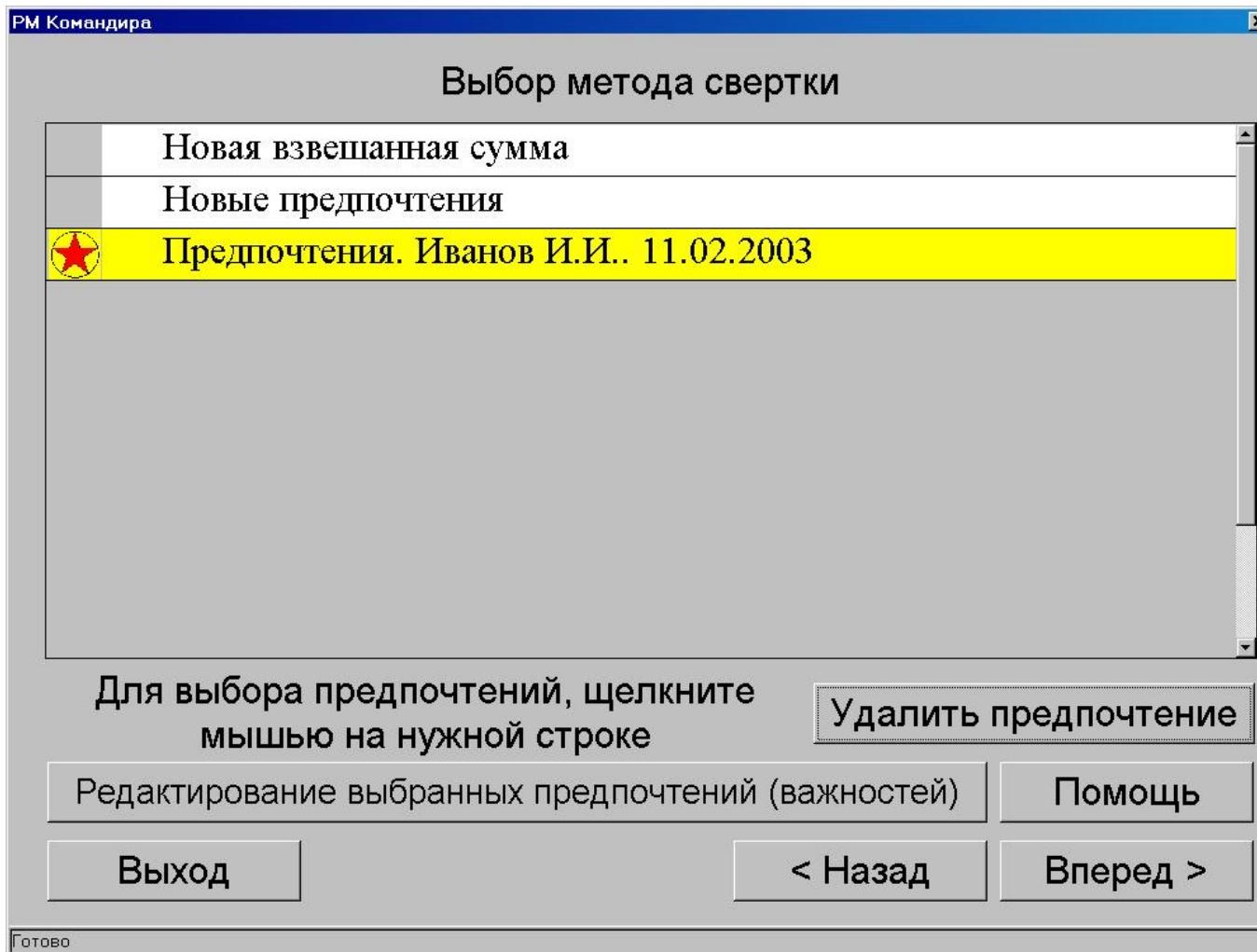
Выход < Назад Вперед >

2 compute

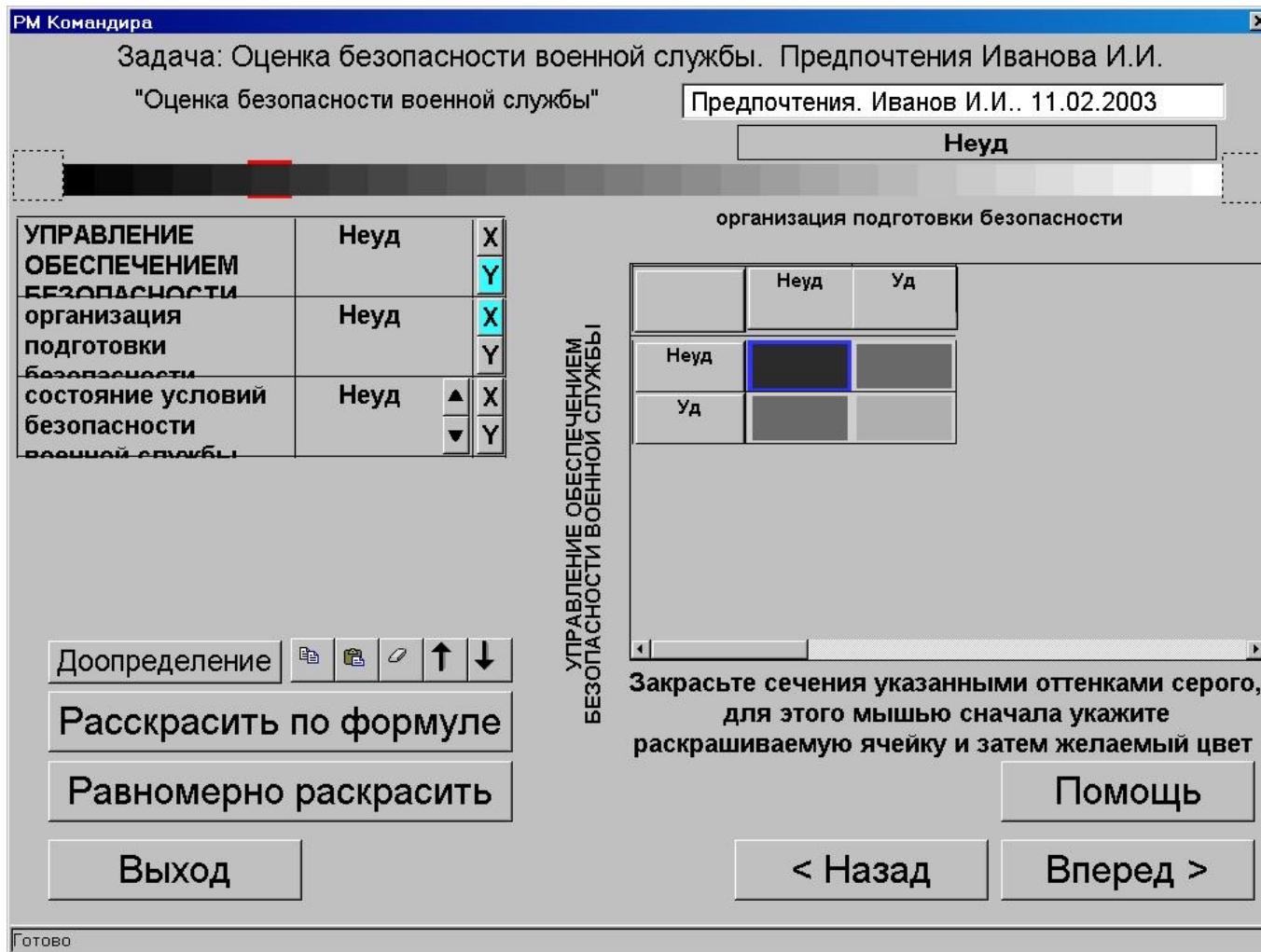
АСКУ



АСКУ



АСКУ



АСКУ

РМ Командира

"Оценка безопасности военной службы" (взвешенная сумма)
"УПРАВЛЕНИЕ ОБЕСПЕЧЕНИЕМ БЕЗОПАСНОСТИ" | Иванов И.И.. 16.04.2003

Сравниваемый показатель: Решение командира воинской части на обеспечение

| | Учёт сведений о состоянии безопасности военной службы | Планирование мероприятий обеспечения безопасности военной службы | Анализ состояния безопасности военной службы |
|--|---|--|--|
| Решение командира воинской части на обеспечение безопасности военной | 0.8 | 0.6 | 0.7 |

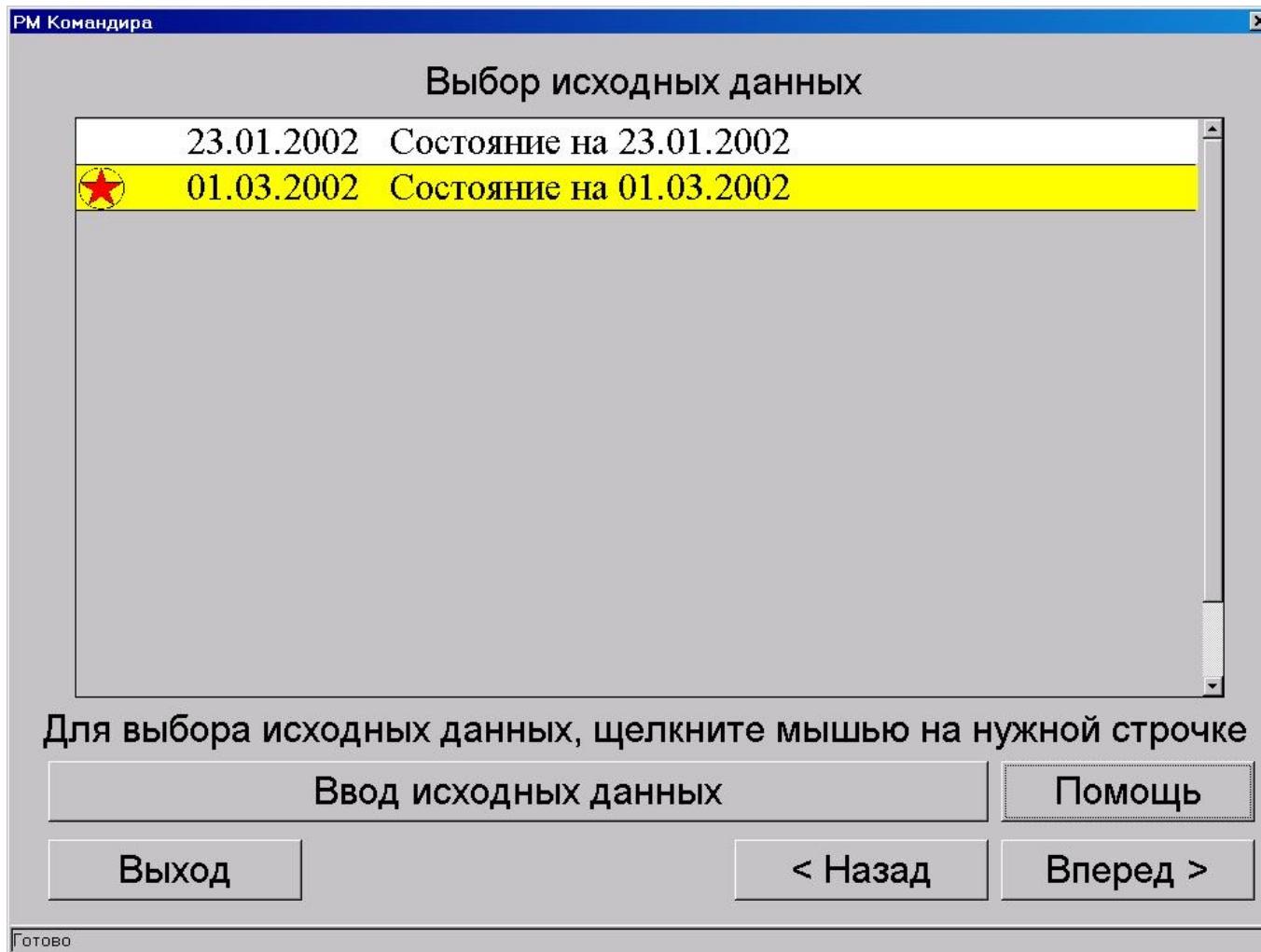
Введите, во сколько раз показатель в заголовке строки важнее показателя в заголовке столбца, для этого целкните мышью по кнопкам со стрелками

Помощь

Выход < Назад Вперед >

Готово

АСКУ



АСКУ

РМ Командира

Ввод исходных данных

| Показатель | Значение | Исх. |
|---|----------|-------------------------------------|
| Организация планирования работы по совершенствованию СВ | Нет | <input checked="" type="checkbox"/> |
| Организация подготовки и издания приказа по СВ | Нет | <input checked="" type="checkbox"/> |
| Организация работы по контролю выполнения приказов | Да | <input checked="" type="checkbox"/> |
| Организация подготовки суточного наряда | Нет | <input checked="" type="checkbox"/> |
| Организация контроля за несением КС и ВС | Да | <input checked="" type="checkbox"/> |
| Оценка недостатков организаторской работы командира | Хор | <input type="checkbox"/> |

все подчиненные

Сохранить **Заполнить** **Помощь**

Выход **< Назад** **Вперед >**

в/ч 01123
отдел ЭМО
ОТК
РО
в/ч 08317
в/ч 32103

в/ч 01123
в/ч 08317
в/ч 32103

Оценка состояния службы войск
Оценка организаторской работы командира по рукам
Оценка недостатков организаторской работы командира
Оценка организации и состояния ГС
Оценка организации и состояния КС
Оценка организации и состояния внутренней службы
Обеспечение сохранности стрелкового оружия и броеного имущества

Задание

Скачайте или выпишите предложения по продаже тех или иных товаров со значениями критериев их оценки из Интернет-магазинов. Товаров не менее 10, критериев не менее 7. Варианты заданий:

1. Ноутбуки.
2. Фотоаппараты.
3. Телефоны.
4. Планшеты.
5. Принтеры.
6. Холодильники.
7. Стиральные машины.

Допустимо предложить свой вариант товара.

Получить обобщённые оценки альтернатив(товаров) средствами электронных таблиц или системы программирования используя методы:

- 1) оптимальность по Парето,
- 2) взвешенная сумма,
- 3) взвешенная сумма с парными сравнениями,
- 4) свертка Гермейера,
- 5) мультипликативная свертка,
- 6) метод идеальной точки.

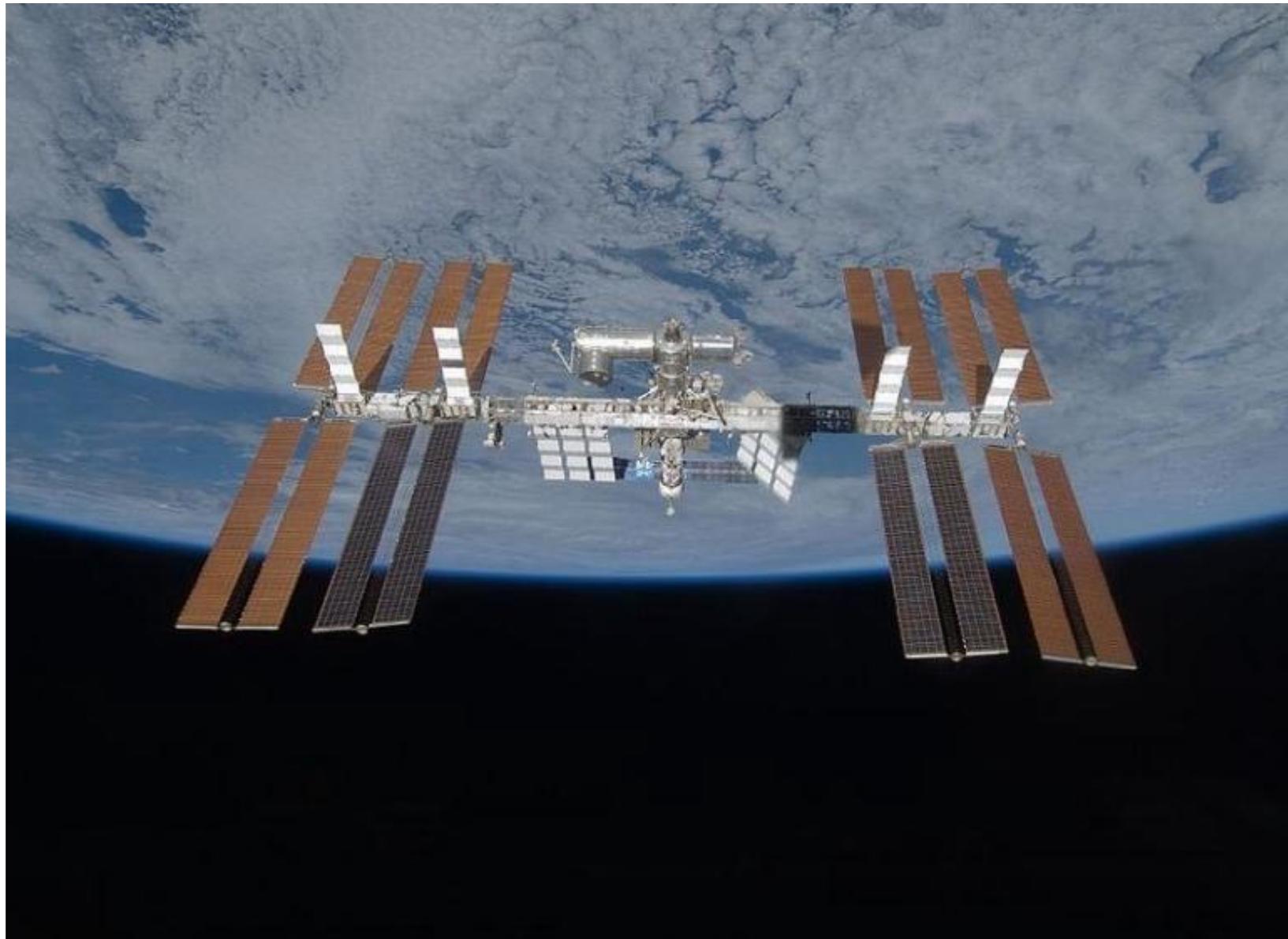
При решении допустимо предполагать, что число критериев и альтернатив фиксировано. Изменения значений критериев должны автоматически приводить к изменению итоговых рангов. В расчетах необходимо использовать агрегирующие функции везде где это возможно (СУММ(), СУММПРОИЗВ(), МИН() и аналогичные).

Применение какого метода дало лучшие, на Ваш взгляд, результаты? Чем это можно объяснить?

Модели исследования операций

- *Операция* — всякое мероприятие (система действий), объединённое единым замыслом и направленное к достижению какой-то цели
- Решение — всякий определённый набор зависящих от человека значений параметров
- Оптимальное — решение, которое по тем или другим признакам предпочтительнее других
- Цель исследования операций — предварительное количественное обоснование оптимальных решений с опорой на показатель эффективности

Планирование космических экспериментов (КЭ) на РС МКС



Редактор критериев

Яндекс

knts.tsniimash.ru/DSS/Criteria.aspx

Редактор критериев

Пользователь
tester

Настройки Выход

- Система поддержки принятия решений
- Задачи
- Альтернативы
- Критерии
- Экспертные оценки
- Ранжирование
- Пользователи
- Роли

Дерево критериев

- Оценка КЭ
 - Значимость эксперимента
 - Актуальность КЭ
 - Значимость результатов КЭ
 - Научный эффект КЭ
 - Прикладной эффект КЭ
 - Безопасность
 - Готовность КЭ к включению программы
 - Доступность информации по КЭ
 - Космическая деятельность
 - КЭ является международным
 - Наличие заинтересованного заказчика
 - Наличие КЭ в подпрограмме исследований
 - Соответствие КЭ направлениям других программ
 - Соответствие КЭ направлениям ФКП
 - Социальный эффект
 - Экологический эффект
 - Экономический эффект
 - Технологическая инновационность
 - Использование существующей техники
 - Наличие международного патента

Параметры

| Наименование | Описание |
|-----------------|----------|
| Актуальность КЭ | |

Родитель: Значимость эксперимента

Шкала

| № | Градация | Ранг | Удалить |
|---|--|------|---------|
| 1 | Оценка невозможна | 1 | X |
| 2 | Эксперимент в принципе полезен, однако его задержка не повлияет на темп исследований в данной области. | 2 | X |
| 3 | Задержка проведения КЭ затормозит, но не остановит дальнейшие исследования в данной области. | 3 | X |
| 4 | Задержка проведения КЭ приведет к соответствующей остановке исследования в данной области. | 4 | X |
| 5 | Срыв проведения эксперимента в заданные сроки приведет к принципиальной невозможности в течение многих лет получить требуемую информацию (например, вследствие изменения активности Солнца, или если аппаратура входит в состав МЛМ и, таким образом, сроки её | 5 | X |

Направление улучшений: Чем больше, тем лучше

Числовой показатель:

Порядковый номер:

Ресурсные ограничения

Редактор ресурсов

192.168.56.101:8080/DSS/Resource.aspx

Редактор ресурсов

Задача: Оценка КЭ
Пользователь: Администратор

Настройки Выход

- Система поддержки принятия решений
- Задачи
- Альтернативы
- Критерии
- Ресурсы
- Экспертные оценки
- Ранжирование
- Планирование
- Пользователи

Ресурсы

| Наименование | Описание | Значение | Период | Единица изменения периода | Критерий | Удалить |
|---|--|----------|--------|---------------------------|---|--------------------------|
| Масса доставляемой на борт НА (кг) | Масса доставляемой на борт НА (кг) | 0 | 2 | квартал | Масса доставляемой на борт НА (кг) | <input type="checkbox"/> |
| Объём доставляемой на борт НА (м³) | Объём доставляемой на борт НА (м ³) | 0 | 2 | квартал | Объём доставляемой на борт НА (м ³) | <input type="checkbox"/> |
| Масса возвращаемых блоков (кг) | Масса возвращаемых блоков (кг) | 0 | 2 | квартал | Масса возвращаемых блоков (кг) | <input type="checkbox"/> |
| Объём возвращаемых блоков (м³) | Объём возвращаемых блоков (м ³) | 0 | 2 | квартал | Объём возвращаемых блоков (м ³) | <input type="checkbox"/> |
| Мощность НА (кВт) | Потребляемая мощность НА (кВт) | 0 | 2 | квартал | Мощность НА (кВт) | <input type="checkbox"/> |
| Энергопотребление НА (кВт·ч) | Суммарное энергопотребление НА (кВт·ч) | 0 | 2 | квартал | Энергопотребление НА (кВт·ч) | <input type="checkbox"/> |
| Время реализации сеансов КЭ (мин) | Суммарное время реализации сеансов КЭ (мин) | 0 | 2 | квартал | Время реализации сеансов КЭ (мин) | <input type="checkbox"/> |
| Требуемое рабочее время экипажа (мин) | В том числе требуемое рабочее время экипажа на проведение КЭ (мин) | 0 | 2 | квартал | Требуемое рабочее время экипажа (мин) | <input type="checkbox"/> |
| Объём передаваемой информации (Гбайт) | Необходимый за время проведения КЭ объём передаваемой информации по радиолинии связи в (Гбайт) | 0 | 1 | день | Объём передаваемой информации (Гбайт) | <input type="checkbox"/> |

Формализация задачи планирования КЭ

- Целевая функция

$$\max_{x_j, t_j (j \in J)} \left\{ \sum_{j=1}^J x_j p_j; \sum_{j=1}^J x_j p_j (T - t_j) / T \right\}$$

- Ограничения

$$\sum_{j=1}^J x_j \left\{ \left(\sum_{t=(n-1)\Delta+1}^{n\Delta} f_j(t - t_j) \right) / \tau_j \right\} \leq Q_k, \quad (k=1,2,\dots,K; n=1,2,\dots,N),$$

$$\sum_{j=1}^J x_j f_j(t - t_j) v_{jl} \leq V_l, \quad (l=1,2,\dots,L; t=1,2,\dots,T),$$

$$t_j \geq \max_s \{ t_{ps[j,s]} + \tau_{ps[j,s]} \}, \quad (j=1, 2, \dots, J),$$

$$x_j \leq t_j \leq T x_j - \tau_j, \quad (j=1, 2, \dots, J),$$

$$x_j = 0 \text{ или } 1, \quad (j=1, 2, \dots, J).$$

Трудноразрешимые задачи, NP-полнота

| Функция временной сложности | Размер n | | | | | |
|-----------------------------|----------------|----------------|----------------|------------------|-----------------------------|----------------------------------|
| | 10 | 20 | 30 | 40 | 50 | 60 |
| n | 0,00001 сек | 0,00002 сек | 0,00003 сек | 0,00004 сек | 0,00005 сек | 0,00006 сек |
| n^2 | 0,0001 сек | 0,0004 сек | 0,0009 сек | 0,0016 сек | 0,0025 сек | 0,0036 сек |
| n^3 | 0,001 сек | 0,008 сек | 0,027 сек | 0,064 сек | 0,125 сек | 0,216 сек |
| n^5 | 0,1 сек | 3,2 сек | 24,3 сек | 1,7 мин | 5,2 мин | 13,0 мин |
| 2^n | 0,001 сек | 1,0 сек | 17,9 мин | 12,7 дней | 35,7 лет | 366 столетий |
| 3^n | 0,059 сек | 58 мин | 6,5 лет | 3855 столетий | 2×10^8 столетий | $1,3 \times 10^{13}$ столетий |

Влияние быстродействия ЭВМ

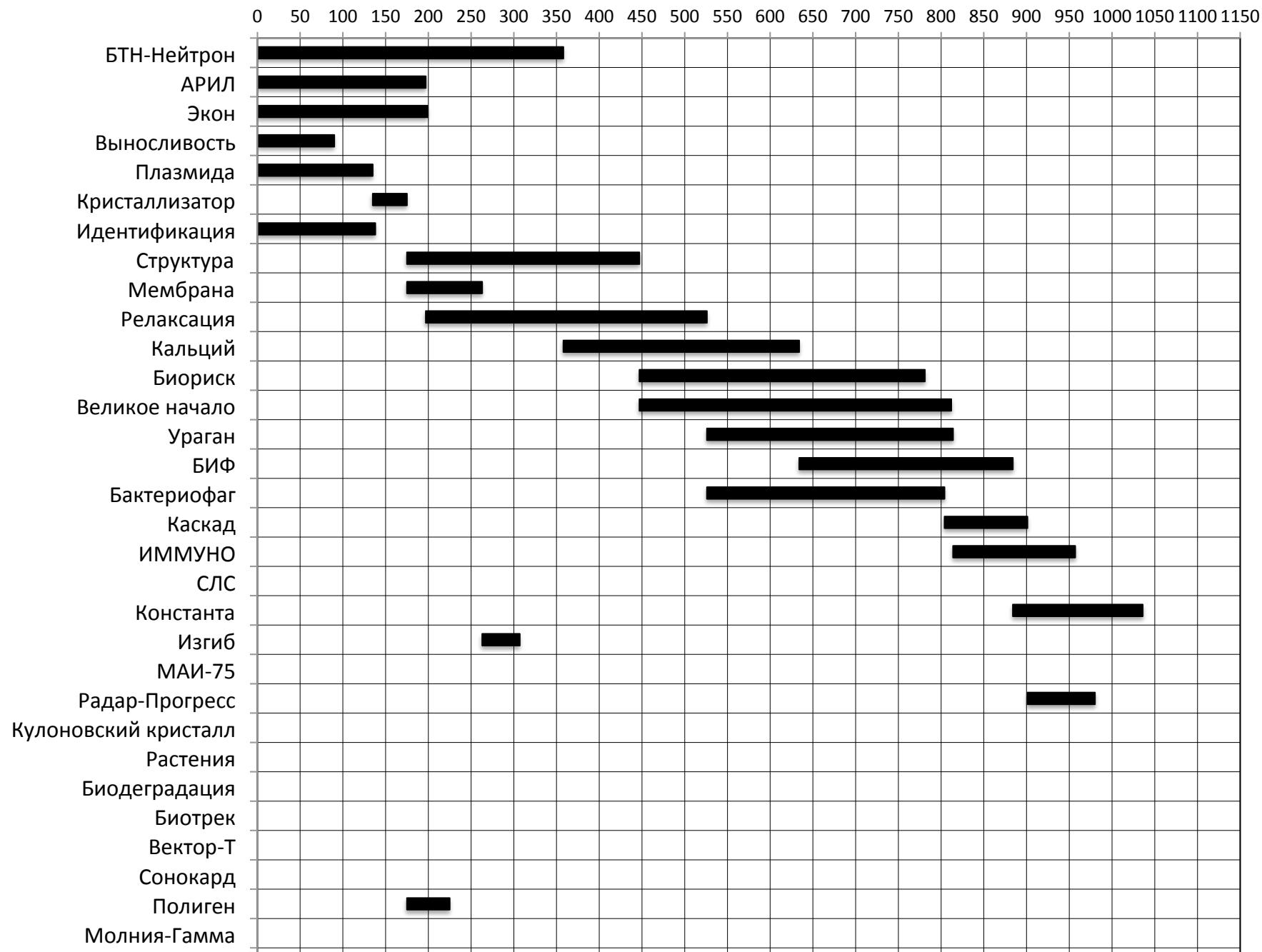
*Размеры наибольшей задачи,
разрешимой за один час*

| <i>Функция временной сложности</i> | <i>На современных ЭВМ</i> | <i>На ЭВМ, в 100 раз более быстрых</i> | <i>На ЭВМ, в 1000 раз более быстрых</i> |
|--|-------------------------------|--|---|
| n | N_1 | $100 N_1$ | $1000 N_1$ |
| n^2 | N_2 | $10 N_2$ | $31,6 N_2$ |
| n^3 | N_3 | $4,64 N_3$ | $10 N_3$ |
| n^5 | N_4 | $2,5 N_4$ | $3,98 N_4$ |
| 2^n | N_5 | $N_5 + 6,64$ | $N_5 + 9,97$ |
| 3^n | N_6 | $N_6 + 4,19$ | $N_6 + 6,29$ |

Экспоненциальная сложность на практике

- Некоторые задачи имеют экспоненциальную сложность, но это сложность для *наихудшего* случая
- Во многих практических задачах эта сложность не проявляется
- Проблема – нет способа спрогнозировать как поведет себя экспоненциальный алгоритм на тех или иных данных

Диаграмма Ганта



Решение транспортных задач в ws-dss

Решение оптимизационных задач в ws-dss



Входные данные:

```
{  
    "trace": 1,  
    "kn": 1,  
    "kria": 1,  
    "kpp": 1,  
    "dir": 1,  
    "c": [-3,-7, 1,-1],  
    "h": [1, 1, 1, 1],  
    "asb": [  
        [2, -1, 1, -1, 2, 1],  
        [-1, 1, -6, -4, 0, -8],  
        [-5, -3, 0, -1, 0, -5]  
    ]  
}
```

Выходные данные:

Метод неявного перебора по векторной решетке
SVL Версия 2.0
КН: ВКЛЮЧЕН
КПИА: ВКЛЮЧЕН
КПП: ВКЛЮЧЕН
Данные по задаче в исходной форме
Направление оптимизации: MAX
Коэффициенты целевой функции: -3 -7 1 -1
Ограничения на опт. переменные сверху: 1 1 1 1
Ограничения:
2 -1 1 -1 >= 1
-1 1 -6 -4 <= -8
-5 -3 0 -1 <= -5
Время начала работы: 11/04/2016 22:10:35
Преобразованная задача:
Направление оптимизации: MIN
Коэффициенты целевой функции: 3 7 1 1

Современные средства решения задач оптимизации

- AMPL - A Modeling Language for Mathematical Programming — язык моделирования для математического программирования
- GAMS - General Algebraic Modeling System – система моделирования для математического программирования и оптимизации
- JULIA - высокоуровневый высокопроизводительный свободный язык программирования с динамической типизацией, созданный для математических вычислений.
- ILOG CPLEX - это решение, предназначенное для быстрой разработки и развертывания моделей математического программирования и программирования в ограничениях. Сочетает в себе полнофункциональную интегрированную среду разработки с поддержкой языка OPL и высокопроизводительные модули решений CPLEX и CP Optimizer.
- SCIP - один из самых быстрых некоммерческих решателей для смешанного целочисленного программирования (MIP) и смешанного целочисленного нелинейного программирования (MINLP).
- SciPy – содержит эффективный решатель MILP задач HiGHS - High-performance parallel linear optimization software

Многокритериальная оценка ЛА

Критерии:

| Технические характеристики | Экономичность | Импортозависимость |
|----------------------------|---------------|--------------------|
| Хорошие 1 | Хорошая 1 | Высокая 1 |
| Удовл. 0.5 | Удовл. 0.5 | Низкая 0 |
| Неудовл. 0 | Неудовл. 0 | |

Альтернативы:

АН-124



ИЛ-76

| Тех. характеристики | Экономичность | Импортозависимость |
|---------------------|---------------|--------------------|
| 0.25 | 0.8 | 1 |

| Тех. характеристики | Экономичность | Импортозависимость |
|---------------------|---------------|--------------------|
| 0.5 | 0.35 | 0 |

Области предпочтения:

| Уровень предпочтения | Низкий | Средний | Высокий |
|----------------------|--------------------|--|--|
| | Значения критериев | Тех. характеристики: 0.5, 0 Экономичность: 0.5, 0 Импорт.: 1 | Тех. характеристики: 0.5 Экономичность: 1, 0.5 Иморт.: 1 |

Ранги:

Ранг АН-124: 0.8256

Ранг ИЛ-76: 0.3484

Использование областей предпочтений

Области предпочтения

Удалить Уровень предпочтения Неуд 2.0

| | | | |
|----------------------------|----------------|---------------|-------------|
| Технические характеристики | Хорошие 0.5 | Плохие 0.3 | Неуд 0.0 |
| Экономичность | Уд 0.3 | Неуд 0.0 | Хор 0.5 |

Удалить Уровень предпочтения Уд 3.0

| | | | |
|----------------------------|----------------|---------------|-------------|
| Технические характеристики | Хорошие 0.5 | Плохие 0.3 | Неуд 0.0 |
| Экономичность | Уд 0.3 | Неуд 0.0 | Хор 0.5 |

Удалить Уровень предпочтения Хорошо 4.0

| | | | |
|----------------------------|----------------|---------------|-------------|
| Технические характеристики | Хорошие 0.5 | Плохие 0.3 | Неуд 0.0 |
| Экономичность | Уд 0.3 | Неуд 0.0 | Хор 0.5 |

Удалить Уровень предпочтения Неуд 2.0

| | | | |
|----------------------------|----------------|---------------|-------------|
| Технические характеристики | Хорошие 0.5 | Плохие 0.3 | Неуд 0.0 |
| Экономичность | Уд 0.3 | Неуд 0.0 | Хор 0.5 |

[Добавить область предпочтения](#)

[Сохранить](#)

Оптимальное размещение грузов на борту воздушных судов

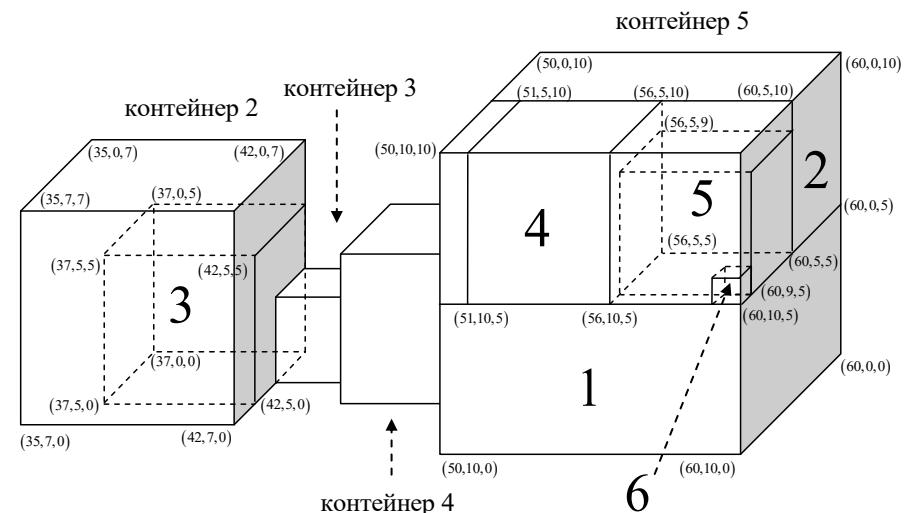
$$\max \alpha \left(\sum_{i=1}^n x_i^1 \right) + \beta \left(\sum_{i=1}^n \sum_{j=2}^g m_i r_{ij} \right)$$

$$(-1)^e x_v^{\left\lfloor \frac{e}{2} \right\rfloor} + (-1)^{e+1} x_i^{\left\lfloor \frac{e}{2} \right\rfloor} + \sum_{d=1}^D (-1)^{f(e,d)+1} M a_{vi}^d$$

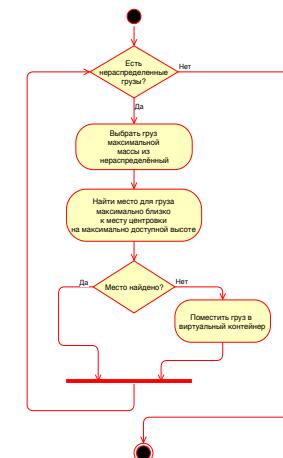
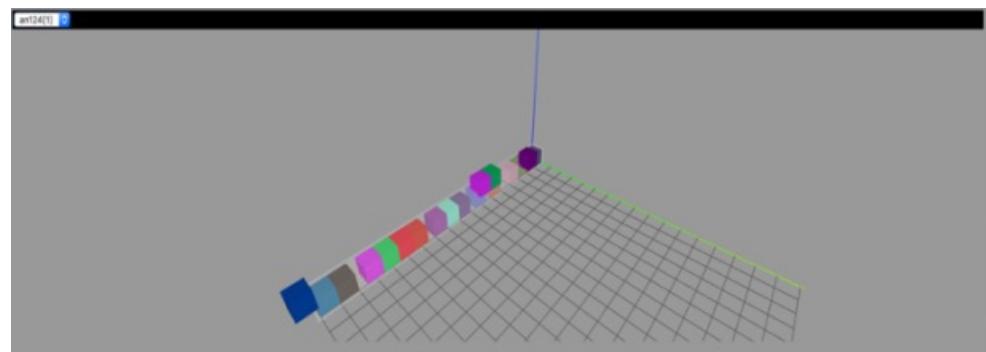
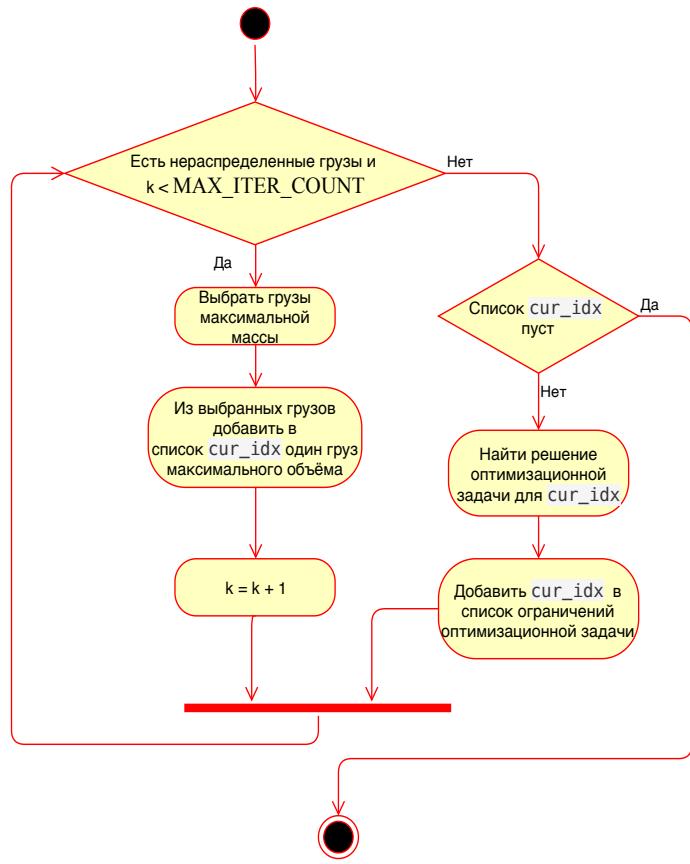
$$\leq -l_{u(e,v,i)}^{\left\lfloor \frac{e}{2} \right\rfloor} + \sum_{d=1}^D f(e, d) M$$

$$\sum_{v=1}^n (m_v q_{vij}^d) - m_i p_{ij}^d - \frac{1}{2} m_i l_i^d r_{ij} = 0$$

$$\sum_{i=1}^n \prod_{d=1}^D l_i^d r_{ij} \leq \prod_{d=1}^D L_j^d .$$



Алгоритм декомпозиции, эвристика, мультиагентный подход



Формирование расписаний авиаперевозок

$$\min_{x_{ijl}^k} \left(\sum_k \sum_i \sum_{j \in J_i^k} \alpha^k x_{ij1}^k + \sum_k \sum_i \sum_{j \in J_i^k} \sum_l \beta_{ij}^k x_{ijl}^k \right)$$

$$\forall k \forall l \sum_i \sum_{j \in J_i^k} x_{ijl}^k \leq 1, \forall k \forall l < L_k \forall j \sum_{i \in I_j^k} x_{ijl}^k \leq \sum_{u \in J_j^k} x_{ju(l+1)}^k + \sum_{u \in J_j^k} x_{ju1}^k$$

$$\forall k \sum_i \sum_{j \in J_i^k} \sum_l t_{ij}^k x_{ijl}^k \leq T - C_k, \forall i \forall j \sum_k \sum_l v_k x_{ijl}^k \geq a_{ij},$$

$$\forall i \sum_k \sum_{j \in J_i^k} \sum_l x_{ijl}^k + \sum_k \sum_{j \in I_i^k} \sum_l x_{jil}^k \leq S_i$$

Решение на портале WS-DSS

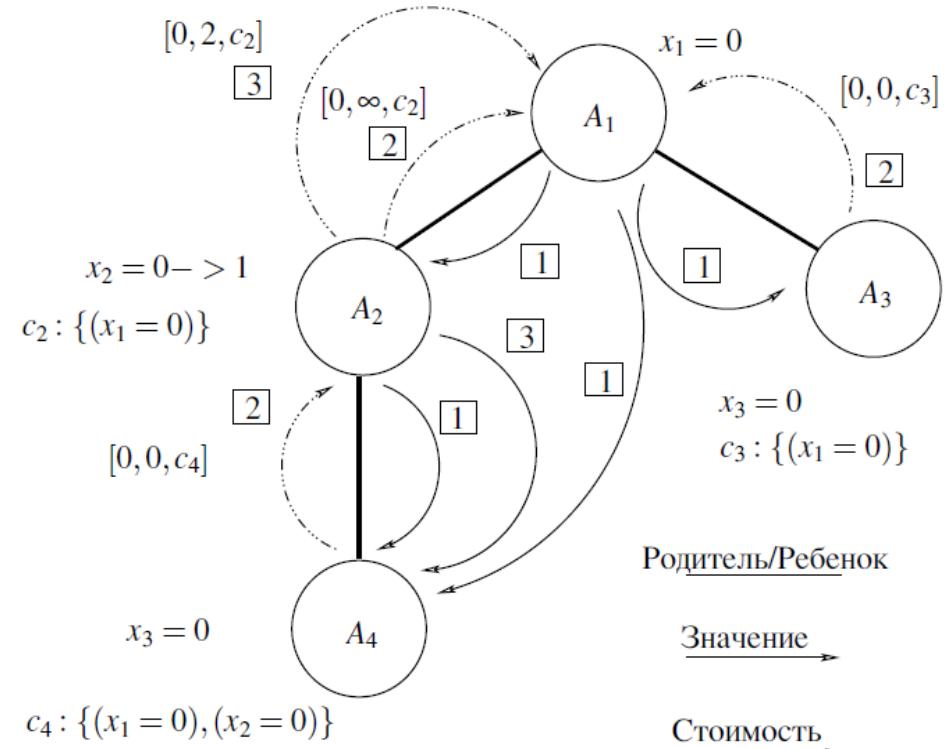
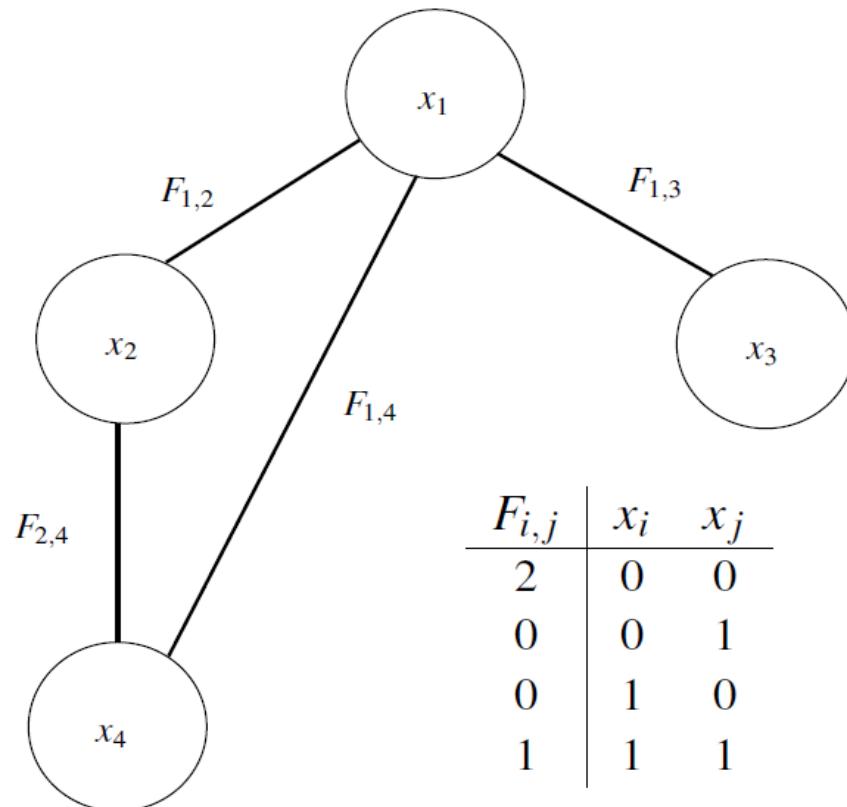


The screenshot shows a web browser window with the URL `ws-dss.com`. The main content area displays a large JSON object. Below the browser, there is a text input field labeled "Выходные данные:" containing a smaller JSON object.

```
ws-dss.com
{
    "GDG": {"MRV": 20000, "CSY": 10000},
    "MRV": {"GDG": 19000, "CSY": 5000},
    "CSY": {"MRV": 7000, "GDG": 11000}
},
"available": {
    "GDG": {"MRV": ["mc21", "sj100", "il96", "tu204"], "CSY": ["mc21", "sj100", "tu204"]},
    "MRV": {"GDG": ["mc21", "sj100", "il96", "tu204"], "CSY": ["mc21", "sj100", "tu204"]},
    "CSY": {"MRV": ["mc21", "sj100", "tu204"], "GDG": ["mc21", "sj100", "tu204"]}
},
"people": {
    "GDG": {"MRV": 250, "CSY": 30},
    "MRV": {"GDG": 200, "CSY": 300},
    "CSY": {"MRV": 200, "GDG": 100}
},
"T": 60,
"trace": 0,
"version": 1
}

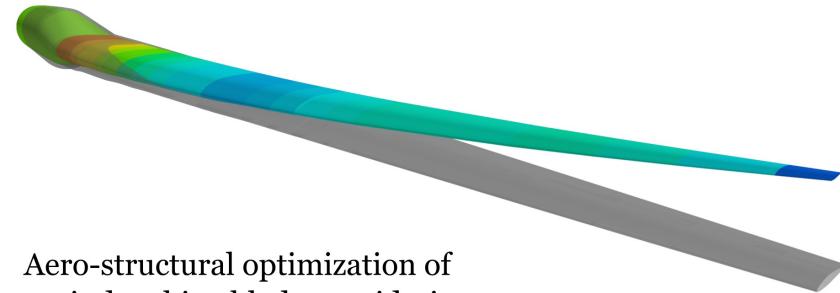
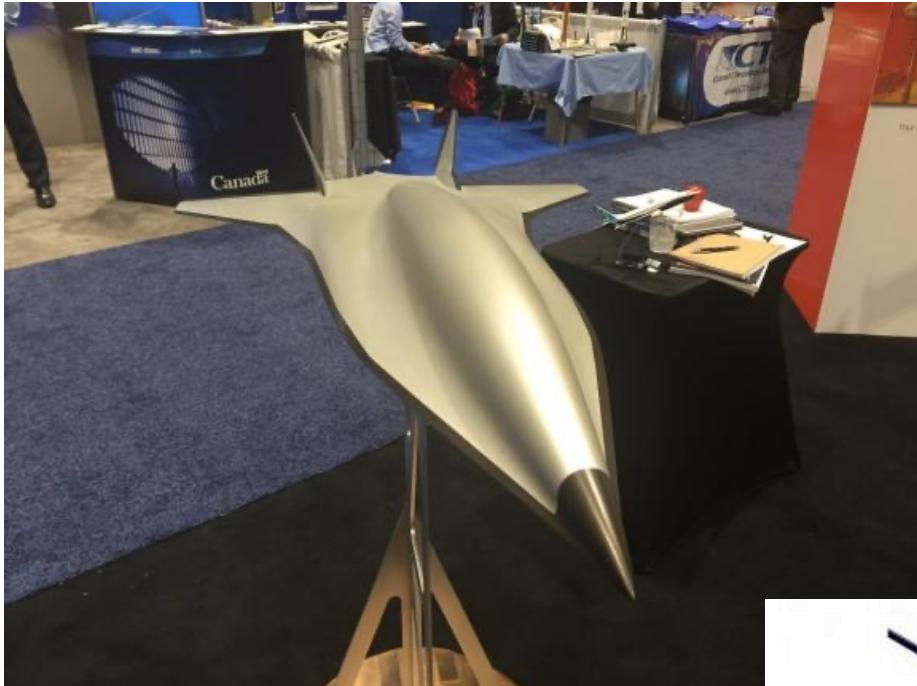
Выходные данные:
{
    "aircraft": {
        "il96": [
            {
                "aircraft_num": 0,
                "airport": "GDG"
            }
        ],
        "tu204": [
            {
                "aircraft_num": 1,
                "airport": "MRV"
            },
            {
                "aircraft_num": 2,
                "airport": "GDG"
            }
        ]
    }
}
```

Мультиагентное моделирование в задачах DCOP

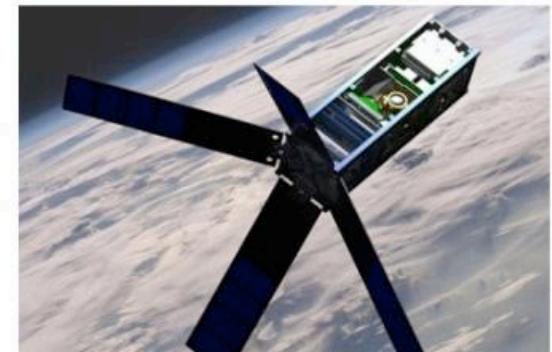
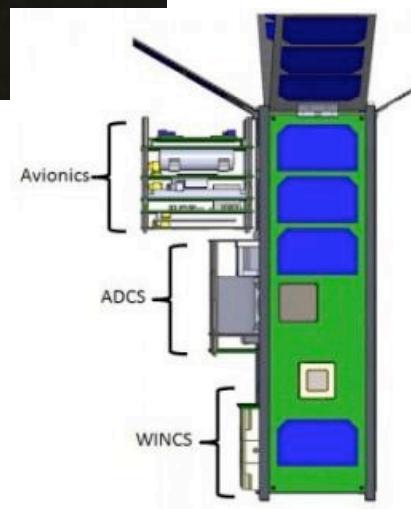


Пакет pyDCOP

Мультидисциплинарная оптимизация



Aero-structural optimization of
a wind turbine blade considering
pre-curve jig shape



Пакет OpenMDAO

Производственная задача

Рассматривается некоторая производственная система, способная производить несколько видов продукции. Для производства используется ряд сырьевых ресурсов, имеющихся в системе в ограниченном количестве. От реализации произведенной продукции система получает прибыль. Требуется так составить производственный план (определить, какие виды продукции и в каком количестве производить), чтобы при имеющихся ограничениях на сырьевые ресурсы получить максимальную прибыль.

Формализованная постановка

n - количество видов продукции, которую может производить система;

m - количество видов сырья, используемого при производстве продукции;

$c[j]$, ($j = 1, 2, \dots, n$) - прибыль, получаемая от реализации произведенной единицы j -ого вида продукции;

$b[i]$, ($i = \overline{1, m}$) - количество имеющегося в наличии сырья i -ого вида;

$a[i, j]$, ($i = \overline{1, m}; j = \overline{1, n}$) - технологический коэффициент затрат i -ого вида сырья на производство единицы продукции j -ого вида;

$x[j]$, ($j = \overline{1, n}$) - планируемого количества производимой продукции j -ого вида (оптимизационная переменная).

С учетом введенных обозначений формализованная запись имеет вид:

$$\max_x z = \sum_{j=1, n} c[j] x[j],$$

$$\sum_{j=1}^n a[i, j] x[j] \leq b[i], \quad (i = \overline{1, m}), \quad (1.1)$$

$$x[j] \geq 0, \quad (j = \overline{1, n}).$$

Здесь z - суммарная прибыль от реализации произведенной продукции, $\sum_{j=1}^n a[i, j] x[j]$ - затраты i -ого вида сырьевого ресурса на реализацию всего производственного плана.

Задача о поднятии плит

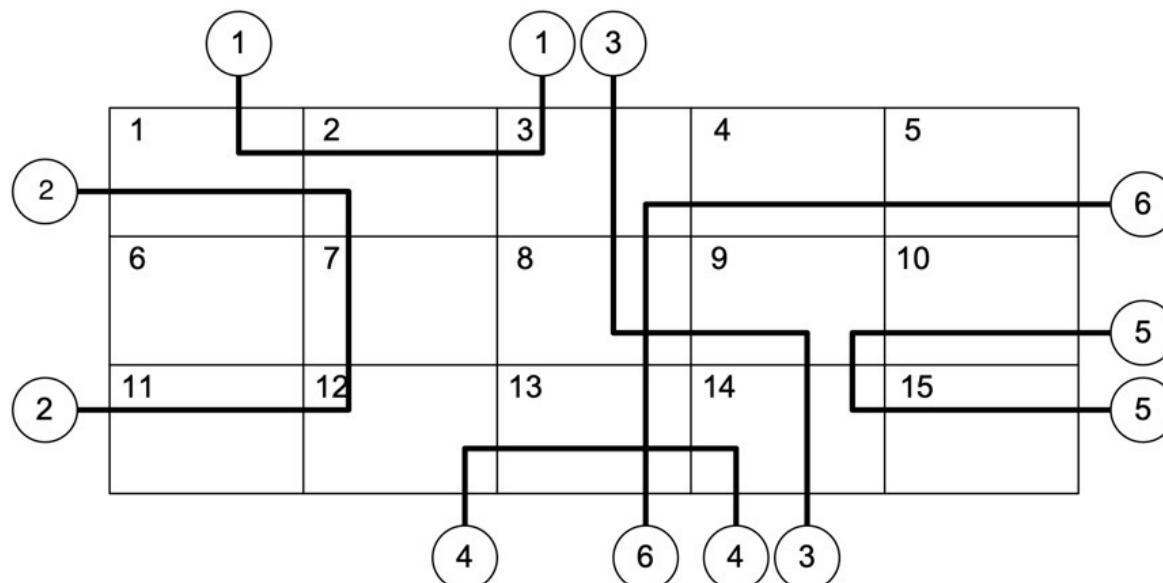
Необходимо проверить состояние кабельных линий, которые находятся под полом, состоящим из плит. Для проверки каждого кабеля достаточно получить к нему доступ в любом месте, для чего нужно поднять соответствующую плиту. На поле расположено оборудование. Поэтому с поднятием каждой плиты пола связан определенный объем работ по демонтажу и перемещению оборудования, задаваемый в человеко-часах. Необходимо определить плиты, которые нужно поднять таким образом, чтобы обеспечить доступ ко всем кабелям, а суммарный объем работ, связанный с поднятием плит фальшпола, был бы минимальным.

Вход: стоимость поднятия плит, какие кабели под какими плитами лежат.

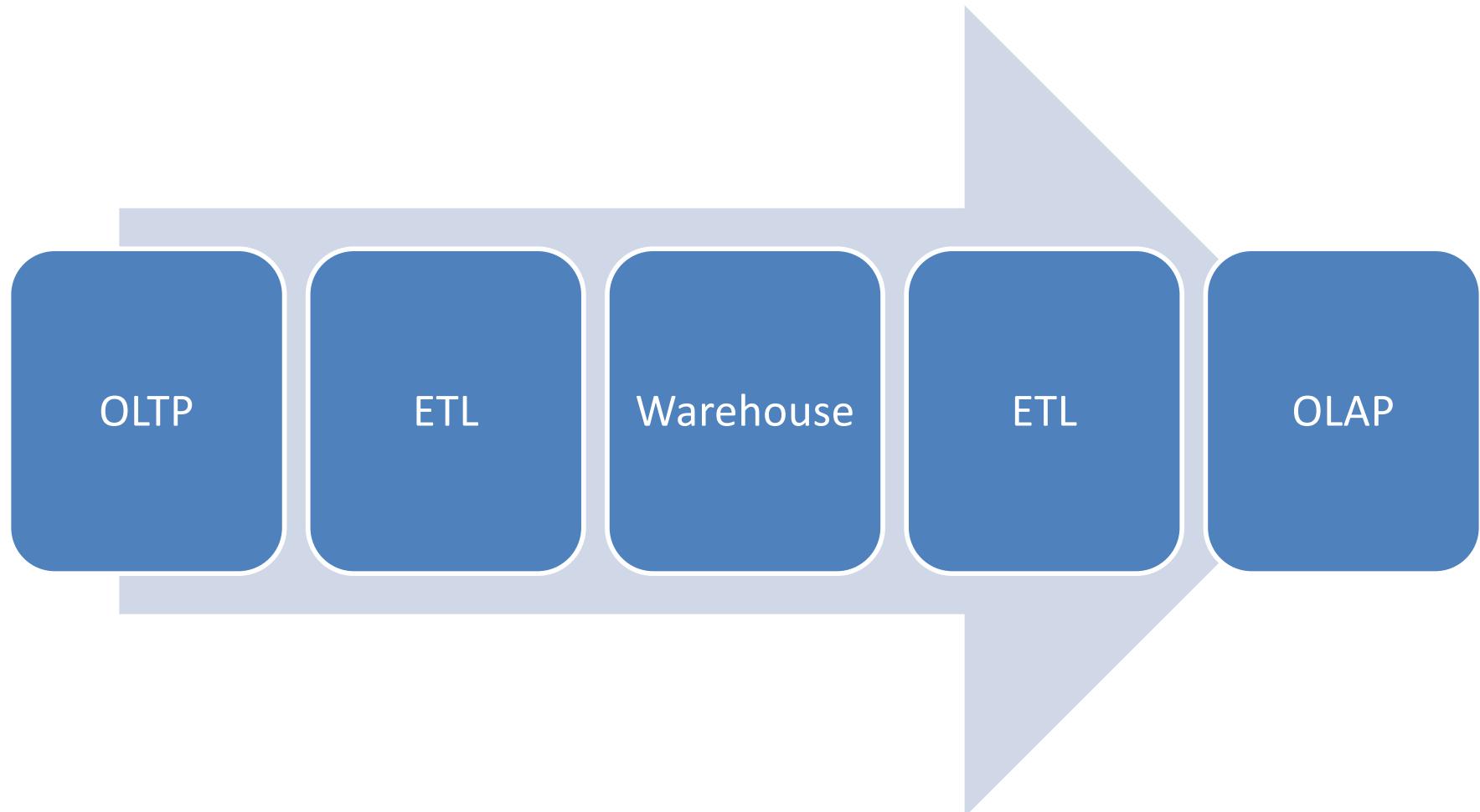
Выход: Какие плиты поднимаем.

Программа должна работать с произвольными корректными данными.

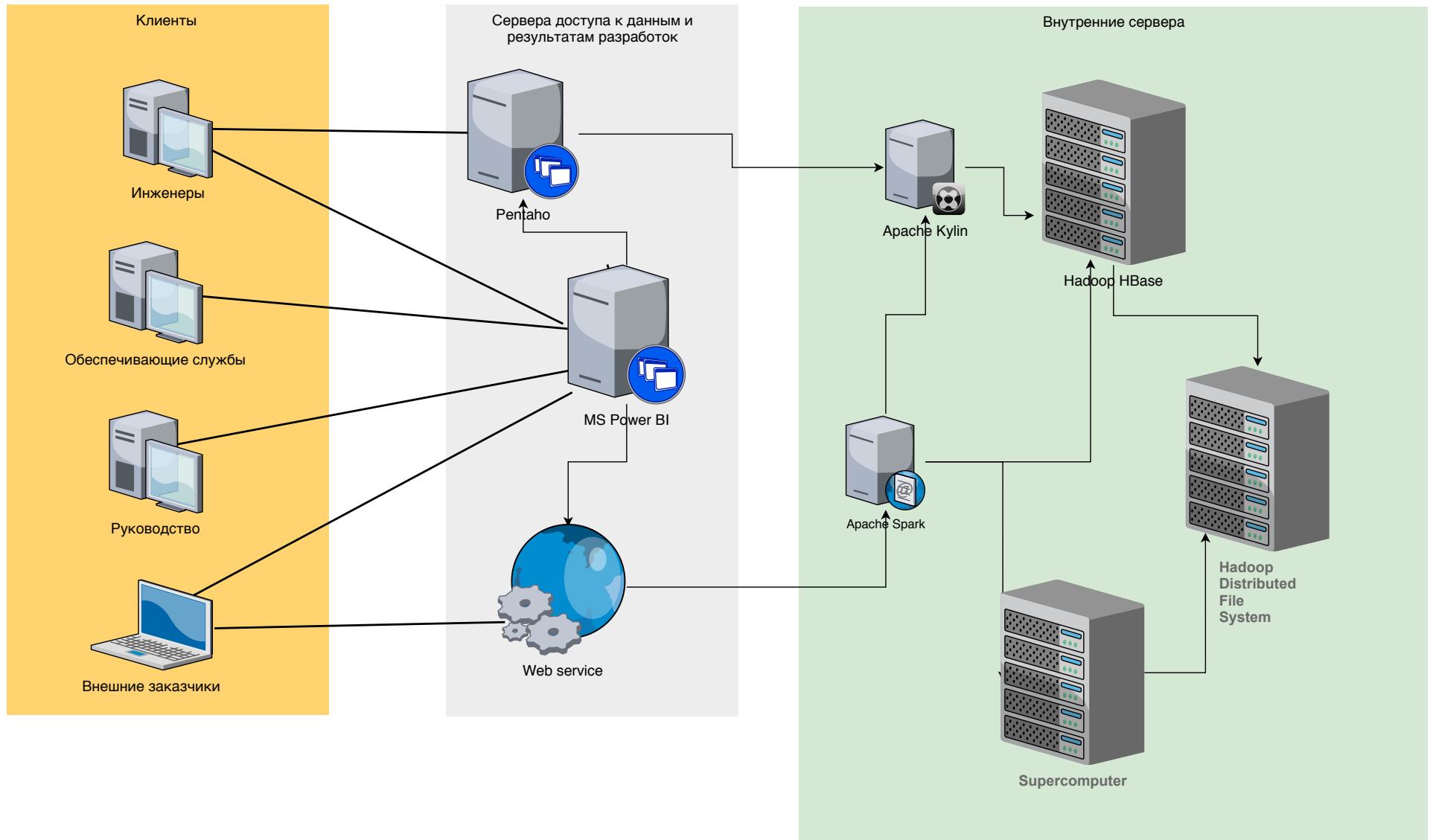
Пример исходных данных:



Типовой процесс обработки информации



Архитектура аналитической платформы



OLAP. Определение

OLAP – это технология обработки данных, заключающаяся в подготовке агрегированной информации на основе больших массивов данных, структурированных по многомерному принципу.

Реализации технологии OLAP являются компонентами программных решений класса Business Intelligence (BI)

FASMI тест

- Fast
- Analysis
- Shared
- Multidimensional
- Information

Куб

- **OLAP-куб** — многомерный массив данных, как правило, разрежённый и долговременно хранимый. Может быть реализован на основе универсальных реляционных СУБД или специализированным программным обеспечением.
- Индексам массива соответствуют измерения (dimensions) или оси куба, а значениям элементов массива — меры (measures) куба:

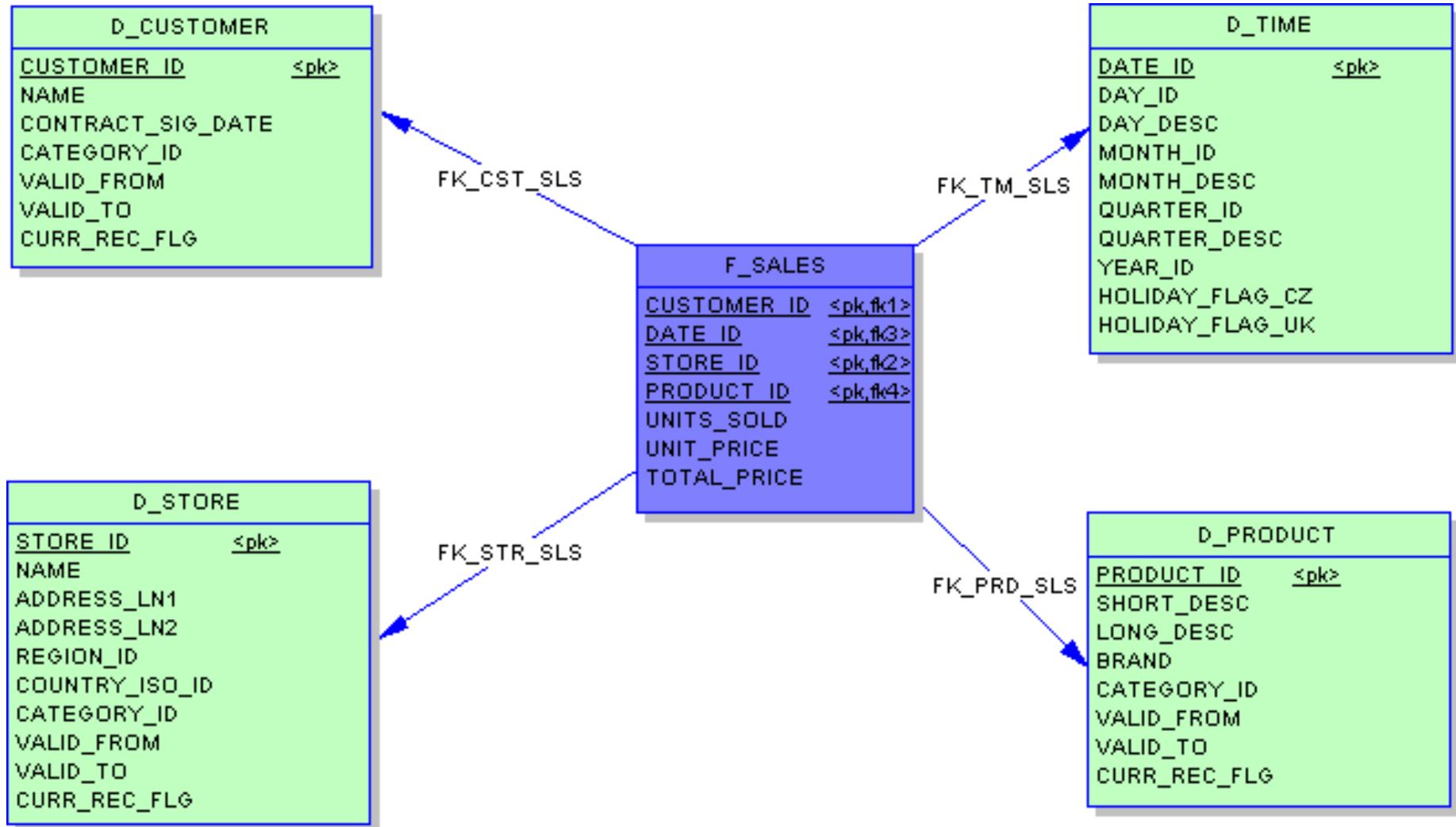
$$w : (x,y,z) \rightarrow w_{xyz},$$

где x, y, z — измерения, w — мера.

- В отличие от обычного массива в языке программирования, доступ к элементам- OLAP-куба может осуществляться как по полному набору индексов-измерений, так и по их подмножеству, и тогда результатом будет не один элемент, а их множество.

$$W : (x,y) \rightarrow W = \{w_{z1}, w_{z2}, \dots, w_{zn}\}$$

Схема звезды



Описание многомерного пространства

- Измерение (dimension) – описывает элемент данных по которому производится анализ. Например, Время.
- Элемент (Member) — соответствует одной точке на измерении. Например, в измерении Время (Time) понедельник будет элементом измерения;
- Атрибут (Attribute) — это полная коллекция элементов Member одного типа. Например, все дни недели будут атрибутом измерения Время (Time);
- Размер (Size), или Кардинальность (Cardinality) измерения — это количество элементов, которое оно содержит. Например, измерение Время (Time), состоящее из дней недели, будет иметь размер 7.

Определения многомерного пространства

- Кортеж (Tuple) — это координата в многомерном пространстве; Кортеж — это набор членов одного или нескольких разных измерений. Задавая кортеж, мы указываем на конкретную ячейку или набор ячеек внутри куба. Таким образом, кортеж — это декартово произведение (т.е. пересечение) множеств, определенных на различных измерениях куба. Tuple is an ordered collection of one or more members from different dimensions. Tuples can be specified enumerating the members, e.g.
(
 [Time].[Fiscal].[Month].[August],
 [Customer].[By Geography].[All Customers].[USA],
 [Measures].[Sales]
)
- Срез (Slice) — это секция многомерного пространства, которая может быть определена кортежем. Срез (Slice) — это созданное пользователем подмножество гиперкуба, получившееся в результате фиксации значения одного или более измерений не входящих в это подмножество.
(*, *, [Май])

MDX

SELECT в MDX имеет три основных предложения: SELECT, FROM и WHERE:

- Предложение SELECT определяет многомерное пространство, которое будет результатом запроса.
- Предложение FROM определяет источник данных, которым может быть либо название куба, содержащего данные, либо другой запрос.
- Предложение WHERE определяет правила, ограничивающие результаты запроса подпространством данных. Процесс ограничения результатов называется сечением (slicing). В Analysis Services 2005 сечение может выполняться не только по одной плоскости, но и по более сложным фигурам.
Предложение WHERE опционально и может быть опущено.

MDX

select <содержимое оси> on axis(0),

< содержимое оси> on axis(1),

< содержимое оси> on axis (2),

< содержимое оси> on axis (n)

from <имя_куба>

MDX

SELECT

{([Drink],[1997]),([Drink],[1998]),([Food],[1997]),([Food],[1998])} ON COLUMNS,

{[Measures].[Store Sales],[Measures].[Store Cost],[Measures].[Unit Sales]} ON ROWS

FROM [Warehouse and Sales] WHERE
[Store].[Stores].[Store Country].[Mexico]

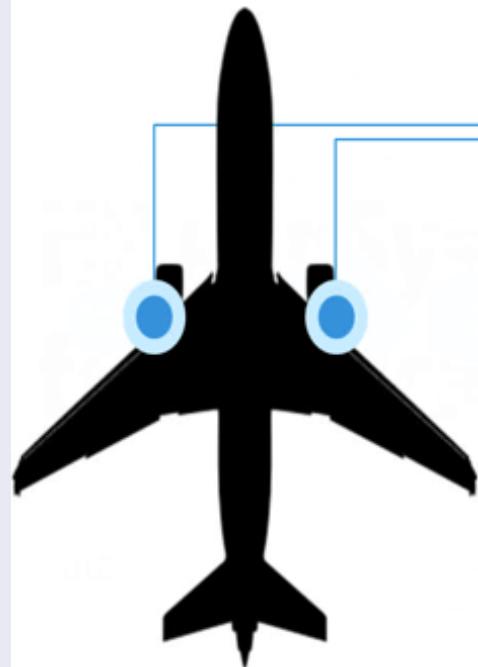
Определение

Большие данные (англ. big data) — обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами

Пример больших данных

Data in USA cross-country commercial flights

Sensor data from a cross-country flight



20 TB × 2 × 6 × 28,537 × 365

20 terabytes of
information per
engine every hour

twin-engine
Boeing 737

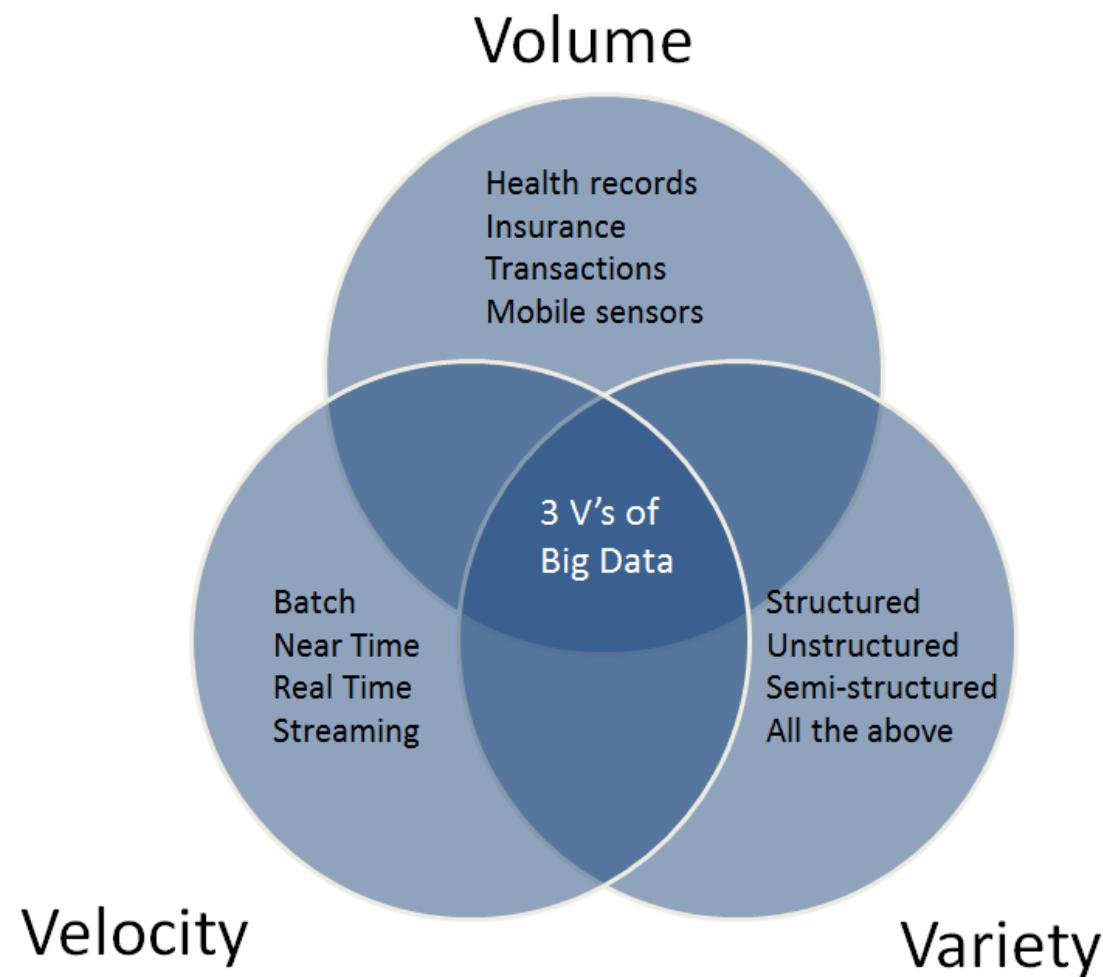
six-hour, cross-
country flight from
New York to Los
Angeles

of commercial
flights in the sky in
the United States on
any given day.

days in a year

= 2,499,841,200 TB

Признаки больших данных



Теорема CAP

— в любой реализации распределённых вычислений возможно обеспечить не более двух из трёх следующих свойств:

- согласованность данных (англ. consistency) — во всех вычислительных узлах в один момент времени данные не противоречат друг другу
- доступность (англ. availability) — любой запрос к распределённой системе завершается корректным откликом, однако без гарантии, что ответы всех узлов системы совпадают
- устойчивость к разделению (англ. partition tolerance) — расщепление распределённой системы на несколько изолированных секций не приводит к некорректности отклика от каждой из секций

NoSQL

NoSQL (от англ. not only SQL — не только SQL) — обозначение широкого класса разнородных систем управления базами данных, существенно отличающихся от реляционных СУБД с доступом средствами SQL

Применяется к системам, в которых делается попытка решить проблемы масштабируемости и доступности за счёт полного или частичного отказа от требований атомарности и согласованности данных

Типы NoSQL

СУБД ключ — значение

используются для хранения изображений, создания специализированных файловых систем, в качестве кэшей для объектов, а также в системах, спроектированных с прицелом на масштабируемость.

Примеры: Berkeley DB, MemcacheDB, Redis, Riak, Amazon DynamoDB

СУБД семейство столбцов

в таких системах данные хранятся в виде разреженной матрицы, строки и столбцы которой используются как ключи. Типичным применением этого типа СУБД является веб-индексирование, а также задачи, связанные с большими данными, с пониженными требованиями к согласованности

Примеры: Google BigTable, Apache HBase, Apache Cassandra, ScyllaDB

Документно-ориентированная СУБД

служат для хранения иерархических структур данных

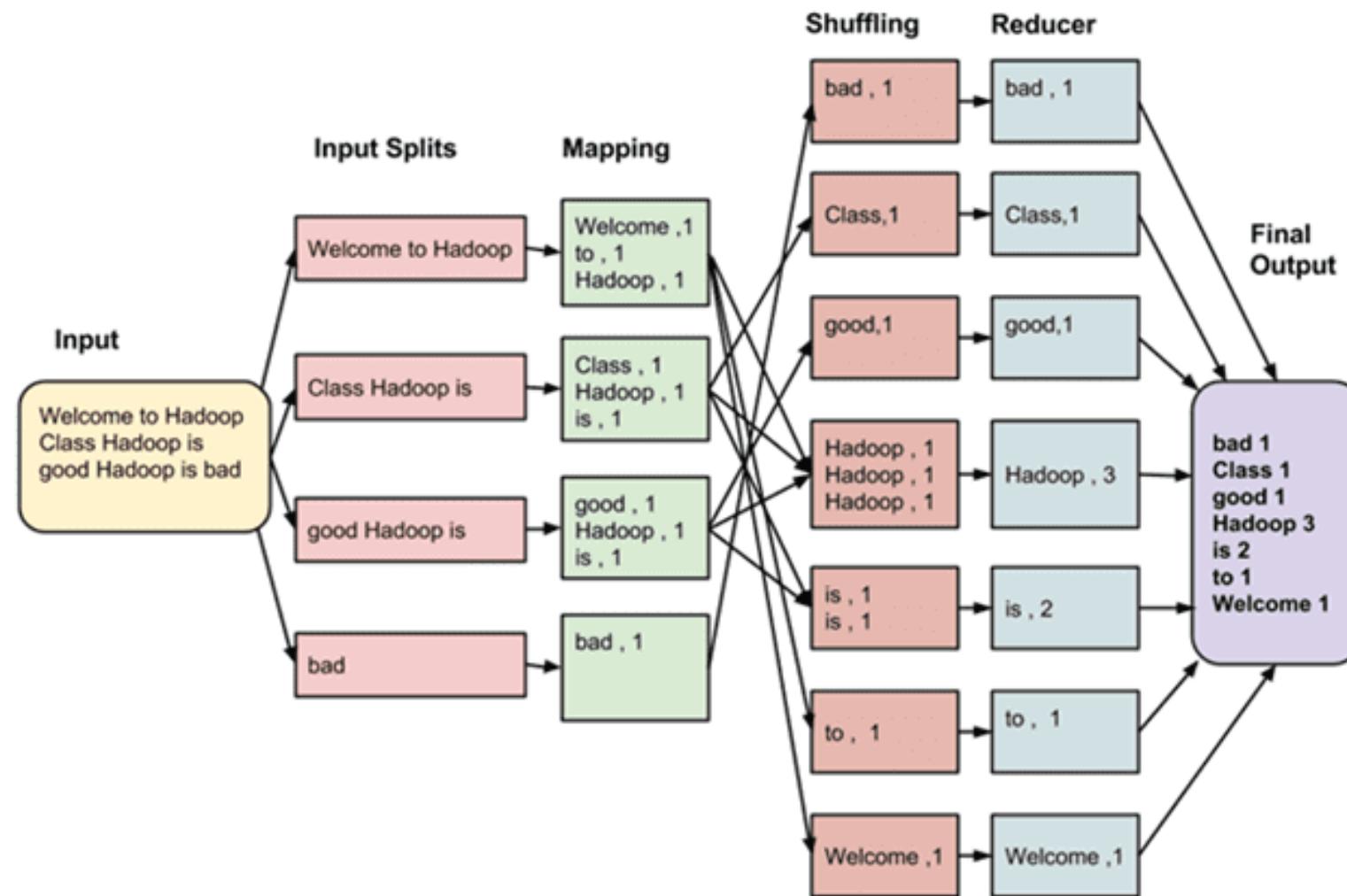
Примеры: CouchDB, Couchbase, MongoDB

Графовая СУБД

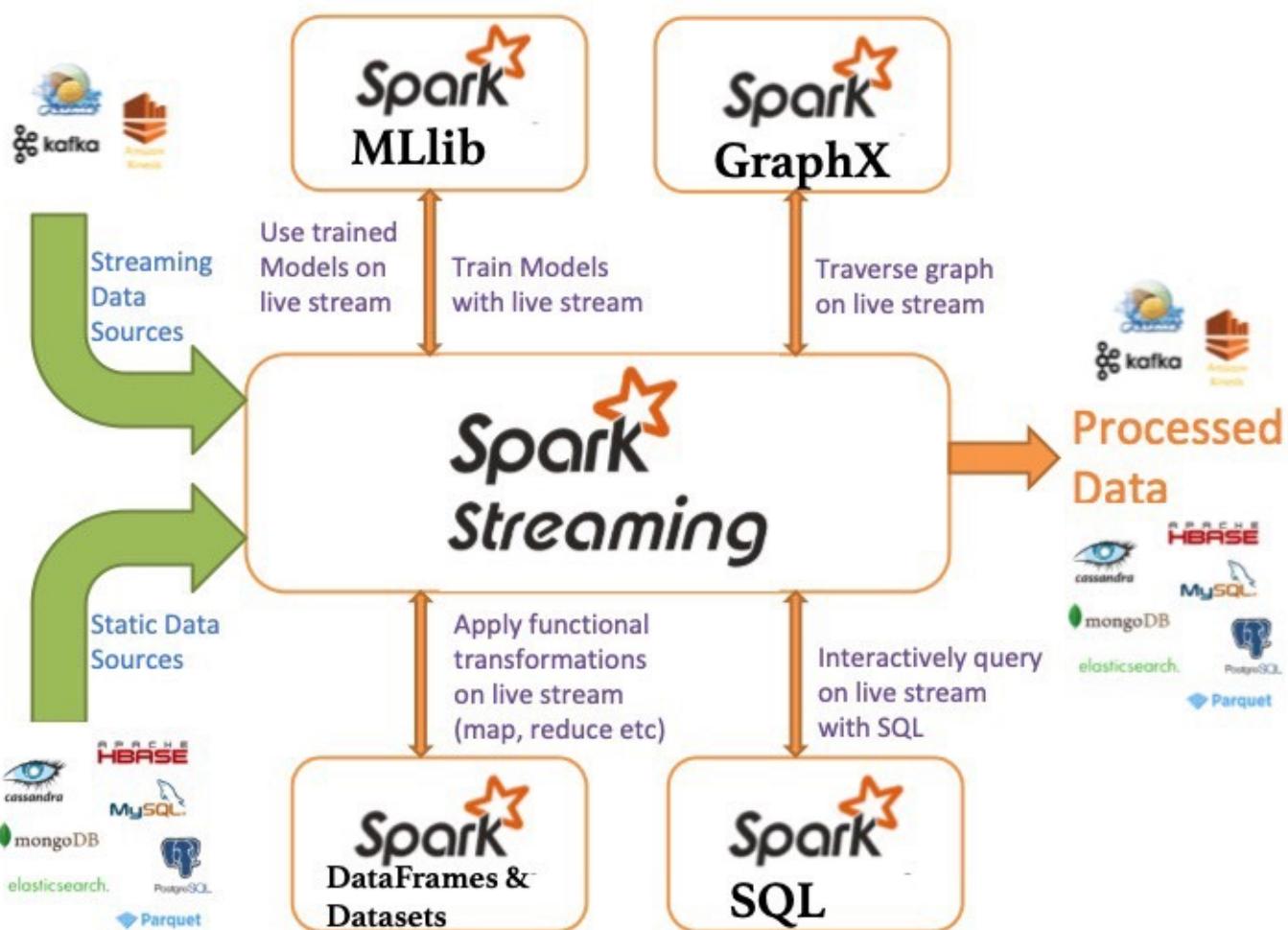
применяются для задач, в которых данные имеют большое количество связей, например, социальные сети, выявление мошенничества

Примеры: Neo4j, OrientDB, AllegroGraph, Blazegraph

Map Reduce



Apache Spark

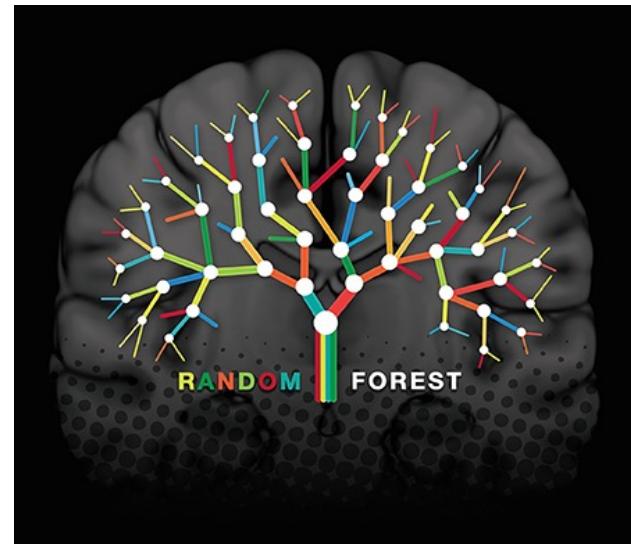


Основные тенденции

- Ансамблевые методы
- Гиперпараметрическая оптимизация
- AutoML
- RL



Yandex
CatBoost



XGBoost



LightGBM

Итоги

- Определите цели
- Измеряйте свое приближение к целям
- Ведите разведочный анализ, готовьте данные
- Выбирайте модели
- Обучайте модели
- Рассчитывайте метрики
- Внедряйте хорошие модели в бизнес-процессы
- Постоянно повышайте качество

Пишите: sudakov@ws-dss.com

Спасибо за внимание!