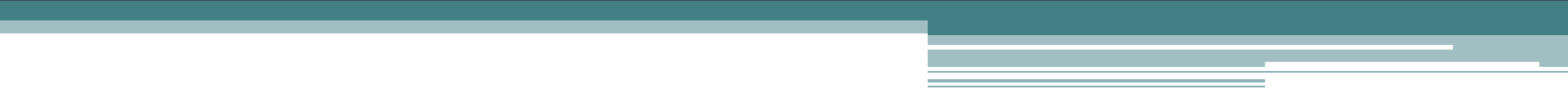


# Интеллектуальная система анализа геолого-промысловых данных

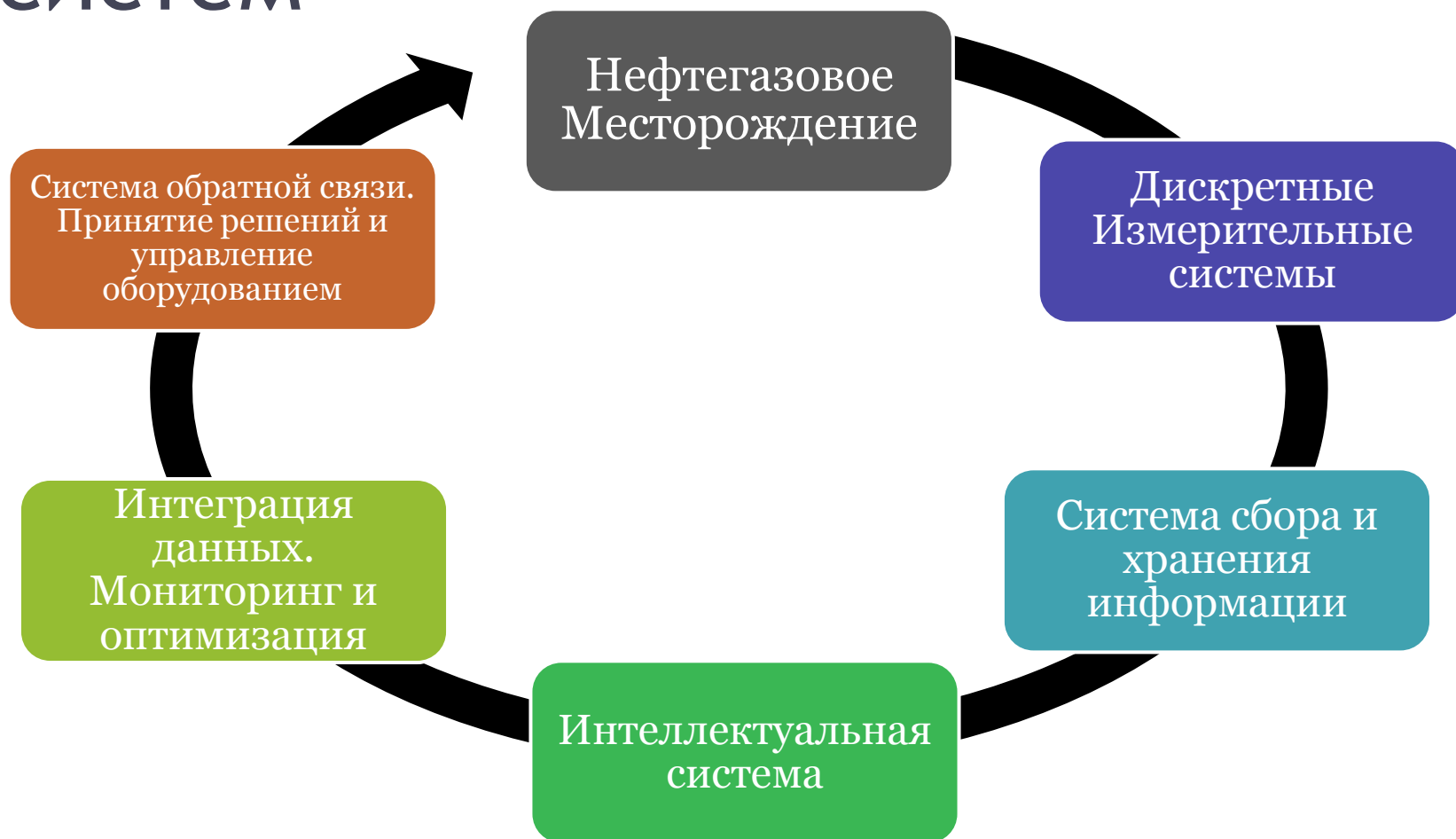


Докладчик: Татарников В.В.  
Руководитель: Загоруйко Н. Г.

# Проблемы индустрии

- Падение добычи и рост издержек
- Усложнение географических условий
- Высокая степень неопределённости данных

# Этапы развития информационных систем

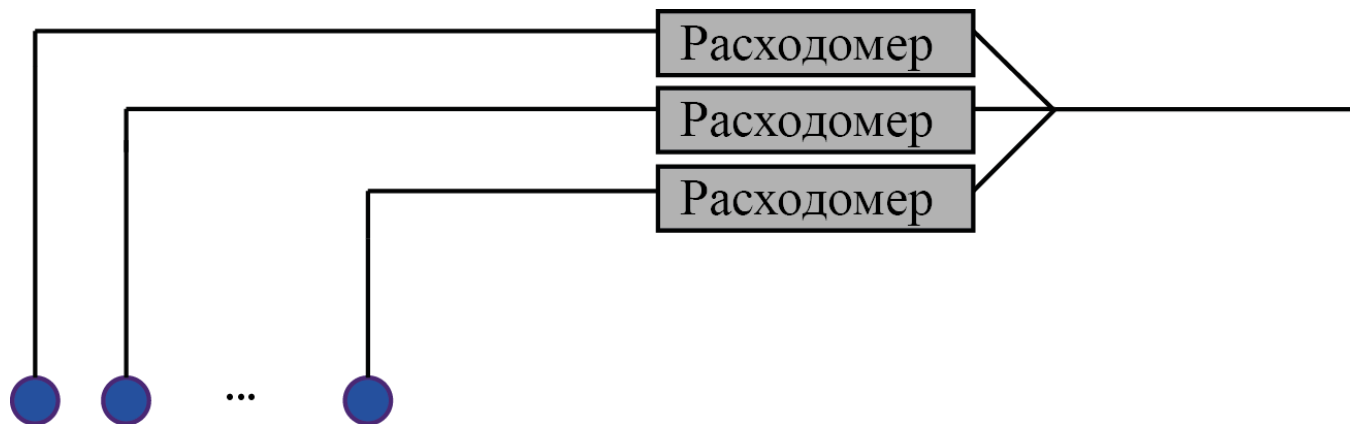


# Цель работы

Разработка интеллектуальной системы,  
обеспечивающей снижение степени  
неопределённости геолого-промысловых  
данных

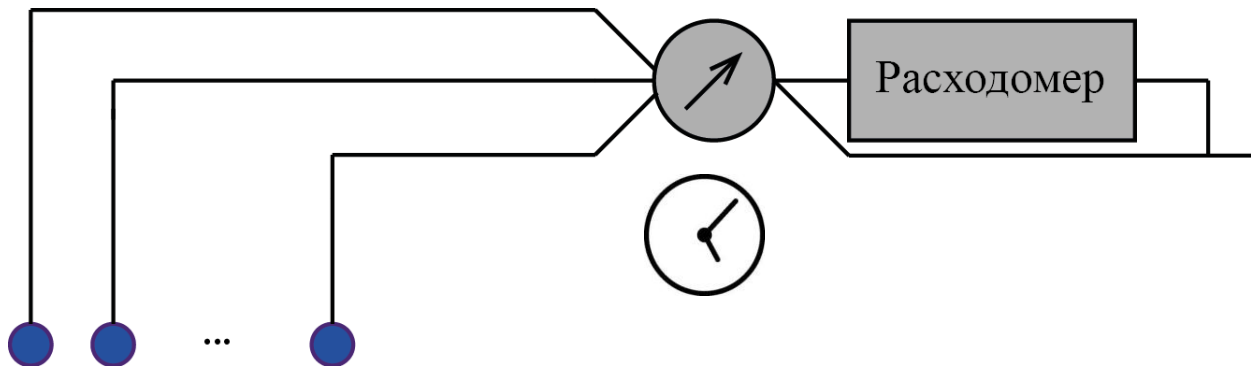
# Неопределённость данных?

- Установить полный комплекс измерительных приборов на каждую скважину



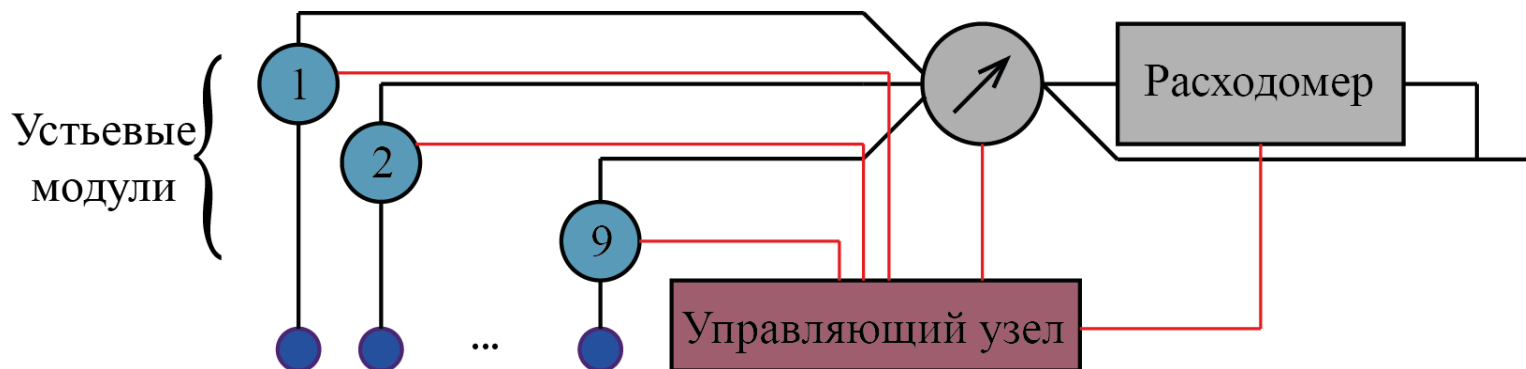
# Неопределённость данных

- Около 70% скважин в РФ низкодебитные
- Используется переключение по расписанию
- Дебит по скважинам расписывается вручную



# Неопределённость данных

- Переключение по событиям
- Снижает потери информации
- Показания устьевых модулей содержат ошибки и пробелы

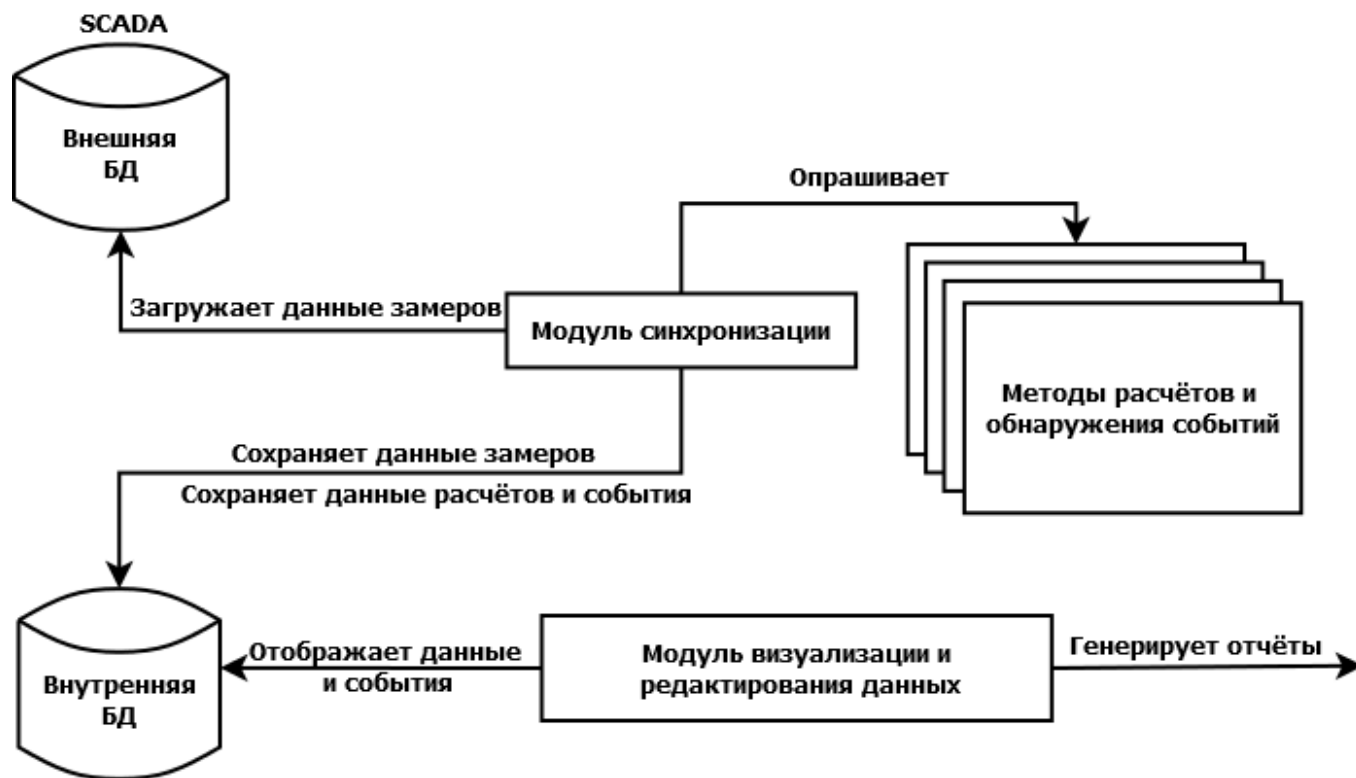


# Задачи

- Обеспечить сбор и хранение, редактирование и визуализацию данных, создание отчётов
- Организовать обработку рассчитываемых параметров (выделение событий)
- Обеспечить предобработку данных измерений скважин (заполнение пробелов, исправление ошибок)

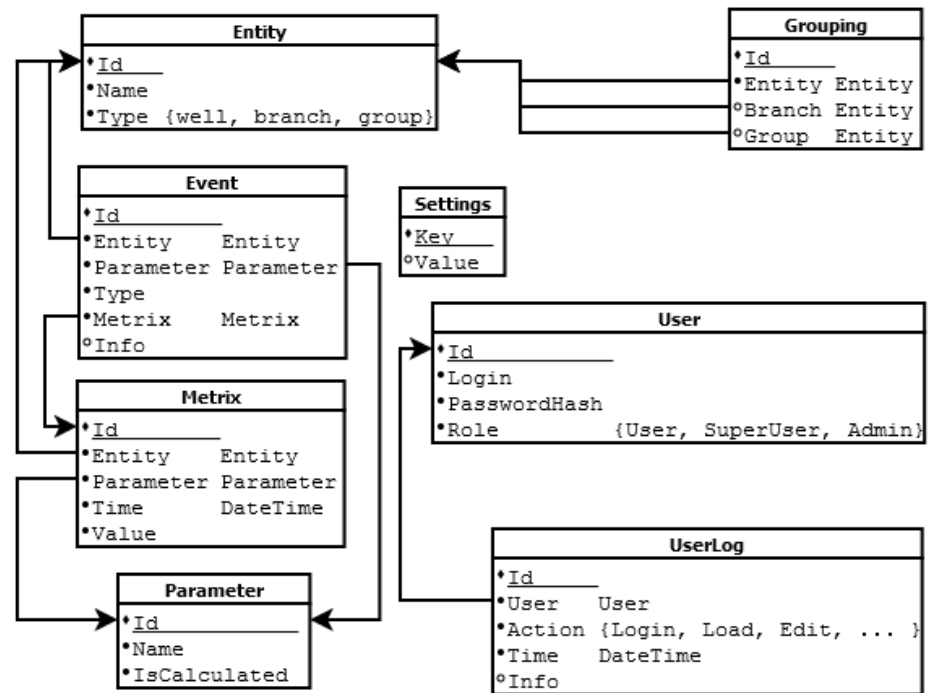


# Модульная организация ИС



# Данные

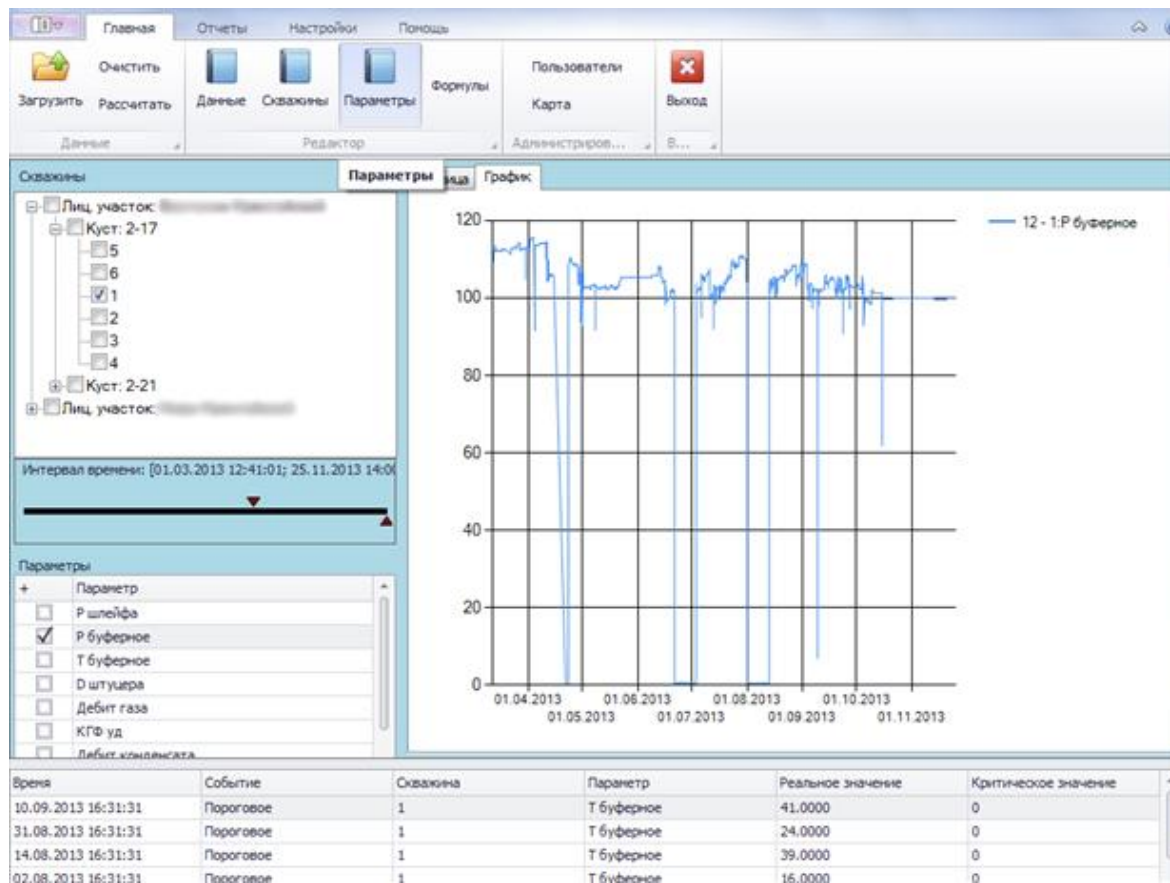
- Схема:
  - Скважины
  - Параметры
  - Измерения
  - События
  - Пользователи
- Импорт данных:
  - Автоматический
  - Ручной



# Редактирование и визуализация

- Работает с внутренней БД
- Ручная загрузка данных
- Ручное редактирование данных, скважин и параметров
- Визуализация данных
- Отображение событий
- Формирование отчётов

# Редактирование и визуализация



# Методы выделения событий

- Реализованы в виде подключаемых модулей

```
public interface EventChecker
```

```
{
```

```
    string Name { get; }
```

```
    string Help { get; }
```

```
    string Version { get; }
```

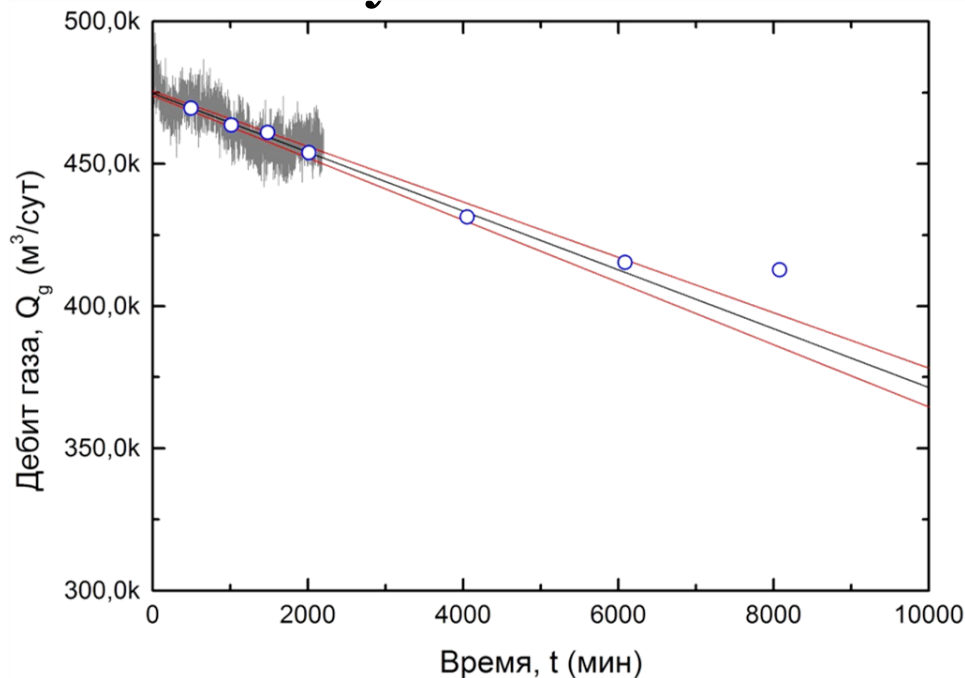
```
    ...
```

```
    Event[] findEvents(DateTime d, DataHelper helper);
```

```
}
```

# Методы выделения событий

- Скачок значения выше заданного порога
- Обнаружена ошибка или пробел
- Доверительный конус

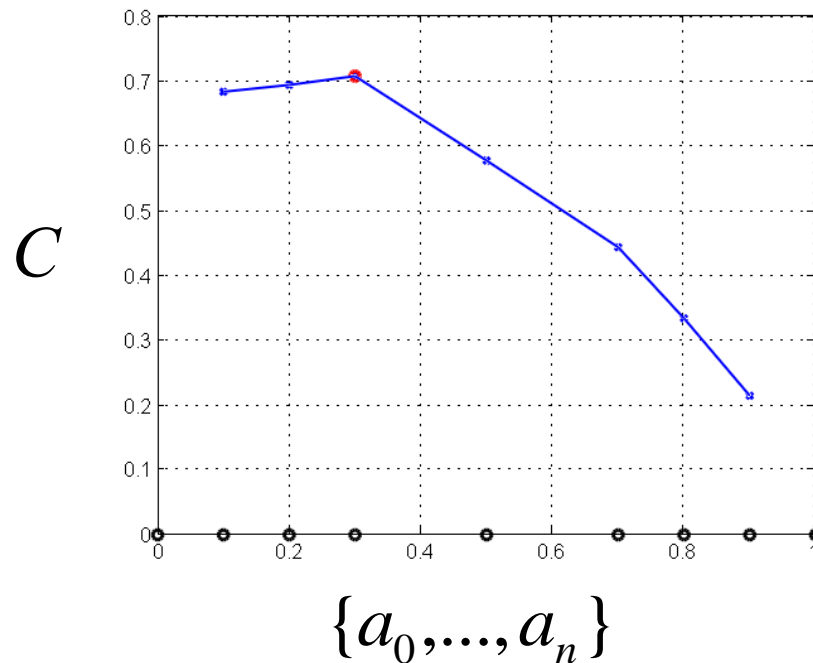


# Заполнение пробелов

- Обзор существующих методов
  - Простые
  - Сложные
- Разработка собственного метода
  - Вариант алгоритма ZET для кубов данных
  - FRiS-функция
  - Количественная оценка компактности подкуба
- Критерий ошибки и реализация модуля для системы

# Заполнение пробелов

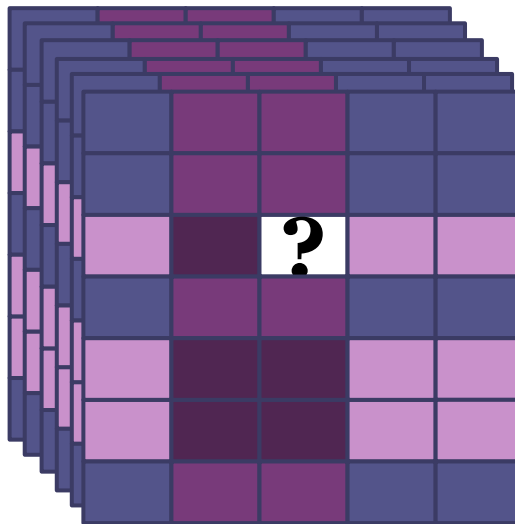
1. Алгоритм ищет наиболее компактный подкуб данных





# Заполнение пробелов

## 2. Заполняет пробел на основании модели



- Модели работают лучше в локальной окрестности исследуемого объекта

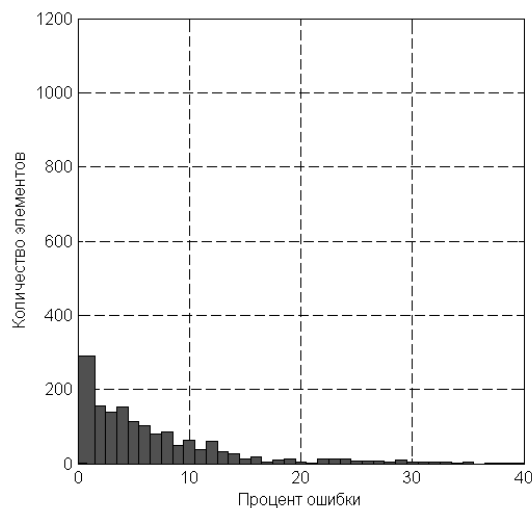
# Заполнение пробелов

- Эксперимент с перекрёстной проверкой при помощи многомерной регрессии
- Отобраны данные по кусту(13) скважин: дебит, давление, температура, ...(11 параметров) за 122 дня
- Каждая клетка дебита полагалась пробелом и заполнялась алгоритмом
- Критерий качества
  - Средняя относительная ошибка

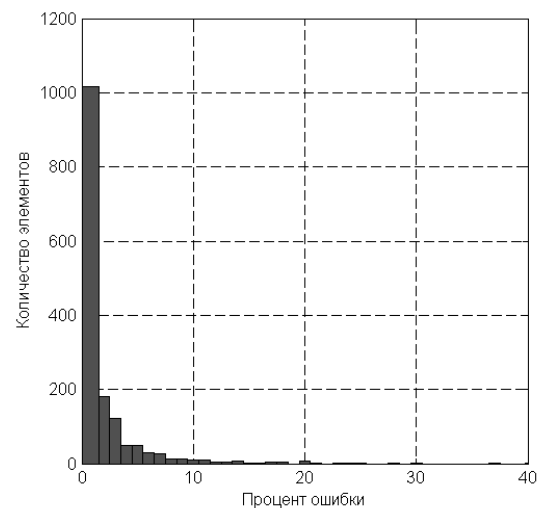
# Заполнение пробелов

- Средняя ошибка: 7% / 2%
- В интервале [0; 10]: 76% / 95%

Без построения подкуба



С построением подкуба



# Результаты - 1

- Предложен собственный метод для заполнения пробелов и определения ошибок в геолого-промысловых данных.
  - Имеет потенциал для его применения в задачах анализа других данных .

# Результаты - 2

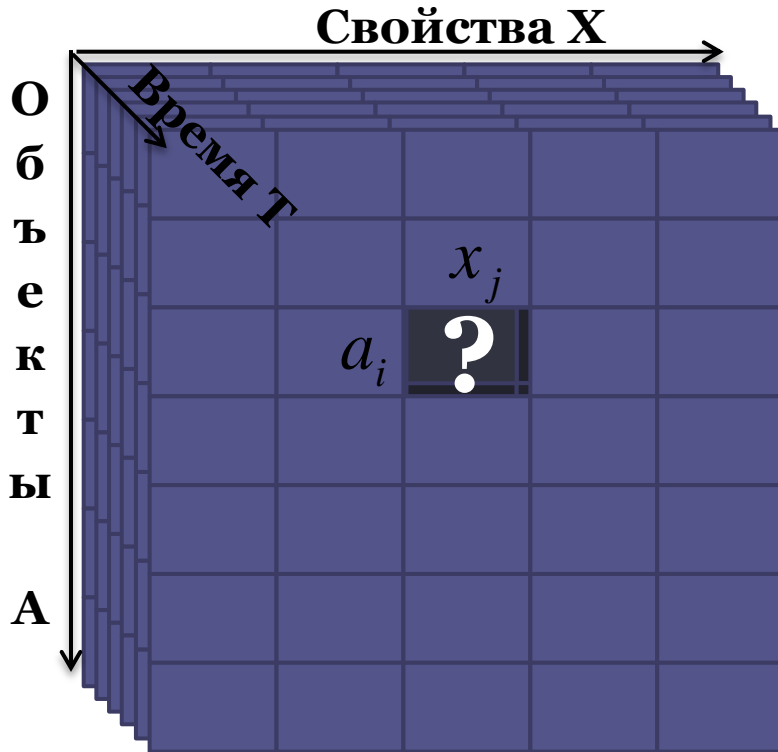
- Разработана интеллектуальная система анализа геолого-промысловых данных с возможностью выделять события для реализации замерной схемы по расписанию.
  - Слабая зависимость компонентов
  - Расширяемость

# Публикации по теме работы

- МНСК-2014 (тезисы)
- Гео-Сибирь-2014 (доклад)
- Сибирский журнал индустриальной математики. Апрель–июнь, 2014 (статья)

**Спасибо за внимание!**

# Пробелы и ошибки



$\langle A, X, T \rangle$  – куб данных

$\langle a_i, x_j, t_k \rangle$  – элемент куба

$\langle a_i, X, T \rangle, \langle A, x_j, T \rangle, \langle A, X, t_k \rangle$  –  
целевые сечения куба

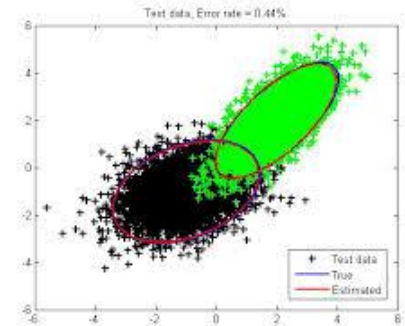
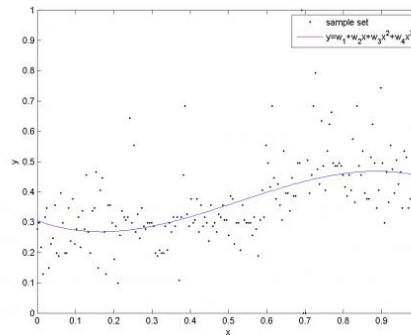
Задача:

- Если элемент не определён, то заполнить.
- Если элемент определён, не содержит ли он ошибку?



# Алгоритмы заполнения

- Простые:
  - Интерполяция
  - Скользящее среднее
- Сложные
  - Регрессия
  - EM-алгоритм



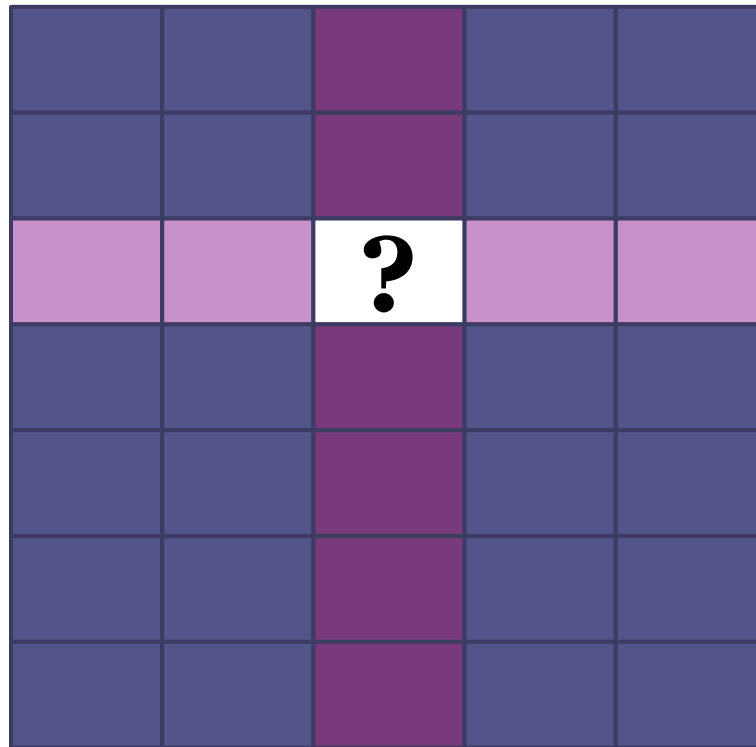
- Действуют глобально: зависимость должна быть реализована на всех объектах выборки

# Алгоритм ZET

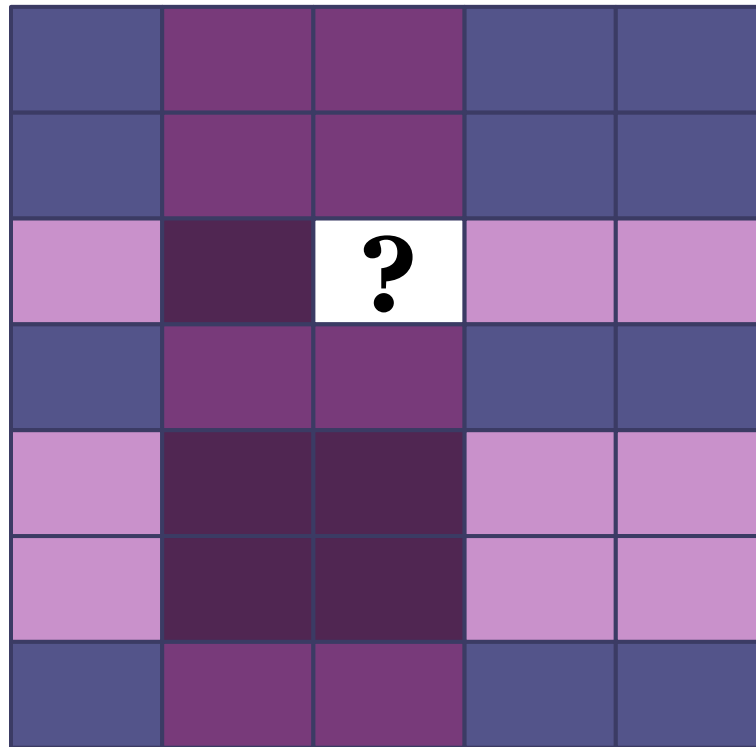
- Используется для двумерных таблиц
- Основные предположения:
  - Избыточность
  - Локальная компактность
  - Линейные зависимости в данных
- Действует локально:  
Выбирает подтаблицу и прогнозирует элемент

Загоруйко Н. Г., Елкина В. Н., Темиркаев В. С. Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм ZET) // Эмпирическое предсказание и распознавание образов . Новосибирск, 1975, Вып.61 , Вычислительные системы. С .3-27.

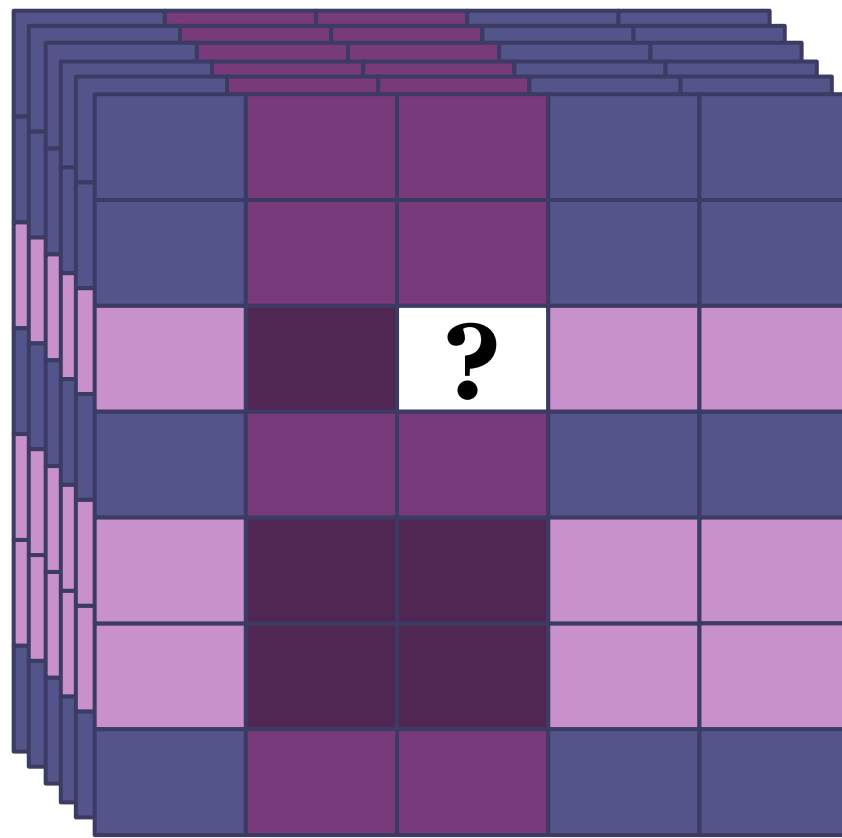
# Алгоритм ZET



# Алгоритм ZET

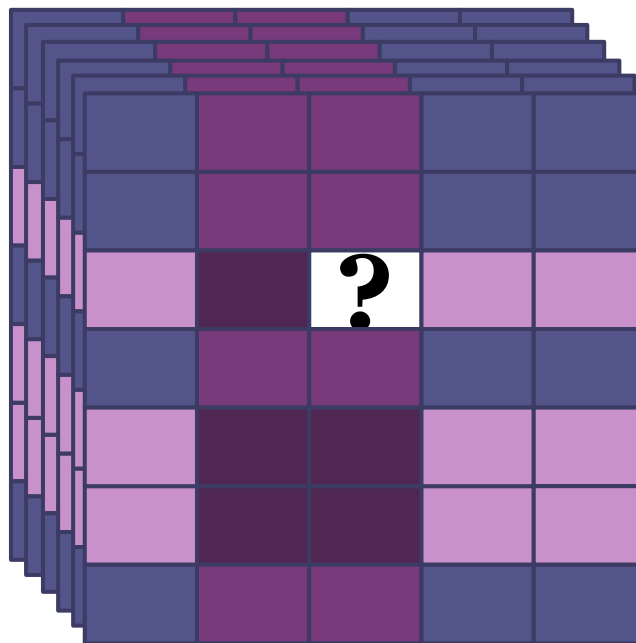


# Переход к кубу данных



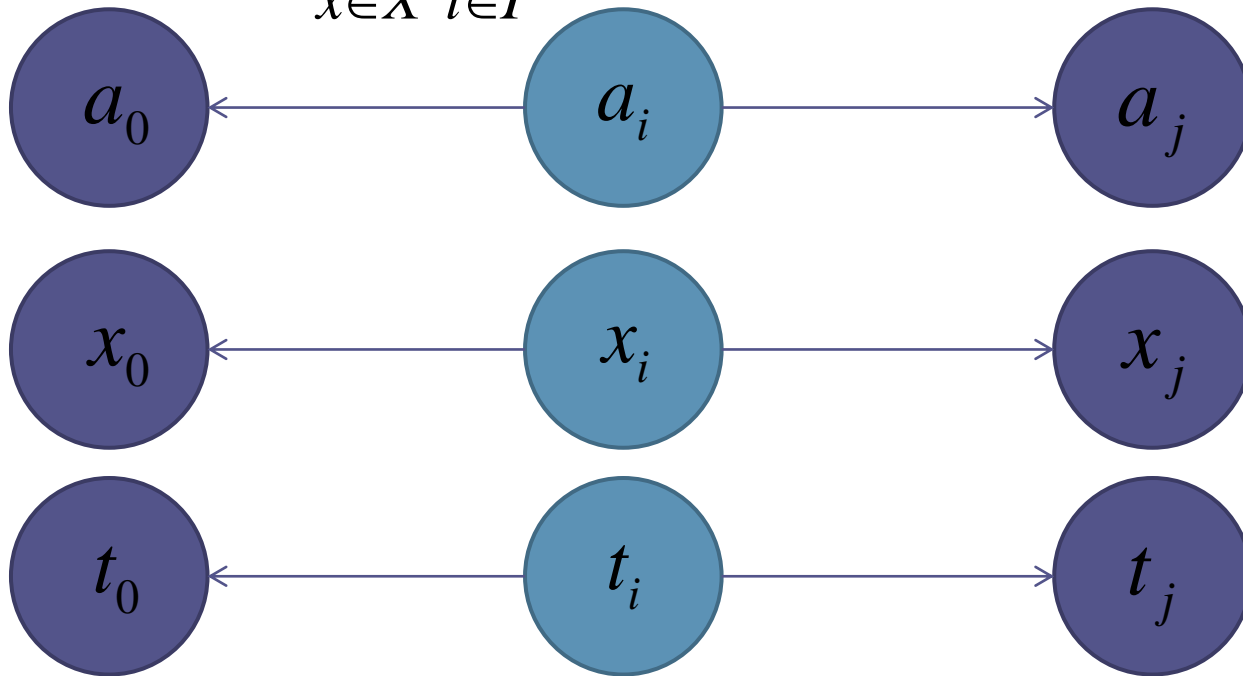
# Переход к кубу данных

- Предложить меру сходства сечений
- Предложить метод оценки компактности



# Сходство сечений

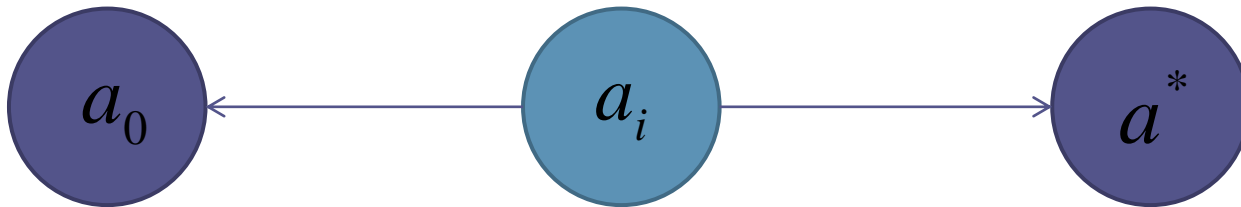
$$r^2(a_i, a_j) = \sum_{x \in X} \sum_{t \in T} \left( \langle a_i, X, T \rangle - \langle a_j, X, T \rangle \right)^2$$



# Конкурентное сходство - 1

$$F(a_i, a_0 | a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

$$F(a_i, a_0 | a^*) \approx 0$$

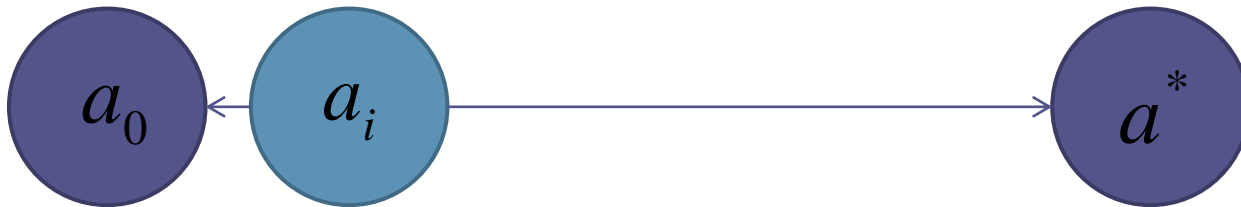




## Конкурентное сходство - 2

$$F(a_i, a_0 | a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

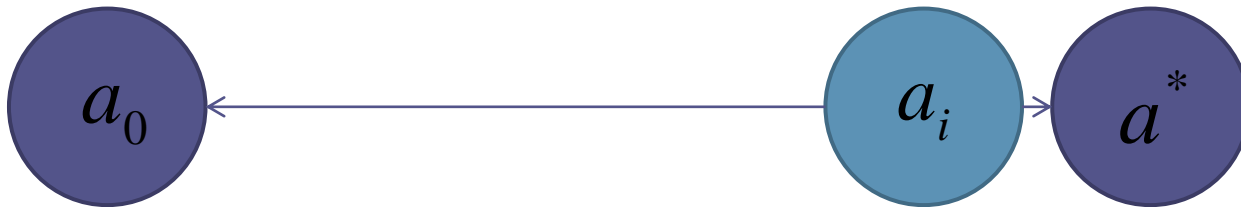
$$F(a_i, a_0 | a^*) \approx 1$$



## Конкурентное сходство - 3

$$F(a_i, a_0 | a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

$$F(a_i, a_0 | a^*) \approx -1$$

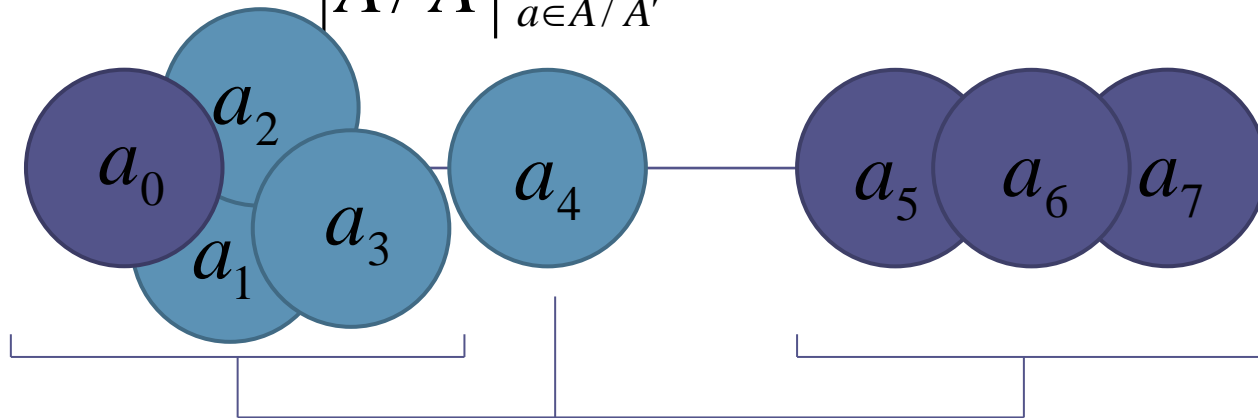


# Компактность сечений

$$C(A') = \sum_{a \in A'} F(\bar{a}_0, a \mid \bar{a}^*)$$

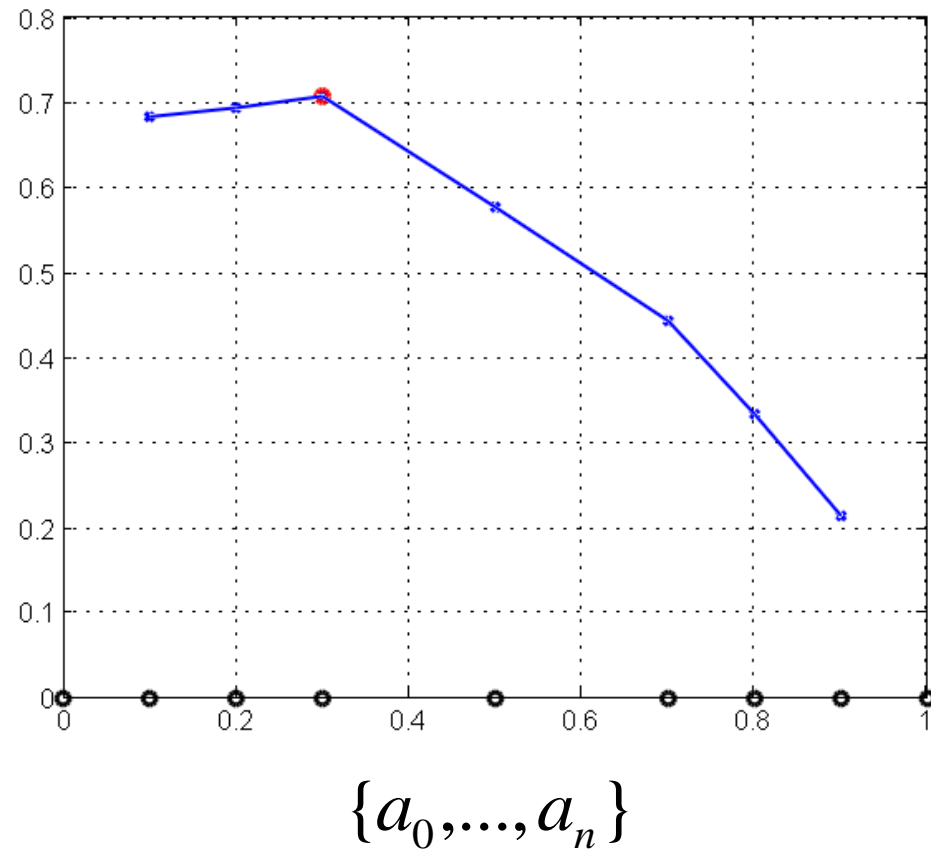
$$\bar{a}_0 : r(a_0, \bar{a}_0) = \frac{1}{|A'|} \sum_{a \in A'} r(a_0, a)$$

$$\bar{a}^* : r(a_0, \bar{a}^*) = \frac{1}{|A/A'|} \sum_{a \in A/A'} r(a_0, a)$$



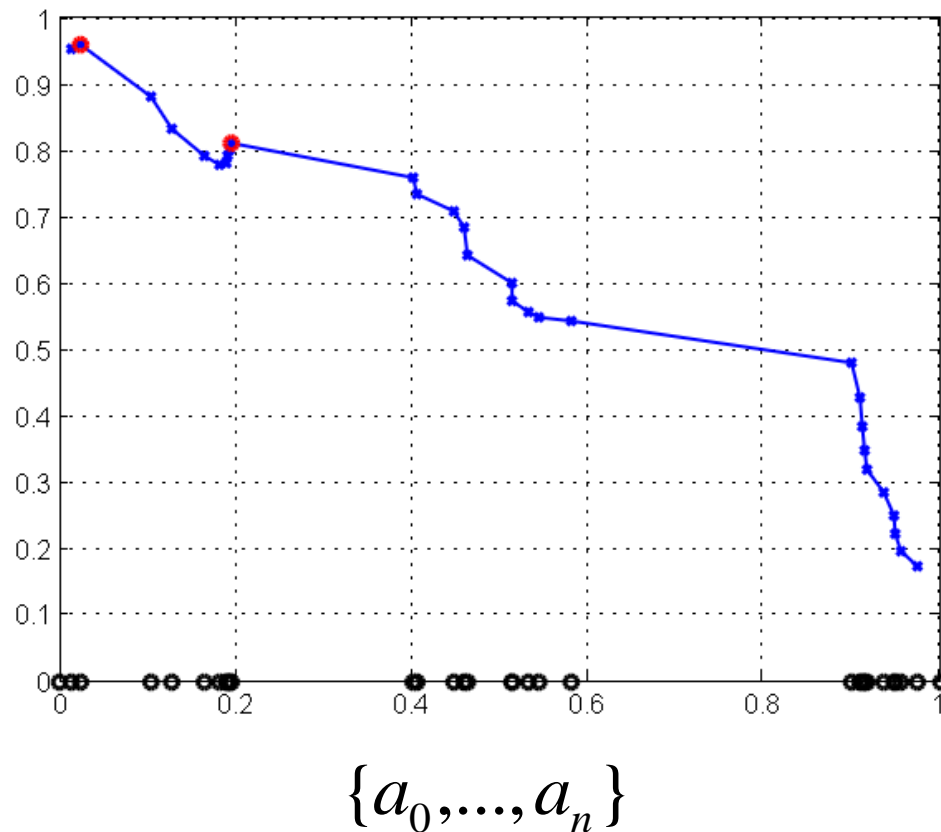
# Компактность сечений

$C(\{a_0, \dots, a_i\})$



# Компактность сечений

$C(\{a_0, \dots, a_i\})$



# Построение подкуба

- Дано:

$\langle A, X, T \rangle$  – куб данных.

Значение  $\langle a_0, x_0, t_0 \rangle$  – ?

- Найти:

$\langle A', X', T' \rangle$  – компактный подкуб.

$\langle a_0, x_0, t_0 \rangle = M_{\langle A', X', T' \rangle}(a_0, x_0, t_0)$

$C\langle A', X', T' \rangle \xrightarrow[\substack{A' \subset A \\ X' \subset X \\ T' \subset T}]{} \max$

# Приближённый алгоритм - 1

$$\langle A', X', T' \rangle = \langle A, X, T \rangle$$

- Положим искомый подкуб равным исходному кубу:

$$C\langle A', X', T' \rangle$$

- Повторять:
  - Найдём наиболее удалённые от целевых сечения  $a^*, x^*, t^*$ .
  - Вычислим компактность
  - Если не выполнен критерий остановки:
    - Исключаем сечения с отрицательным сходством
  - Иначе: конец.

# Приближённый алгоритм - 2

- Критерии остановки:

1. Локальный максимум
2. Достижение подкубом минимального размера
3. Если по одному из направлений количество сечений с отрицательным сходством слишком велико и их исключение приведёт к значительному уменьшению размеров подкуба, целесообразно не исключать сечения на этом шаге, а подождать следующего. Выполнение условия 3 по всем направлениям останавливает алгоритм.



# Приближённый алгоритм - 3

- После итерации удаления плоскостей целесообразно пополнить подкуб небольшим числом (<50%) случайно выбранных сечений отброшенных на предыдущих шагах

**Средство защиты от локальных экстремумов**

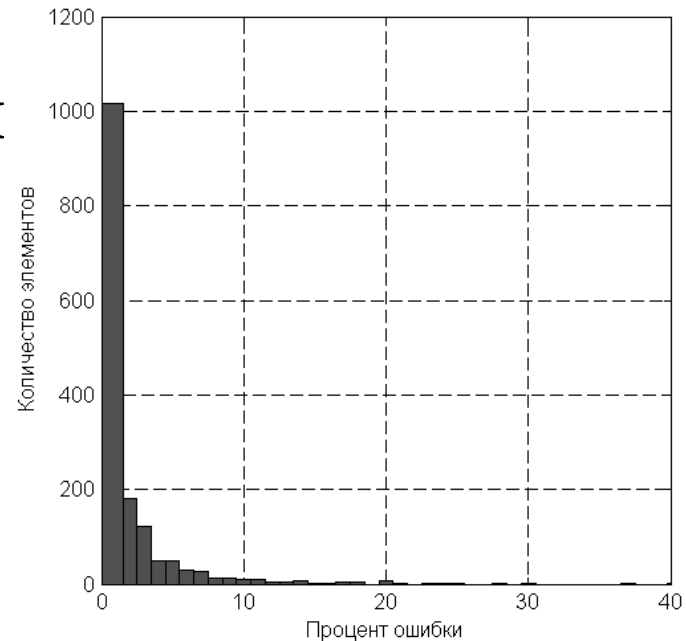
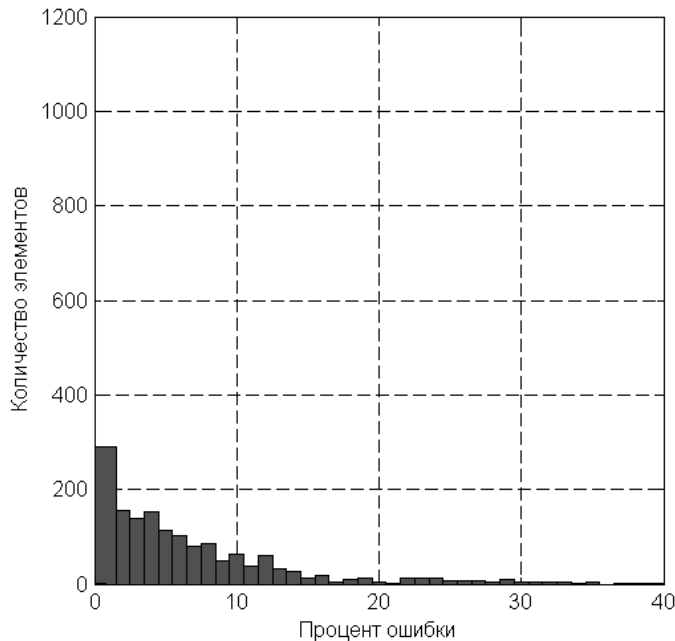
# Заполнение пробела

$$M_{\langle A', X', T' \rangle}(a_0, x_0, t_0) \approx \langle a_0, x_0, t_0 \rangle$$

- Выбор модели прогнозирования
$$\langle a_0, x_0, t_0 \rangle = \sum_{t \in T'} \langle a_0, x_0, t \rangle$$
- Зависит от природы данных и характера задачи:
  - Усреднение по времени:
  - Усреднение по объектам (алгоритм kNN)
  - Регрессия по строкам/столбцам
  - ...



# Данные



- 7%/2% средняя относительная ошибка
- 76%/95% клеток в интервале [0;10]