

# Правдоподобные рассуждения в анализе данных

О.М.Аншаков

РГГУ, кафедра МЛиИС

15 октября 2009 г.

Интеллектуальный анализ данных — аналог терминов  
*Data Mining, Knowledge Discovery in Databases* и  
*Information Extraction.*

Интеллектуальный анализ данных — аналог терминов  
*Data Mining, Knowledge Discovery in Databases* и  
*Information Extraction.*

Data Mining — термин ввел Г. Пятецкий-Шапиро в 1992  
году в статье *W. Frawley, G. Piatetsky-Shapiro,*  
*C. Matheus (1992) Knowledge Discovery in*  
*Databases: An Overview, AI Magazine, pp.*  
*213-228*



**Data Mining** (по Г. Пятецкомку-Шапиро) — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

**Data Mining** (по Г. Пятецкомку-Шапиро) — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

**Data Mining** (определение из Wikipedia) — Data mining is the process of extracting *patterns* from data.

**Data Mining** (по Г. Пятецкомку-Шапиро) — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

**Data Mining** (определение из Wikipedia) — Data mining is the process of extracting *patterns* from data.

Термин *pattern* используется здесь вместо *knowledge*, *information* или *regularity*.

# Данные (факты) и знания

## Основные свойства



Данные (факты) — конкретные утверждения о конкретных предметах.

# Данные (факты) и знания

## Основные свойства

**Данные (факты)** — конкретные утверждения о конкретных предметах.

**Знания** — описания общих закономерностей.

**Данные (факты)** — конкретные утверждения о конкретных предметах.

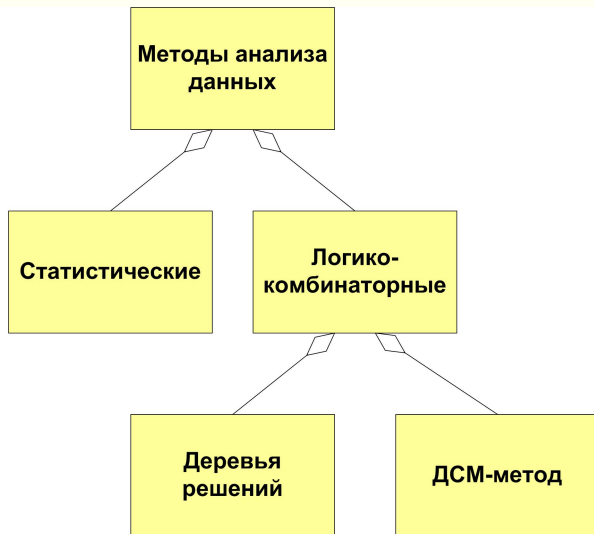
**Знания** — описания общих закономерностей.

Таблица: Особенности данных и знаний

	<b>Данные (факты)</b>	<b>Знания</b>
<b>Описание</b>	Конкретные утверждения о конкретных предметах	Описания общих закономерностей
<b>Источник</b>	С помощью наблюдений и экспериментов	С помощью анализа данных экспериментов и наблюдений и с помощью построения теории
<b>Представление</b>	С помощью атомарных формул без переменных	С помощью универсальных формул или правил с переменными

# Методы анализа данных

## Разновидности



Статистические и логико-комбинаторные методы имеют свои особенности, области применения, преимущества и недостатки.

Таблица: Особенности методов анализа данных

	<b>Статистические</b>	<b>Логико-комбинаторные</b>
<b>Размер данных</b>	Большой	Малый и средний
<b>Устойчивость</b>	Устойчивы к случайным ошибкам	Чувствительны к качеству данных
<b>Область применения</b>	Для обнаружения статистических закономерностей	Для обнаружения закономерностей детерминистского характера

Можно указать следующие особенности приведенных выше примеров логико-комбинаторных методов анализа данных (деревья решений и ДСМ-метод).

**Таблица:** Особенности логико-комбинаторных методов анализа данных

	<b>Деревья решений</b>	<b>ДСМ-метод</b>
<b>Логика</b>	Классическая	Неклассическая
<b>Метод порождения правил</b>	Разбор случаев	Индукция и фальсификация
<b>При плохих данных</b>	Порождаются ошибочные правила	Не порождаются правила

ДСМ-метод — это результат синтеза трех познавательных процедур:

- индукции,
- аналогии и
- абдукции.

Особенности ДСМ-метода:



Особенности ДСМ-метода:

- Систематическое использование правдоподобных (недостовверных) рассуждений.

### Особенности ДСМ-метода:

- Систематическое использование правдоподобных (недостовверных) рассуждений.
- Успешное применение правила ДСМ-метода можно интерпретировать как замену *неопределенности* на *определенность*.

### Особенности ДСМ-метода:

- Систематическое использование правдоподобных (недостовверных) рассуждений.
- Успешное применение правила ДСМ-метода можно интерпретировать как замену *неопределенности* на *определенность*.
- Постепенное сужение области неопределенности в данных за счет правдоподобных рассуждений вплоть до достижения стабилизации.

**Рассуждения** — это процесс получения заключений из уже обоснованного набора посылок по частично или полностью формализованным правилам, а также история этого процесса, представленная в частично или полностью формализованном виде.

**Рассуждения** — это процесс получения заключений из уже обоснованного набора посылок по частично или полностью формализованным правилам, а также история этого процесса, представленная в частично или полностью формализованном виде.

Правила рассуждений называются *достоверными*, если из истинных посылок *обязательно* получается истинное заключение.

**Рассуждения** — это процесс получения заключений из уже обоснованного набора посылок по частично или полностью формализованным правилам, а также история этого процесса, представленная в частично или полностью формализованном виде.

Правила рассуждений называются *достоверными*, если из истинных посылок *обязательно* получается истинное заключение.

Правила рассуждений называются *правдоподобными*, если из истинных посылок *не обязательно* получается истинное заключение.

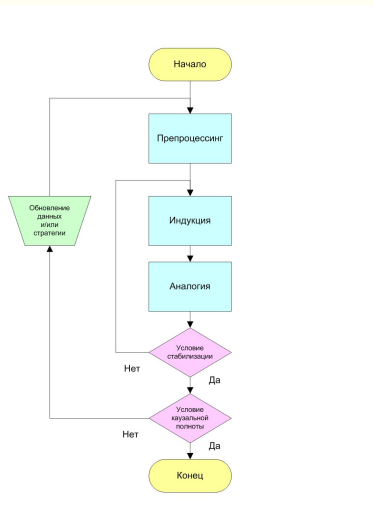
Рассуждения называются *достоверными*, если при их проведении используются только достоверные правила.

Рассуждения называются *достоверными*, если при их проведении используются только достоверные правила. В противном случае рассуждения называются *правдоподобными*.



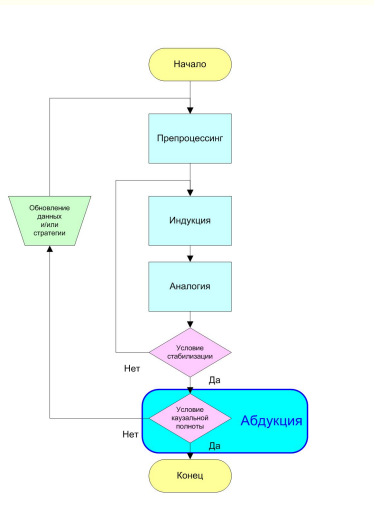
# ДСМ-метод

## Основной алгоритм



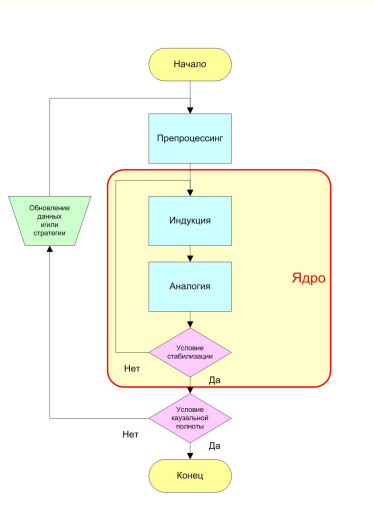
# ДСМ-метод

## Основной алгоритм



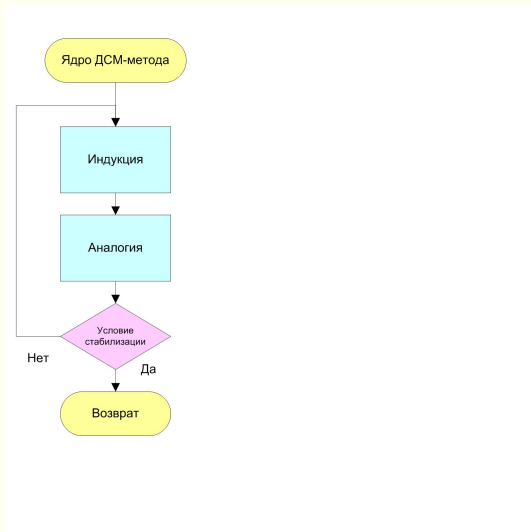
# ДСМ-метод

## Основной алгоритм



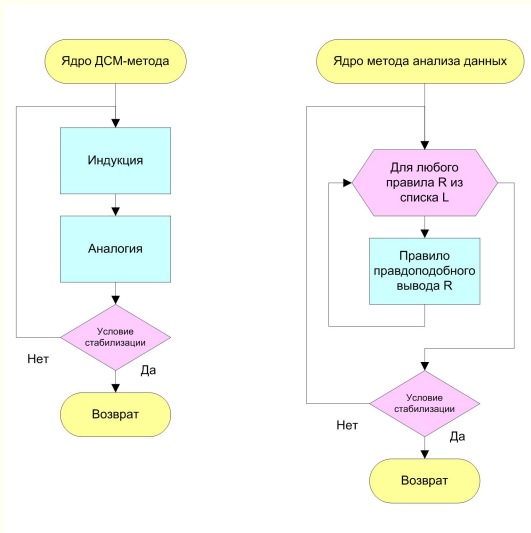
# ДСМ-метод

## Ядро



# ДСМ-метод

## Ядро



# Особенности предлагаемого подхода

## Общая характеристика

# Особенности предлагаемого подхода

## Общая характеристика

Данный подход был предложен при работе над книгой

**О. Anshakov, T. Gergely (2010)**

**Cognitive Reasoning: A Formal Approach**

# Особенности предлагаемого подхода

## Общая характеристика

Данный подход был предложен при работе над книгой  
**О. Anshakov, T. Gergely (2010)**

### **Cognitive Reasoning: A Formal Approach**

В этой книге предлагается общий подход к построению систем анализа данных и его логико-математическое обоснование. Особенности данного подхода.



Данный подход был предложен при работе над книгой  
**О. Anshakov, T. Gergely (2010)**

### **Cognitive Reasoning: A Formal Approach**

В этой книге предлагается общий подход к построению систем анализа данных и его логико-математическое обоснование. Особенности данного подхода.

- Предлагается платформа, а не конкретное решение.

Данный подход был предложен при работе над книгой  
**О. Anshakov, T. Gergely (2010)**

### **Cognitive Reasoning: A Formal Approach**

В этой книге предлагается общий подход к построению систем анализа данных и его логико-математическое обоснование. Особенности данного подхода.

- Предлагается платформа, а не конкретное решение.
- Заранее не предъявляется никаких содержательных требований к списку правил правдоподобных рассуждений. В частности, не утверждается, что эти правила должны соответствовать определенной разновидности правдоподобных рассуждений (индукции, аналогии, абдукции и т.п.). Выбор как типов правил, так и конкретная формулировка правил относится к решениям на базе предлагаемой платформы.
- К списку правил правдоподобных рассуждений предъявляется ряд формальных требований.