

Министерство образования Российской Федерации
Казанский государственный технический университет им. А.Н.Туполева
Кафедра АСОИУ

Интеллектуальные информационные системы

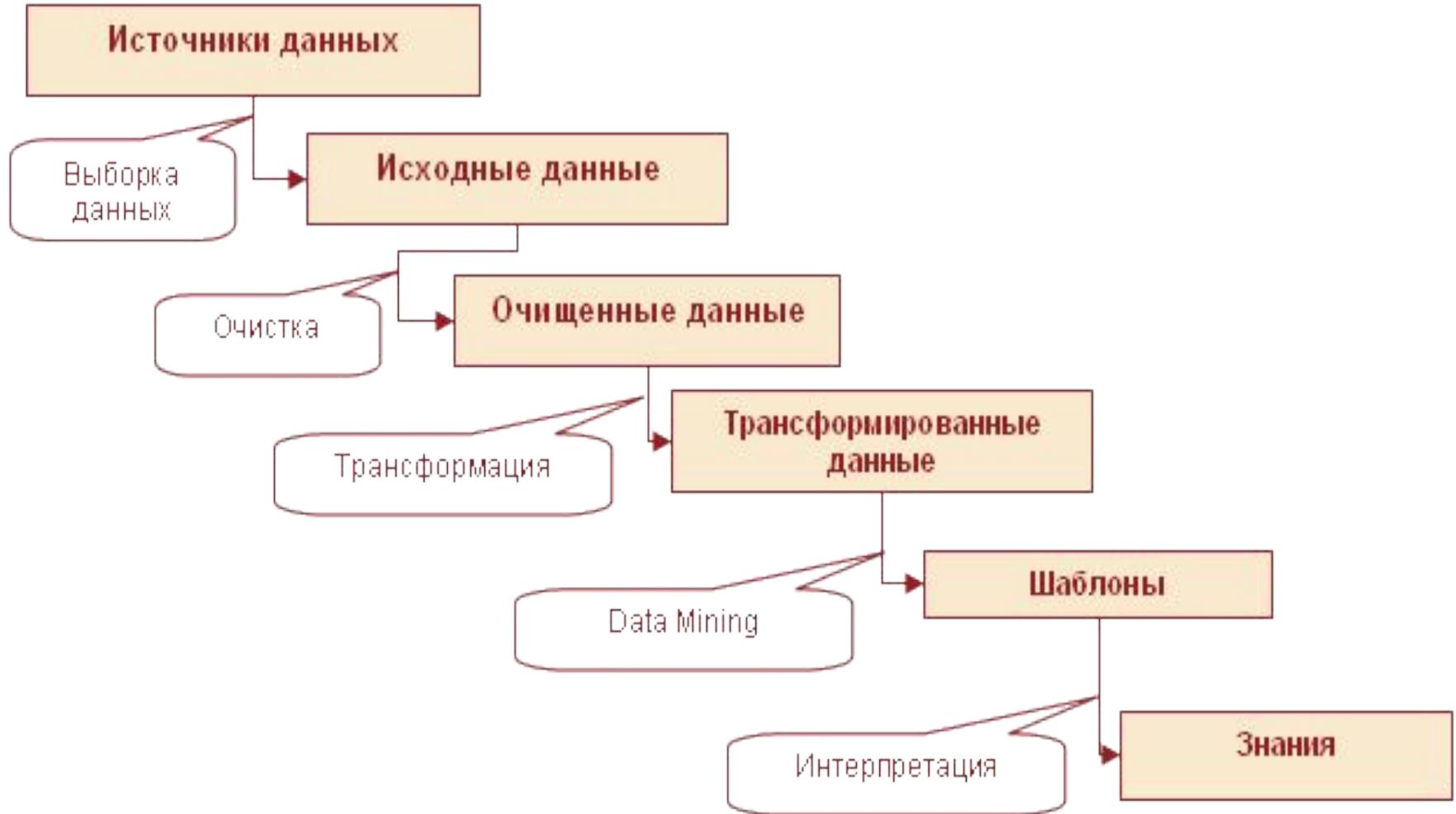
Извлечение знаний из эмпирических данных

2011

Методы приобретения знаний

Источники знаний	Методы приобретения знаний
Знания и опыт экспертов	Коммуникационные
Тексты	Текстологические
Экспериментальные данные	KDD

Knowledge Discovery in Databases



Обучение

Примеры

Начальные
знания

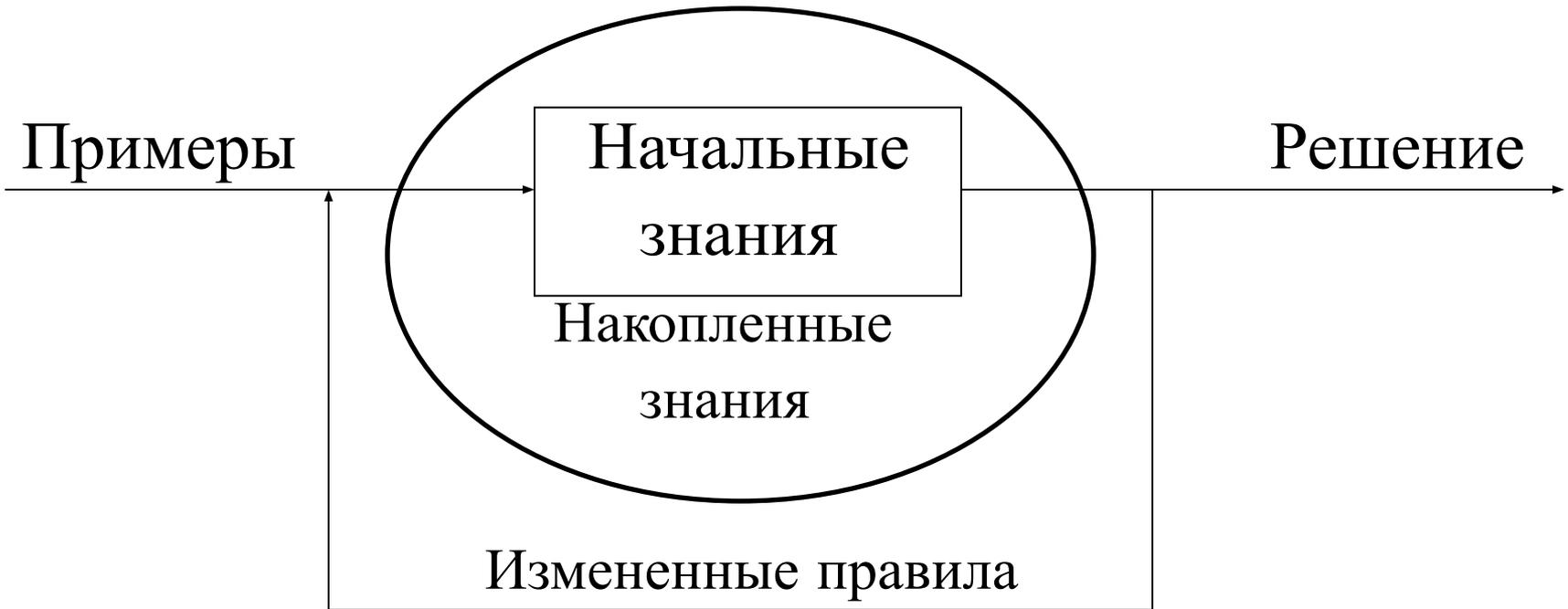
Решение



Обучение



Обучение



Обучение



Эффективность решения:

- более качественное, точное решение;
- уменьшение времени решения;
- расширение круга задач.

Этапы обучения

- получение знаний без логических выводов
- получение знаний с правилами
- формирование обобщений

Методы обобщения

АНАЛОГИЯ Вывод подобного заключения, если существуют подобные посылки

ИНДУКЦИЯ Вывод из имеющихся данных объясняющего их правила

Обучение с учителем: поставляются гипотезы.

Обучение без учителя: гипотезы генерируются
автоматически

Формирование обобщений

Обобщение - переход от единичного объекта o или некоторого множества объектов O к рассмотрению множества объектов V :

$$o \in V \quad \text{или} \quad O \subset V$$

Понятие - результат обобщения

$$O = \{ o_1, o_2, \dots, o_n \} \quad \text{- объем понятия}$$

Если $o_i \in O$, o_i – положительный пример,

$o_i \notin O$, o_i – отрицательный пример.

Формирование понятий

Z - некоторая система знаний

$$Z = \{ Z_1, Z_2, \dots, Z_k \}$$

Z_i - характеризуется множеством значений:

$$Z_i = \{ z_{i1}, z_{i2}, \dots, z_{im} \} \quad 1 \leq i \leq k$$

Типы

признаков:- количественные,

- качественные с упорядоченными значениями,

- качественные с неструктурированными

значениями

Объект: $O_i = \{ z_{ij} \}, \quad z_{ij} \in Z_i, \quad 1 \leq j \leq m, \quad 1 \leq i \leq k$

Задача индуктивного формирования понятий

Пусть R – множество объектов,

V – множество положительных примеров,

W – множество отрицательных примеров.

Очевидно, $R = V \cup W$ $V \cap W = \emptyset$

K – обучающая выборка $K \subset R$

$K = K^+ \cup K^-$ $K^+ \subset V$ $K^- \subset W$

Задача индуктивного формирования понятий

1. В результате обобщения формируется некоторое разделяющее правило. Используется обучающая выборка.
2. Правильность обобщения проверяется на экзаменационной выборке. Экзаменационная выборка не совпадает с обучающей !
3. При неудаче в п.2 экзаменационная выборка добавляется к обучающей выборке и повторяется п.1.

Качественное обобщение

Z_{ij} - j -е значение i -го признака

Каждому Z_{ij} поставим в соответствие булеву переменную h_{ij}
:

$$h_{ij} = \begin{cases} 1, & \text{если } i\text{-й признак имеет } j\text{-е значение} \\ 0, & \text{если } i\text{-й признак не имеет } j\text{-е значение} \end{cases}$$

Конъюнктивное обобщение – выражение вида: $\& h_{ij}$

Дизъюнктивное понятие: $\cup \alpha_k$, где α_k - k -ое конъюнктивное обобщение

Алгоритм качественного обобщения. Пример

Обучающая выборка

	a	b	c	d
1	1	1	1	1
2	1	0	1	0
3	1	0	1	1
4	1	1	1	1
5	0	1	0	1
-6	0	1	1	1
-7	1	0	0	1
-8	0	0	1	0
-9	0	0	1	1

Алгоритм качественного обобщения. Пример

Обучающая выборка

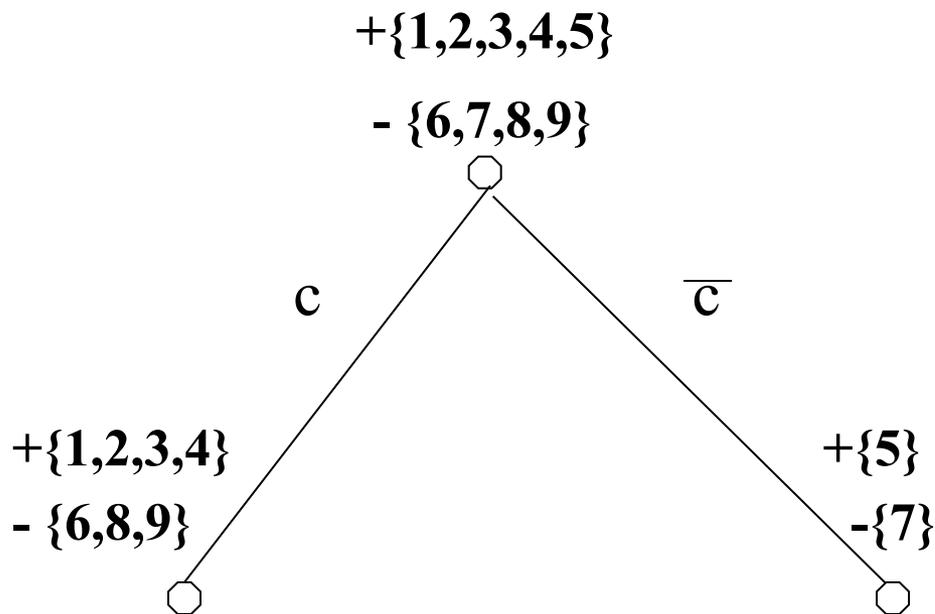
	a	b	c	d
1	1	1	1	1
2	1	0	1	0
3	1	0	1	1
4	1	1	1	1
5	0	1	0	1
- 6	0	1	1	1
- 7	1	0	0	1
- 8	0	0	1	0
- 9	0	0	1	1

a	1	5/9
	0	4/9
b	1	4/9
	0	5/9
c	1	7/9
	0	2/9
d	1	7/9
	0	2/9

Алгоритм качественного обобщения. Пример

Обучающая выборка

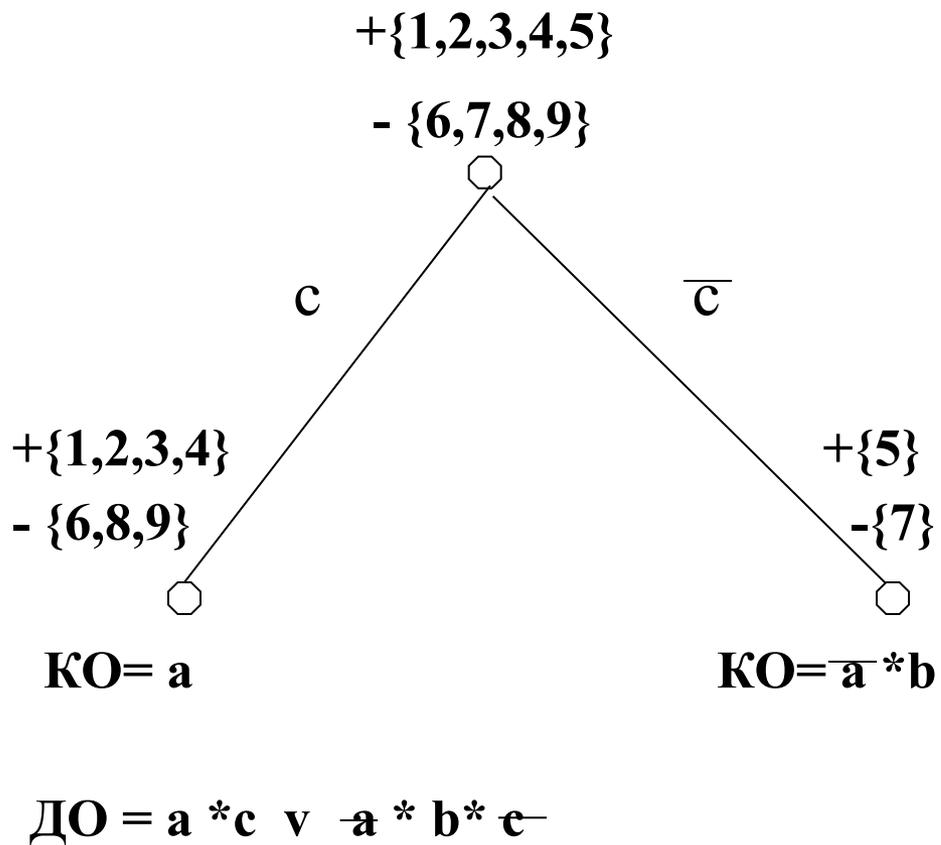
	a	b	c	d
1	1	1	1	1
2	1	0	1	0
3	1	0	1	1
4	1	1	1	1
5	0	1	0	1
-6	0	1	1	1
-7	1	0	0	1
-8	0	0	1	0
-9	0	0	1	1



Алгоритм качественного обобщения. Пример

Обучающая выборка

	a	b	c	d
1	1	1	1	1
2	1	0	1	0
3	1	0	1	1
4	1	1	1	1
5	0	1	0	1
-6	0	1	1	1
-7	1	0	0	1
-8	0	0	1	0
-9	0	0	1	1



Data Mining (DM)

- Знания должны быть новые
- Знания должны быть нетривиальные
- Знания должны быть практически полезны
- Знания должны быть доступны для

понимания

Средства представления знаний в DM – модели:

- правила
- деревья решений
- кластеры
- математические функции

Этапы Data Mining



Задачи Data Mining

- Классификация
- Кластеризация
- Регрессионный анализ
- Ассоциации

Методы Data Mining

- Методы классификация
- Методы кластеризация
- Регрессионный анализ
- Прогнозирование временных рядов
- Ассоциации
- Нейронные сети
- Эволюционные алгоритмы