硕士研究生课程《智能信息处理》

# 形式概念分析

大连海事大学信息科学技术学院

### 第2章

## 形式背景

形式背景

形式背景上的计算

形式背景: 进一步讨论



(对象集<sub>n</sub>,属性集<sub>n</sub>)

世界

(对象的集合) 世界是普遍联系的 人脑

(概念的集合) 概念之间有关系

#### 世界

#### 问题置换

人脑

(对象1,属性集1)

(对象,,属性集,)

(对象,,属性集,)

概念1

概念2

• • •

概念,

使概念联系起来■

(形式对象1,形式属性集1)

(形式对象2,形式属性集2)

(形式对象<sub>N</sub>,形式属性集<sub>N</sub>)

形式概念1形式概念2

形式概念N

 $N=2^n$ 

#### 表形式

#### 形式概念

台	然	抽工	人
H	1	1170	15

属性对象	语言	吠
人	<b>√</b>	
狗		<b>√</b>

对象组合 N=2<sup>n</sup> 穷尽枚举

对象属性	语言	吠
人,狗		
人	<b>√</b>	
狗		<b>√</b>
Ø	<b>√</b>	<b>√</b>

#### 图形式

({人},{语言}) ({狗},{吠}) ({人},{语言}) ○ ○

({人,狗},Ø)

({狗},{吠})

(Ø, {语言,吠})

#### 人脑中的客观世界: 形式概念集

形式概念1:(形式对象1,形式属性集1)

形式概念2:(形式对象2,形式属性集2)

形式概念N: (形式对象N, 形式属性集N)

#### 人脑中的形式概念集: 置于二维表

行

列

#### 形式对象集》形式属性集

形式概念1:(形式对象1,形式属性集1)

形式概念2:(形式对象2,形式属性集2)

形式概念N: (形式对象N, 形式属性集N)

## 形式背景(formal context)

	形式属性1	形式属性2	•••	形式属性 <sub>M</sub>
形式对象1	<b>√</b>	<b>√</b>		<b>√</b>
形式对象2		<b>√</b>	<b>√</b>	
•••		•••		
形式对象 <sub>N</sub>	<b>√</b>		<b>√</b>	✓

形式概念表

概念表

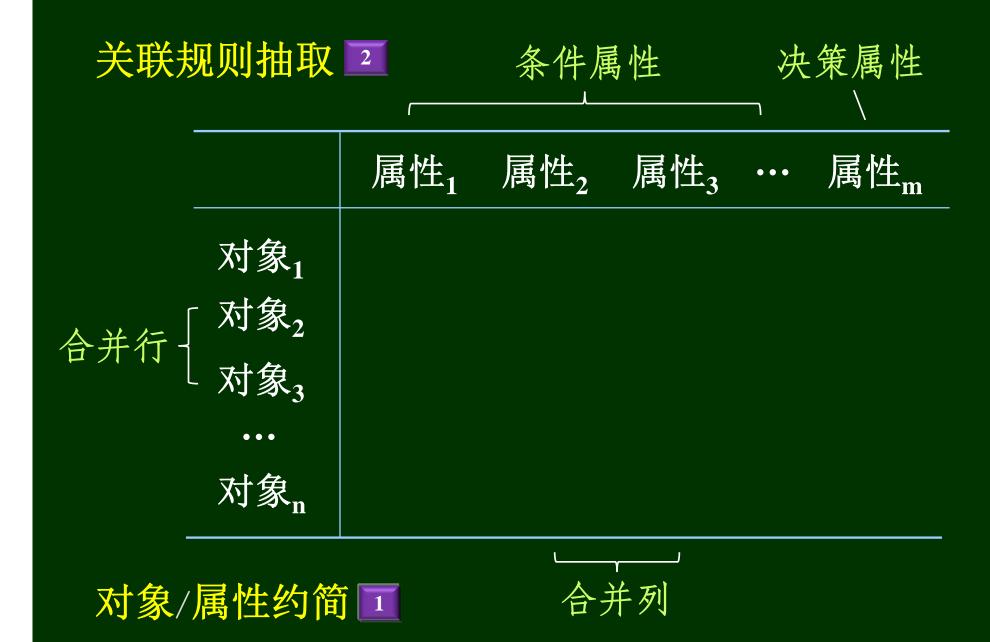
#### 形式背景的特例:背景■

	属性1	属性2	•••	属性 <sub>m</sub>
对象1				
对象2				
•••				
对象n				

#### 形式概念 置于其中的环境

## 形式背景 (formal context)

形式概念置于其中的二维表



背景:某办公用品销售公司的业务活动

	设备	计算机	复印机	打字机	专用装置
咨询	$\mathbf{Y}$	$\mathbf{Y}$	$\mathbf{Y}$	$\mathbf{Y}$	$\mathbf{Y}$
方案	Y	Y			
安装	Y	Y	Y	Y	Y
指导		Y	Y	Y	Y
培训		Y			
配件	Y	Y	Y	Y	Y
维修	Y	Y	Y	Y	Y
服务		Y	$\mathbf{Y}$	Y	

#### 某公司业务活动的背景\_对象约简/属性约简

	设备	计算机	复印机	打字机	专用装置
咨询	Y	Y	Y	Y	$\mathbf{Y}$
方案	Y	Y			
安装	<b>-Y</b>	Y	Y	Y	<u>Y</u>
指导		Y	Y	Y	Y
培训		Y			
配件	<u>-Y</u>	Y	Y	Y	Y
维修	<del>-Y</del> -	Y	Y Y	<del>Y</del>	Y
服务		Y	Y	Y	

#### 某办公用品销售公司业务活动的约简背景

	设备	计算机	复印机/打字机	专用装置
咨询/安装配件/维修	Y	$\mathbf{Y}$	Y	Y
方案	Y	$\mathbf{Y}$		
指导		Y	Y	Y
培训		Y		
服务		Y	Y	

#### 举例:基于形式背景的关联规则抽取

	a	þ	条 <sub> </sub> 件 C	属 d	ŧ e	f	决策 <b>D</b> <sub>1</sub>	属性 <b>D</b> 2
1	0	1	1	0	1	0	1	1
2	0	0	1	0	0	0	1	0
_3	1	_1_	1	0	1	1	1	1
4	1	1	•	1	-0	1	0	1
5	0	0	1	0		0	1	0

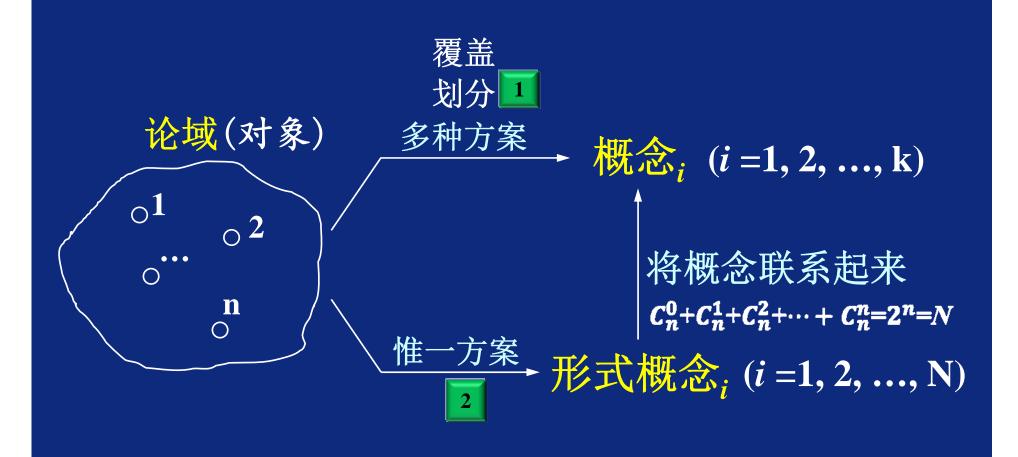
关联规则 If c Then  $D_1$  If b Then  $D_2$  If e Then  $D_1D_2$ 

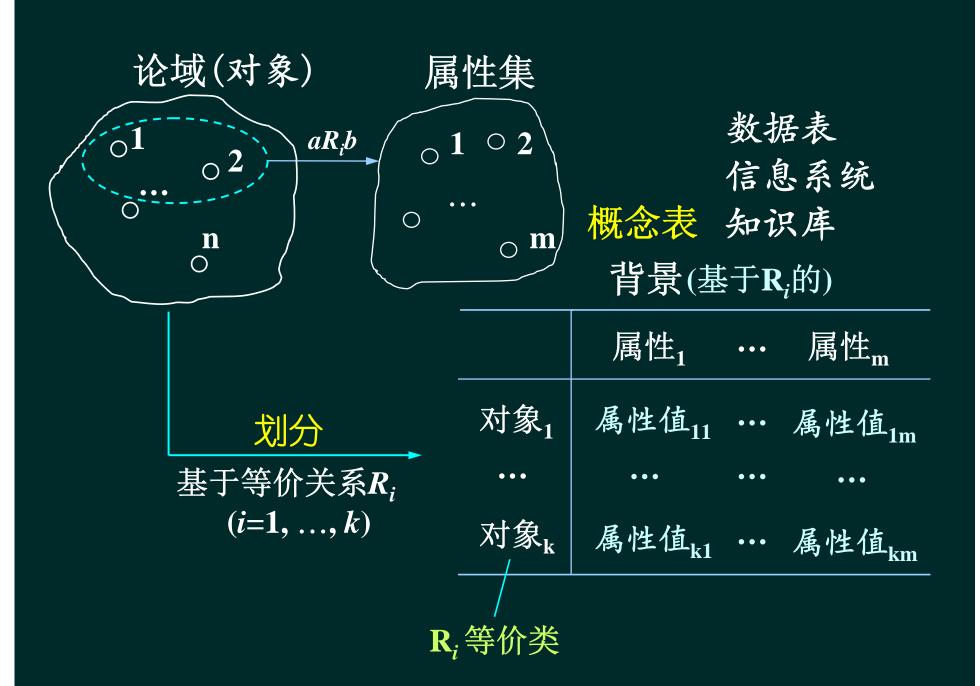
### 形式背景: 进一步的讨论

形式背景的生成

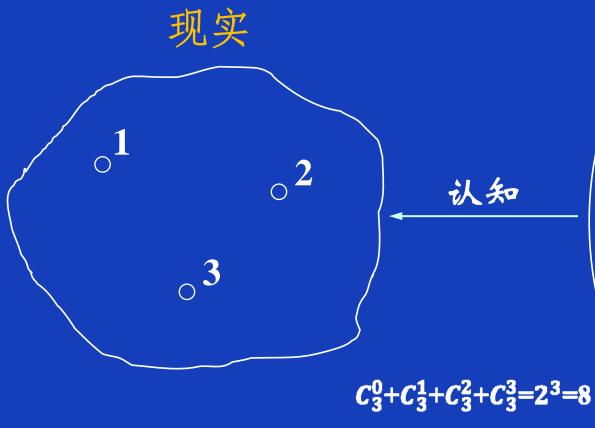
形式背景的类型

形式背景的应用









## **论域 U** (元素 n个)

## 人脑 {1} {2} {3} {1,2} {1,3}

## 幂集 P(U)

{2, 3}

 $\{1, 2, 3\}$ 

(子集 2<sup>n</sup>个)

#### 形式背景

#### 形式对象

{ }

**{1}** 

**{2}** 

**{3}** 

*{*1*,* 2*}* 

*{*1*,* 3*}* 

{2, 3}

 $\{1, 2, 3\}$ 

#### 论域

#### 对象

1

2

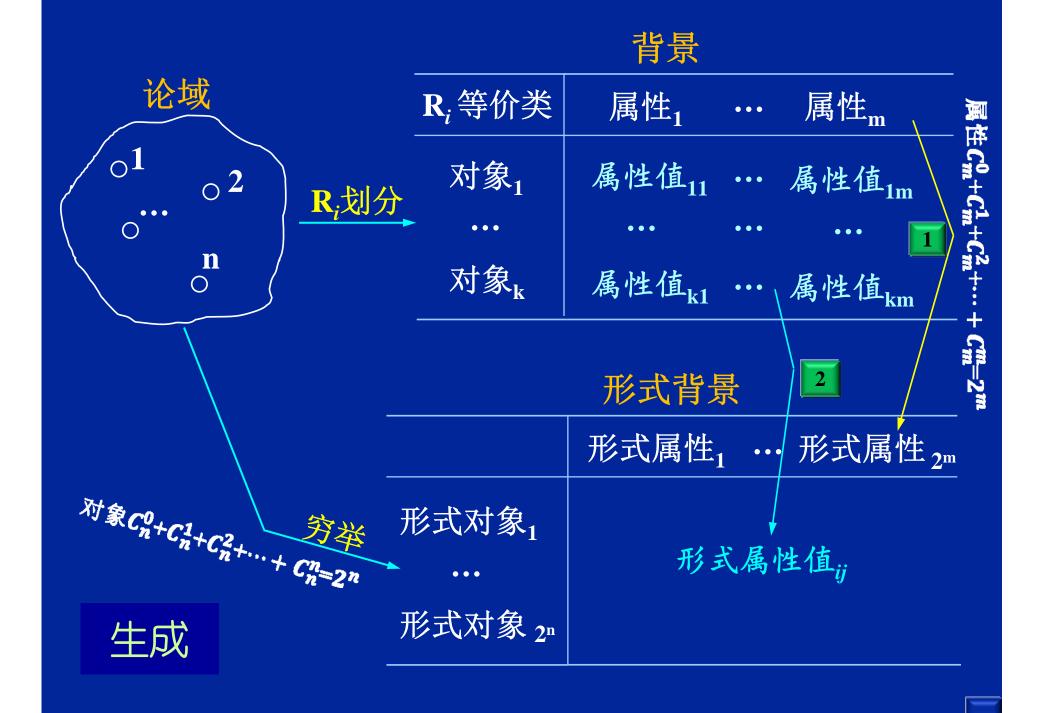
3

#### 组合穷举

$$C_3^0 + C_3^1 + C_3^2 + C_3^3 = 2^3 = 8$$

#### n个 对象

2n个 形式对象



月上	是
月	尽

属性<sub>1</sub> 属性<sub>2</sub> 属性<sub>3</sub>

属性 m 个 形式属性 2<sup>m</sup> 个

#### 形式背景

Ø 属性<sub>1</sub> 属性<sub>2</sub> 属性<sub>3</sub> 属性<sub>12</sub> 属性<sub>13</sub> 属性<sub>23</sub> 属性<sub>123</sub>

过户	見
H	ూ

	属性1	•••	属性m
对象1	属性值 <sub>11</sub>	•••	属性值 <sub>1m</sub>
•••	•••	•••	•••
对象 <sub>k</sub>	属性值 <sub>k1</sub>	•••	属性值 <sub>km</sub>

概述

子类 {1, ..., k}

隶属度[0,1]

#### 形式背景

#### 形式属性值的确定方法

	形式属性1 …	形式属性2m
形式对象1		at the
•••	形式属性	生值 <sub>ij</sub>
形式对象 <sub>2</sub> n		

\ <del>\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\</del>	<b></b>	形式背暑
> Tml/		

- 4	

论域		背	景		
对象		对象	属性1		形式
$x_1$		$\boldsymbol{x_1}$	1		Ø
$x_2$	划分	$x_2$	2	扩展	$\{x_1\}$
$x_3$		$x_3$	2		$\{x_2\}$
					$\{x_3\}$
数据表					$\{x_1, x_1, x_2, \dots, x_n\}$
					$\{x_1, x_1, x_2, x_3, x_4, x_5, x_6, x_6, x_6, x_6, x_6, x_6, x_6, x_6$
知识库					$\{x_2, x_3\}$

形式	背景	$C_1^0 + C_1^1 = 2$
形式对象	Ø	属性1
Ø	<b>√</b>	1, 2
$\{x_1\}$	<b>√</b>	1
$\{x_2\}$	<b>√</b>	2
$\{x_3\}$	<b>√</b>	2
$\{x_1, x_2\}$	<b>√</b>	1, 2
$\{x_1, x_3\}$	<b>√</b>	1, 2
$\{x_2, x_3\}$	<b>√</b>	2, 2
$\{x_1, x_2, x_3\}$	<b>√</b>	1, 2, 2

论域	背	景	形式	背景	
对象	等价类	属性1	形式对象	Ø	属性1
x <sub>1</sub> 划分	$\{x_1\}$	1	Ø	√	1, 2
$x_2$	$\{x_2, x_3\}$	2	$\{x_1\}$	√	1
$x_3$			$\{x_2\}$	<b>√</b>	?
			$\{x_3\}$	✓	?
			$\{x_1, x_2\}$	✓	1,?
粗糙集 等	价类-不可	丁定义集	$-\{x_1, x_3\}$	✓	1,?
	等价类-可			✓	2
	(等价类及	(其并集)	$\{x_1, x_2, x_3\}$	1	1, 2

#### 基于等价类: $\{x_1\} = 1$ ; $\{x_2, x_3\} = 2$ 确定形式属性值

to the second			
下近似集	形式对象	形式属性	上近似集
Ø	Ø	1, 2	Ø
$\{x_1\}$	$\{x_1\}$	1	$\{x_1\}$
Ø	$\{x_2\}$	可能2	$\{x_2, x_3\}$
Ø	$\{x_3\}$	可能2	$\{x_2, x_3\}$
$\{x_1\}$	$\{x_1, x_2\}$	1,可能2	$\{x_1\} \cup \{x_2, x_3\}$
$\{x_1\}$	$\{x_1, x_3\}$	1,可能2	$\{x_1\} \cup \{x_2, x_3\}$
$\{x_2, x_3\}$	$\{x_2, x_3\}$	2	$\{x_2, x_3\}$
$\{x_1\} \cup \{x_2, x_3\}$	$\{x_1, x_2, x_3\}$	1, 2	$\{x_1\} \cup \{x_2, x_3\}$

#### (上近似集-下近似集)中等价类的值

形式属性值

 $\{ 肯定V_{T},可能V_{L-T} \}$ 

下近似集中等价类的值

#### 论域 $U=\{x_1, x_2, x_3\}$ 划分 $U/R_i = \{\{x_1\}, \{x_2, x_3\}\}$

$R_i$ 下近似集	形式对象	$R_i$ 上近似集	
Ø	Ø	Ø	$R_i$ 精确集
$\{x_1\}$	$\{x_1\}$	$\{x_1\}$	$R_i$ 精确集
Ø	$\{x_2\}$	$\{x_2, x_3\}$	$R_i$ 粗糙集
Ø	$\{x_3\}$	$\{x_2, x_3\}$	$R_i$ 粗糙集
$\{x_1\}$	$\{x_1, x_2\}$	$\{x_1, x_2, x_3\}$	$R_i$ 粗糙集
$\{x_1\}$	$\{x_1, x_3\}$	$\{x_1, x_2, x_3\}$	$R_i$ 粗糙集
$\{x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	$R_i$ 精确集
$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	$R_i$ 精确集

#### 论域 $U=\{x_1, x_2, x_3\}$ 划分 $U/R_i = \{\{x_1\}, \{x_2, x_3\}\}$

$\mathbf{R}_i$ 下近似集	形式概念	$R_i$ 上近似集	粗糙度■
Ø	Ø	Ø	0
$\{x_1\}$	$\{x_1\}$	$\{x_1\}$	0
Ø	$\{x_2\}$	$\{x_2, x_3\}$	1
Ø	$\{x_3\}$	$\{x_2, x_3\}$	1
$\{x_1\}$	$\{x_1, x_2\}$	$\{x_1, x_2, x_3\}$	2/3
$\{x_1\}$	$\{x_1, x_3\}$	$\{x_1, x_2, x_3\}$	2/3
$\{x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	0
$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	0

#### 粗糙集的粗糙度的计算

粗糙度=1- 下近似集的基数上近似集的基数

#### 背景

#### 形式背景

14	<u> </u>			
对象	属性1	形式对象	属性1	粗糙度
$\{x_1\}$	1	Ø	1, 2	0
$\{x_2, x_3\}$	2	$\{x_1\}$	1	0
		$\{x_2\}$	2(可能)	1
		$\{x_3\}$	2(可能)	1
		$\{x_1, x_2\}$	1,2(可能)	2/3
		$\{x_1, x_3\}$	1,2(可能)	2/3
		$\{x_2, x_3\}$	2	0
结果		$\{x_1, x_2, x_3\}$	1, 2	0



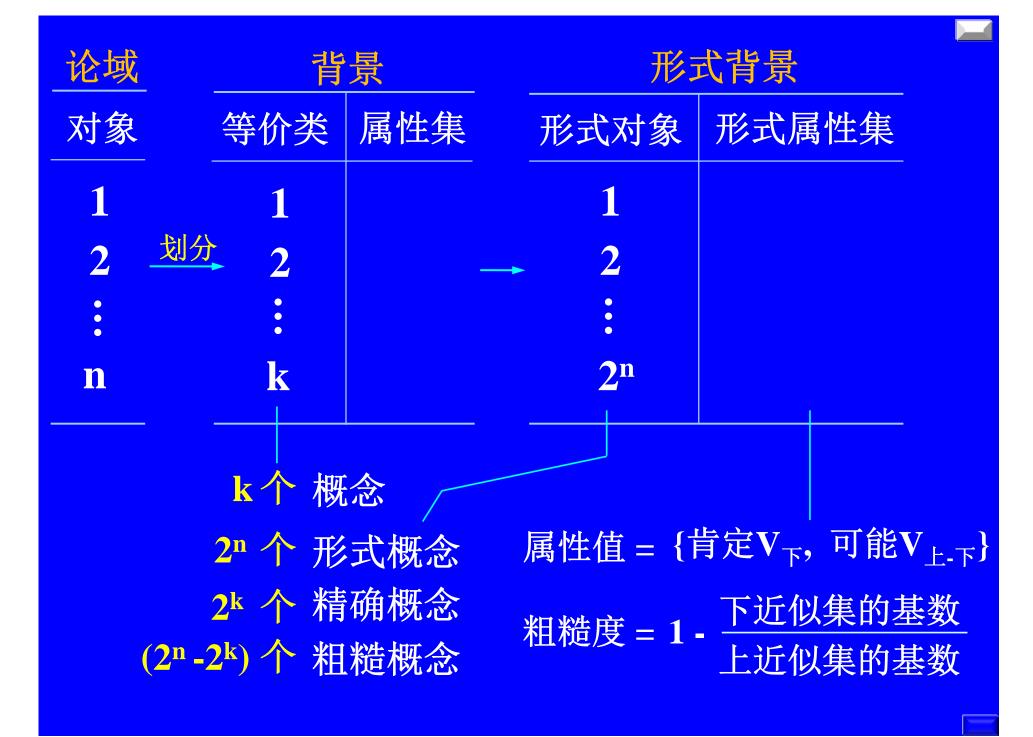
张三 李夫

李妻

#### 基于家庭划分

对象	收入
张三	3000
李夫,李妻	5000

形式对象	收入
Ø	3000, 5000
{张三}	3000
{李夫}	5000(可能)
{李妻}	5000(可能)
{张三,李夫}	3000,5000(可能)
{张三,李妻}	3000,5000(可能)
{李夫,李妻}	5000
{张三,李夫,李妻}	3000, 5000



对象	属性	
$x_1$	1	
$x_2$	0	

x1     1     高个       x2     0     低个     子类     清晰集       x1     2     2     2	对象	身高		
	_		子类 {1,, <i>k</i> }	清晰集

対象
 高个

 
$$x_1$$
 1
 高个
 隶属度 模糊集

  $x_2$ 
 0
 非高个
  $\mu \in [0, 1]$ 

$$\{x_1, x_2\}$$
  $\mu = 1/x_1 + 0/x_2$ 

 $\{x_1, x_2\}$  1, 0

论域
1
2
:

10

#### 背景

	性别	身高
1	男	高
2	女	低
3	女	中
4	男	中
5	男	高
6	男	中
7	女	低
8	男	低

#### 形式背景

	性别	身高
1		
2		
:		
1024		

论
1
2
3

#### 知识库

	高个
1	0.9
2, 3	0.5

形式概念 2<sup>n</sup> = 8个 模糊概念 2<sup>k</sup> = 4个 粗糙模糊概念 4个

#### 形式背景

	高个	
Ø	0.9, 0.5	
1	0.9	
2	0.5(可能)	
3	0.5(可能)	
1, 2	0.9, 0.5(可能)	
1, 3	0.9, 0.5(可能)	
2, 3	0.5	
1, 2, 3	0.9, 0.5	

对象 属性  $\{x_1\}$  1  $\{x_2, x_3\}$  0.6

清晰 类属确定 模糊 类属不确定

新识对象

已有知识

属性
1
0.6(可能)

基于已有知识精确 完全定义 粗糙 不完全定义



精确清晰概念 粗糙清晰概念 精确模糊概念 粗糙模糊概念

形式背景 粗糙形式背景 模糊形式背景 粗糙模糊形式背景

概念 粗糙概念 模糊概念 粗糙模糊概念 
 论域
 背景
 形式背景

 对象
 等价类 属性集
 形式对象 形式属性集

 1
 1
 1

 2
 2
 2

 :
 :
 :
 2k 个精确值

 n
 k
 2n
 (2n-2k) 个粗糙值

子类 属性值域{1,...,k}

隶属度 属性值域 [0,1]

k=n 多值精确清晰 k<n 多值粗糙清晰 k=n 多值精确模糊 k<n 多值粗糙模糊



# 形式背景的八种基本类型

精确知识 k=n

单值

单值模糊

**{0, 1}** 

多值

多值模糊

其它

粗糙知识 k<n

单值粗糙

单值粗糙模糊

 $\{0, 1\}$ 

多值粗糙

多值粗糙模糊

其它

清晰知识

模糊知识

物质

数据 人知其在

信息 人知其义 数据集成

知识 人知其用 信息集成

语词

语词/语句

语句

事实知识 陈述句 规则知识 因果句

- 有效决策 - 有效行为

### 语句类型:基于结构划分和基于功能划分

按结构划分

单句 复句

主谓句

并列句 因果句 转折句 功能

陈述事实 提出疑问 表达祈使 抒发感叹 按功能划分

陈述句知句何何何何

# 形式背景=知识库

→有效决策

知识类型数量

知识获取途径

知识提供方式

# 规则知识 $_i(i = \{0, 1, ..., d\})$

# 背景

			属性 <sub>1</sub>	属性2	•••	属性 <sub>m</sub>
事实知识1	概念1	对象1	属性值11	属性值 <sub>12</sub>	•••	属性值 <sub>1m</sub>
事实知识2	概念2	对象2	属性值21	属性值22	•••	属性值 <sub>2m</sub>
•••	•••	•••	•••	•••	•••	•••
事实知识 <sub>k</sub>	概念k	对象 <sub>k</sub>	属性值 <sub>k1</sub>	属性值 <sub>k2</sub>	•••	属性值 <sub>km</sub>

#### 知识库

	性别	身高
1	男	高
2	女	低
3	女	中
4	男	中
5	男	高
6	男	中
7	女	低
8	男	低

#### 事实知识

第1人是男性且高个。

第2人是女性且低个。

• • •

第8人是男性且低个。

#### 规则知识

若是高个,则为男性。

若是女性,则为中个或低个。

论域		知识库 背景				以库 式背景
对象		对象	属性集		形式对象	形式属性集
1		1			1	
2 _	<b></b>	2		<b></b>	2	
•		•			•	
n		k			<b>2</b> <sup>n</sup>	

k个事实知识 d个规则知识

> 清晰知识 模糊知识

2<sup>n</sup>个 事实知识 d个 规则知识

清晰知识/粗糙知识 模糊知识/粗糙模糊知识 (k=n) (k<n)

#### 关联 d个规则知识



聚类 k个事实知识

### d个规则知识

关联 论域 背景 属性集 对象 粗糙集 n个对象k k个事实知识

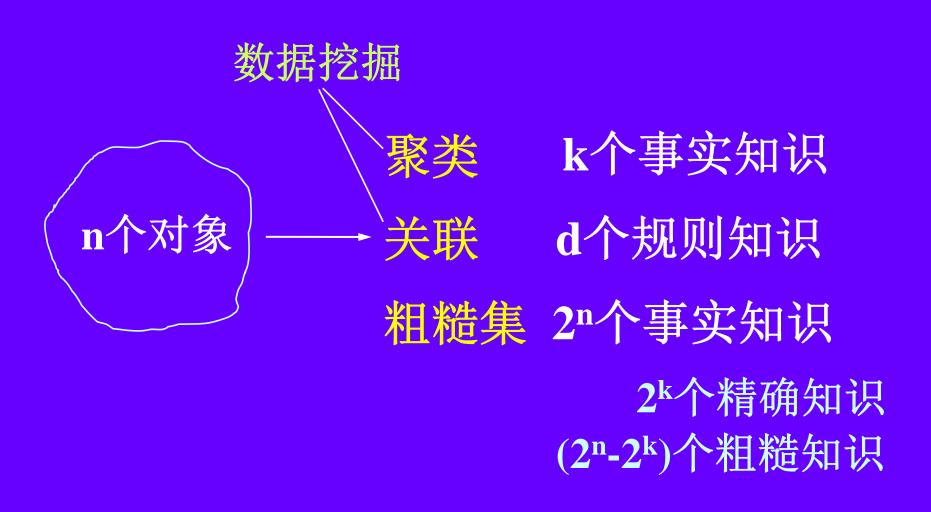
精确知识

形式背景

2<sup>n</sup>个事实知识

精确知识粗糙知识

### 形式背景/知识库的获得:数据挖掘与粗糙集



对象	对象	属性集	
1 2 数据控 : :	1 2 : k		一人类有效决策
		粗糙集	
	形式对象	形式属性集	-
	1		
	2 :		<b>→</b> 人类有效决策
	2 <sup>n</sup>		



