

DOI: 10.11992/tis.201606019

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160808.0831.032.html>

# 数据偏序结构关系中的知识发现可视化方法

郑存芳<sup>1,2</sup>, 洪文学<sup>1</sup>, 李少雄<sup>1</sup>, 任蕴丽<sup>1,3</sup>

(1. 燕山大学 电气工程学院, 河北 秦皇岛 066004; 2. 燕山大学 里仁学院, 河北 秦皇岛 066004; 3. 河北科技师范学院 数学与信息科技学院, 河北 秦皇岛 066004)

**摘 要:**在形式概念分析与偏序结构理论基础之上, 针对决策模式信息表, 提出一种基于认知原理的规则提取与知识发现的可视化新方法——属性偏序决策图。该方法在将决策问题转化为决策模式信息表的基础上, 通过研究对象的属性特征, 将其表现在可视化图形上, 介绍了属性偏序结构图的原理、生成算法及应用实例。实验表明, 属性偏序结构图可以将数据中蕴含的知识和规则得以形象地表示, 通过对属性偏序决策图支路、节点、簇集的分析可以有效地发现数据中蕴含的决策规则。

**关键词:**属性偏序决策图; 偏序结构; 形式概念分析; 可视化; 知识发现

**中图分类号:** TP182 **文献标志码:** A **文章编号:** 1673-4785(2016)04-0475-06

中文引用格式: 郑存芳, 洪文学, 李少雄, 等. 数据偏序结构关系中的知识发现可视化方法[J]. 智能系统学报, 2016, 11(4): 475-480.

英文引用格式: ZHENG Cunfang, HONG Wenxue, LI Shaoxiong, et al. A novel knowledge discovery visualization method based on data partial ordered structure[J]. CAAI Transactions on Intelligent Systems, 2016, 11(4): 475-480.

## A novel knowledge discovery visualization method based on data partial ordered structure

ZHENG Cunfang<sup>1,2</sup>, HONG Wenxue<sup>1</sup>, LI Shaoxiong<sup>1</sup>, REN Yunli<sup>1,3</sup>

(1. School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China; 2. Liren College, Yanshan University, Qinhuangdao 066004, China; 3. College of Mathematics and Information Technology, Hebei Normal University of Science and Technology, Qinhuangdao 066004, China)

**Abstract:** In this paper, the formal concept is first analyzed and partial order structure theory introduced. The decision diagram of attribute partial ordered structure (DDAPOS), a visualization method of rule extraction and knowledge discovery based on cognitive principles, is then proposed. After the decision problem is transformed into a decision pattern information table, the attributes of a research object can be presented in the visualized diagram. This paper introduces the principles, generation algorithm, and application examples of DDAPOS. Experimental results show that the knowledge and rules contained in the data can be represented graphically, and the decision-making rules in the data can be found effectively through analysis of the graph branches, nodes and clusters.

**Keywords:** decision diagram of attribute partial ordered structure; partial ordered structure; formal concept analysis; visualization; knowledge discovery

随着信息技术的发展, 数据的获取变得越来

越简单, 如何有效地利用数据, 从数据海洋中获取有价值的规律、知识、信息成为摆在人们面前的突出问题。一些不依赖于先验知识, 单纯以数据驱动的理论和方法在这一背景下产生和发展, 如形式概念分析<sup>[1]</sup>、粗糙集<sup>[2]</sup>、模糊集<sup>[3]</sup>、商空

收稿日期: 2016-06-06. 网络出版日期: 2016-08-08.

基金项目: 国家自然科学基金项目(61273019, 61473339, 61501397); 河北省自然科学基金重点项目(F2016203443); 燕山大学青年教师自主研究计划课题(13LGB033).

通信作者: 洪文学. E-mail: hongwx@ysu.edu.cn.

间<sup>[4]</sup>等理论,这些理论虽然只有 30 多年甚至更短的时间,但是显示出其强大的生命力,并有相互融合,共同快速发展的趋势<sup>[5]</sup>。

形式概念分析(formal concept analysis, FCA)理论由德国数学家 Will.R 教授于 1982 年提出。FCA 以形式背景为研究对象,通过研究对象和属性之间的二元关系,发现形式背景中蕴含的知识和规则,并通过 Hasse 图的形式形象地表示出来。因在知识发现、规则提取等领域的优势,使得 FCA 理论在知识工程、信息检索等领域有效地被利用<sup>[6]</sup>。

洪文学等<sup>[7-8]</sup>在形式概念分析的基础上,以人类认知机理为出发点,通过对形式背景中属性的属性、对象的对象的研究,提出了偏序结构理论,创新性地提出异于 Hasse 图的形式背景知识可视化表示方法——偏序结构图。偏序结构图与 Hasse 图相比,其突出优势表现在支路之间相互无交叉,可以有效解决 Hasse 图在对象和属性关系复杂时可视化效果变差,失去可视化知识发现利用价值的问题。偏序结构理论已在中医诊疗知识发现<sup>[9]</sup>、英语语言学<sup>[10]</sup>等领域得以应用。

本文在将决策数据转化为决策模式的基础上,将偏序结构理论的思想应用于决策模式信息表,提出一种完全不依赖先验知识、以数据(决策模式信息表)为驱动的决策知识发现和规则提取可视化新方法——属性偏序决策图(decision diagram of attribute partial ordered structure, DDA-POS),介绍其生成原理和算法,通过实例分析其使用方法、特点及有效性。

1 基础定义

**定义 1** 形式背景。称三元组  $(U, A, I)$  为一个形式背景。其中  $U$  是有限非空对象集合,每个  $x_i \in U$  为形式背景的一个对象; $A$  为有限非空属性集合,每个  $a_i \in A$  为形式背景的一个属性; $I$  表示  $U$  与  $A$  之间的二元关系,  $I \subseteq U \times A$ 。

**定义 2** 决策信息表。称五元组  $(U, C, I, D, J)$  为一个决策信息表。其中  $U$  是有限非空对象集合,即论域,  $C$  为有限非空条件属性集合,  $D$  为决策属性,  $I$  表示论域  $U$  与条件属性  $C$  之间的映射关系,  $J$  表示论域  $U$  与决策属性  $D$  之间的映射关系。其中三元组  $(U, C, I)$  为一形式背景。表 1 所示为决策信息表的例子。

表 1 决策信息表  
Table 1 An example of decision information table

$U$	$C$						$D$
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	
$x_1$	1	0	1	0	0	1	$d_1$
$x_2$	1	0	1	0	1	0	$d_1$
$x_3$	0	1	1	0	0	1	$d_2$
$x_4$	1	0	0	1	0	0	$d_2$
$x_5$	0	1	1	0	1	0	$d_3$
$x_6$	0	1	0	1	1	0	$d_2$
$x_7$	1	0	0	1	0	1	$d_2$
$x_8$	1	0	0	1	0	1	$d_2$

在决策信息表中,如存在若干个对象其对应条件属性与决策属性完全相同,称其为同模式对象,同模式对象的个数,称为该模式的度。如表 1 中,  $x_7$  与  $x_8$  即为一组同模式对象,决策信息表中的一个模式记为  $m$ ,所有的模式构成模式集  $M$ 。

**定义 3** 决策模式信息表。决策信息表  $(U, C, I, D, J)$  对应的六元组  $(M, C, I', D, J', \text{De})$  称为决策模式信息表。其中  $M$  是模式集合,  $C$  为条件属性集合,  $D$  为决策属性,  $\text{De}$  为模式的度。  $I'$  表示模式集合  $M$  与条件属性  $C$  之间的映射关系,  $J'$  表示模式集合  $M$  与决策属性  $D$  之间的映射关系。表 1 所对应的决策模式信息表如表 2 所示。

表 2 表 1 对应的决策模式信息表  
Table 2 Decision pattern information table of table 1

$U$	$C$						$D$	$\text{De}$
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$		
$m_1$	1	0	1	0	0	1	$d_1$	1
$m_2$	1	0	1	0	1	0	$d_1$	1
$m_3$	0	1	1	0	0	1	$d_2$	1
$m_4$	1	0	0	1	0	0	$d_2$	1
$m_5$	0	1	1	0	1	0	$d_3$	1
$m_6$	0	1	0	1	1	0	$d_2$	1
$m_7$	1	0	0	1	0	1	$d_2$	2

在决策模式信息表中,若任意两个模式条件属性完全相同时,决策属性亦相同,则此决策模式信息表称决策模式一致信息表。反之,此决策信息表称为决策模式不一致信息表。在决策模式不一致信息表中,条件属性相同的模式是无法区分的,因此需要对决策不一致模式的结果按某种规则进行判别而转化为决策模式一致信息表。

**定义 4** 最大共有条件属性。在决策模式信息表  $(M, C, I', D, J', \text{De})$  所对应的形式背景  $(M, C, I')$  中,如果条件属性  $c \in C$  满足  $g(c) = U$ ,则称属性  $c$  是形式背景  $(M, C, I')$  的最大共有条件属性。

**定义 5** 共有条件属性。在决策模式信息表  $(M, C, I', D, J', De)$  所对应的形式背景  $(M, C, I'$  中,如果属性  $c_0, c_1, \dots, c_k \in C$  满足  $g(c_i) \subseteq g(c_0)$ , 其中  $i = 1, 2, \dots, k, k \geq 2$ , 则称在形式背景  $(M, C, I')$  中,条件属性  $c_0$  是条件属性集合  $\{c_1, c_2, \dots, c_k\}$  的共有条件属性。

2 属性偏序决策图

2.1 属性偏序决策图生成原理

由认知心理学可知:在一定情景和经验的背景下,人类是通过客观事物具有的属性特征来认知事物的。基本的认知原理可以描述为:特征相同靠近相同,特征不同远离相同,在相同特征中找出不同,在不同特征中找出相同。这样的描述蕴含着:相似性原理(相同靠近相同),相异性原理(不同远离相同),自上而下感知原理(相同之中找不同),自下而上感知原理(不同之处找相同)。图 1 给出了一般化认知机理模型示意图<sup>[11]</sup>。

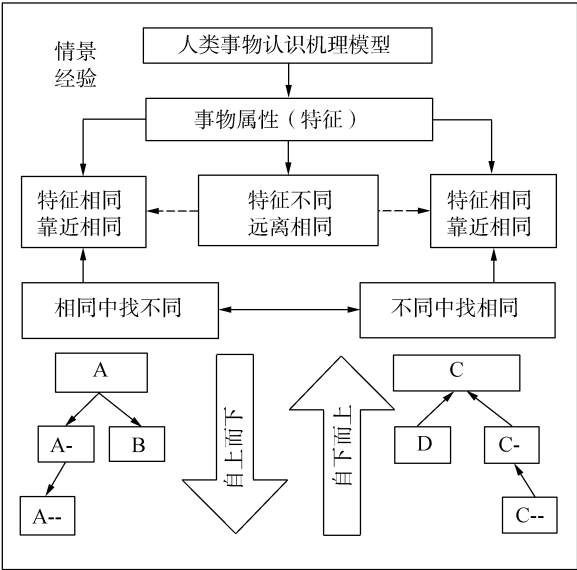


图 1 一般化认知机理模型示意图  
Fig.1 Schematic diagram of general cognitive mechanism model

属性是各类事物特征的表达,属性间的关系表达了研究问题的概念之间的关系。共有属性表达的是一定是事物普遍存在的现象,是共性的表达,具有较大的外延和较浅的内涵。从人类认识模式知识的角度来看,可以构造出以属性特征和对象相似性为指标的层次结构图。

2.2 属性偏序决策图生成算法

根据属性偏序决策图的生成原理,构建属性偏序决策图的算法描述如下所示。

构建属性偏序决策图

- 输入** 决策一致模式信息表  $(M, C, I', D, J', De)$ ;
- 输出** 属性偏序决策图  $(N, E, R)$ ;
- 1) 提取决策模式一致信息表  $(M, C, I', D, J', De)$  中的  $(M, C, I')$  构成形式背景;
  - 2) 生成属性偏序结构图<sup>[12]</sup>;
  - 3) 根据属性偏序结构图每一分支对应的模式标注其决策属性,形成决策层;
  - 4) 从根节点由上至下检索各支路,当检索至某节点覆盖之下各支路模式完全一致时,停止检索此节点之下的支路和节点,并将此节点下各支路上的节点标记为隐藏节点;
  - 5) 直至检索完所有支路,所得结果即为属性偏序决策图。

表 2 所示决策模式信息表对应的属性偏序决策图如图 2 所示。由图可见,属性偏序决策图在结构上分为决策规则表示层和决策规则输出层两个层次。图中虚线表示的节点为隐藏节点,不参与知识发现和规则提取过程。决策表示层是一个典型的树结构,其中每一个节点表示一个条件属性,图中最顶层为最大共有条件属性。树中根据条件属性的偏序关系由上而下形成若干个节点和边构成的支路,其中每一条支路即为一种决策模式。根据属性偏序决策图中支路聚类形成的群结构、子群结构及每条支路对应的决策属性即可完成决策规则的输出。

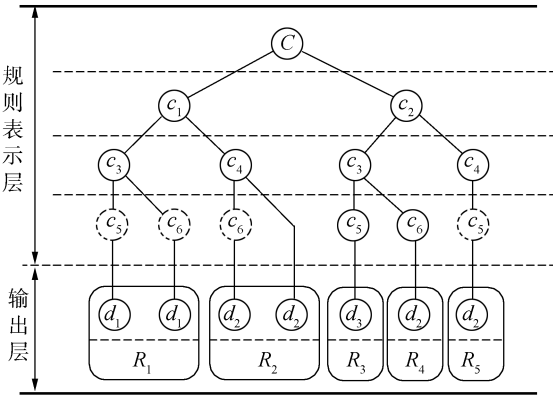


图 2 表 2 所示决策模式信息表对应属性偏序决策图  
Fig.2 DDAPOS of table2

2.3 基于属性偏序决策图的规则提取方法

基于属性偏序决策图自顶向下的构图原则,条件属性中普遍性强在上层,可以从集群结构、集子群结构、支路和节点等不同角度对原始数据进行知识发现。支路条件属性集合表达的一条决策模式,即一个决策模式具有什么属性。图中同一子群结构表达的为模式间相似性;不同子群结构表达决策模式之间的差异性。综合属性偏序决策图条件属性偏序关系及决策属性,即可得出最终的规则提取。图 2 所示的属

性偏序决策图决策结果输出可表示为表 3。

表 3 图 2 所示属性偏序决策图决策结果输出  
Table 3 Decision output of DDAPOS of Fig.2

序号	决策输出结果
$R_1$	IF (( $c_1$ ) AND ( $c_3$ )) THEN ( $D_{\text{result}} = d_1$ );
$R_2$	IF (( $c_1$ ) AND ( $c_4$ )) THEN ( $D_{\text{result}} = d_2$ );
$R_3$	IF (( $c_2$ ) AND ( $c_3$ ) AND ( $c_5$ )) THEN ( $D_{\text{result}} = d_3$ );
$R_4$	IF (( $c_2$ ) AND ( $c_3$ ) AND ( $c_6$ )) THEN ( $D_{\text{result}} = d_2$ );
$R_5$	IF (( $c_2$ ) AND ( $c_4$ ) AND ( $c_5$ )) THEN ( $D_{\text{result}} = d_2$ );

3 实例实验结果与分析

下面通过一个医学诊断的实例介绍属性偏序决策图的具体使用方法,并验证其规则提取的有效性。

3.1 实验数据

选用 UCI 数据库的 Breast Cancer Wisconsin Data Set(BCWD)数据集验证属性偏序决策图的有效性,数据集属性组成如表 4 所示。该数据集共有 699 个样本(其中有 16 个样本存在属性缺失),9 个属性,2 个类(良性、恶性)。本实验剔除了这些有属性缺失的样本,因此实际使用了数据集中样本 683 个。实验采用 10 折交叉验证的方式对所述方法的可行性进行验证,并与主流的模式分类算法进行了对比分析。

表 4 乳腺癌数据集属性构成  
Table 4 Attributes of BCWD

代码	属性	取值范围
CT	Clump Thickness	1~10
$U_{\text{csi}}$	Uniformity of Cell Size	1~10
$U_{\text{csh}}$	Uniformity of Cell Shape	1~10
MA	Marginal Adhesion	1~10
SECS	Single Epithelial Cell Size	1~10
BN	Bare Nuclei	1~10
BC	Bland Chromatin	1~10
NN	Normal Nucleoli	1~10
$M$	Mitoses	1~10

3.2 实验过程

1)原始数据特征(属性)选择。根据原始数据分布,利用 lasso 算法完成特征子集选择,特征相关序列为:BN、 $U_{\text{csh}}$ 、 $U_{\text{csi}}$ 、CT、BC、NN、MA、SECS、 $M^{[13]}$ 。本例选取其中 3 个属性:BN、 $U_{\text{csh}}$ 、 $U_{\text{csi}}$ 参与决策生成。

2)原始数据粒化,生成决策形式信息表。以数据驱动的决策规则生成中,数据的形式可能是数值型、名义型、布尔型、区间型等。所有的原始数据需根据不同粒化规则,转换为布尔型规则。根据原始数据分布,按表 5 规则完成数据粒化生成决策信息表如表 6 所示。

表 5 乳腺癌数据集粒化规则  
Table 5 Graining rules of BCWD

序号	粒化规则
1	IF ( $U_{\text{csi}} \leq 2$ ) THEN ( $c_1 = 1, c_2 = 0$ )
2	IF ( $U_{\text{csi}} > 2$ ) THEN ( $c_1 = 0, c_2 = 1$ )
3	IF ( $U_{\text{csh}} \leq 3$ ) THEN ( $c_3 = 1, c_4 = 0$ )
4	IF ( $U_{\text{csh}} > 3$ ) THEN ( $c_3 = 0, c_4 = 1$ )
5	IF ( BN $\leq 2$ ) THEN ( $c_5 = 1, c_6 = 0$ )
6	IF ( BN $> 2$ ) THEN ( $c_5 = 0, c_6 = 1$ )

表 6 乳腺癌数据决策信息表  
Table 6 Decision information table of BCWD

编号	原始数据			条件属性 $C$						$D$
	$U_{\text{csi}}$	$U_{\text{csh}}$	BN	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	
1190394	1	1	3	1	0	1	0	0	1	N
1272039	1	1	1	1	0	1	0	1	0	N
1173235	3	2	3	0	1	1	0	0	1	N
1133136	1	1	3	1	0	1	0	0	1	N
1207986	8	4	8	0	1	0	1	0	1	P
837480	4	4	10	0	1	0	1	0	1	P
1049837	1	1	1	1	0	1	0	1	0	N
...	...	...	...	...	...	...	...	...	...	...

3)根据决策形式信息表生成决策模式信息表。检索决策信息表中的模式,得到乳腺癌数据决策模式信息表如表 7 所示。

表 7 乳腺癌数据决策模式信息表  
Table 7 Decision pattern information table of BCWD

$U$	$C$						$D$	De
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$		
$m_1$	1	0	1	0	0	1	N	20
$m_2$	1	0	1	0	0	1	P	7
$m_3$	1	0	1	0	1	0	N	305
$m_4$	0	1	1	0	0	1	P	17
$m_5$	0	1	1	0	0	1	N	3
$m_6$	1	0	0	1	1	0	N	5
$m_7$	1	0	0	1	1	0	P	1
$m_8$	0	1	1	0	1	0	N	19
$m_9$	0	1	1	0	1	0	P	2
$m_{10}$	0	1	0	1	1	0	P	17
$m_{11}$	0	1	0	1	1	0	N	3
$m_{12}$	0	1	0	1	0	1	P	140
$m_{13}$	0	1	0	1	0	1	N	5
$m_{14}$	1	0	0	1	0	1	P	2

表 8 为决策模式不一致信息表,对条件属性相同但决策属性不同的模式中,按照模式度较小的模式的决策属性服从模式度较大模式的原则进行修



正,得到表 8 所对应的决策模式一致信息表如表 8。

表 8 乳腺癌数据决策模式一致信息表

Table 8 Consistent decision information table of BCWD

U	C						D	De
	c <sub>1</sub>	c <sub>2</sub>	c <sub>3</sub>	c <sub>4</sub>	c <sub>5</sub>	c <sub>6</sub>		
m <sub>1</sub>	1	0	1	0	0	1	N	27
m <sub>2</sub>	1	0	1	0	1	0	N	305
m <sub>3</sub>	0	1	1	0	0	1	P	20
m <sub>4</sub>	1	0	0	1	1	0	N	6
m <sub>5</sub>	0	1	1	0	1	0	N	21
m <sub>6</sub>	0	1	0	1	1	0	P	20
m <sub>7</sub>	0	1	0	1	0	1	P	125
m <sub>8</sub>	1	0	0	1	0	1	P	2

4)生成属性偏序决策图。根据表 8 决策一致模式信息表生成属性偏序决策图如图 3 所示。

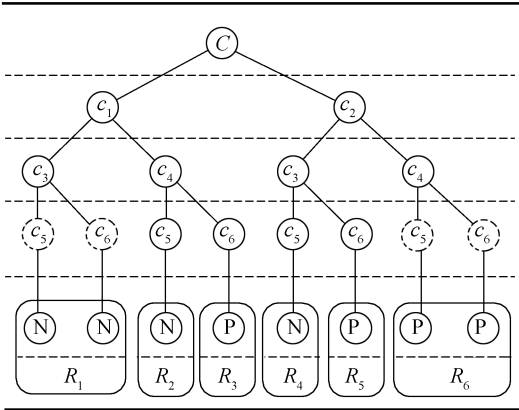


图 3 乳腺癌诊断属性偏序决策图

Fig.3 DDAPOS of BCWD

5)根据属性偏序决策图,完成决策规则提取任务。由图 3 所示属性偏序决策图提取的规则如表 9 所示。

3.3 结果分析

应用上述诊断规则对数据集进行诊断,同时与主流模式分类算法(kNN、Naive Bayes、SVM、C5.0、Random Forests)对比结果见表 10。通过对比可以发现,本文所述属性偏序决策图在只有 3 个属性参与规则提取的条件下,提取的诊断规则在各项指标上均较理想,达到了主流模式分类和知识发现方法的水平。通过增加参与规则提取的属性,改善粒化规则等措施,各项指标仍有提高的空间。与常规模式识别方法(如 kNN、SVM 等)相比,属性偏序决策图可以将决策规则以图形的方式进行明确地表示,这一特性可以有效地沟通领域专家与数据分析专家,降低其在具体领域的应用门槛。

表 9 乳腺癌诊断规则

Table 9 Decision output of DDAPOS

序号	决策输出结果
R <sub>1</sub>	IF (( U <sub>csi</sub> ≤ 2 ) AND ( U <sub>esh</sub> ≤ 3 )) THEN ( D <sub>result</sub> = N );
R <sub>2</sub>	IF (( U <sub>csi</sub> ≤ 2 ) AND ( U <sub>esh</sub> > 3 ) AND ( BN ≤ 2 )) THEN ( D <sub>result</sub> = N );
R <sub>3</sub>	IF (( U <sub>csi</sub> ≤ 2 ) AND ( U <sub>esh</sub> > 3 ) AND ( BN > 2 )) THEN ( D <sub>result</sub> = P );
R <sub>4</sub>	IF (( U <sub>csi</sub> > 2 ) AND ( U <sub>esh</sub> ≤ 3 ) AND ( BN ≤ 2 )) THEN ( D <sub>result</sub> = N );
R <sub>5</sub>	IF (( U <sub>csi</sub> > 2 ) AND ( U <sub>esh</sub> ≤ 3 ) AND ( BN > 2 )) THEN ( D <sub>result</sub> = P );
R <sub>6</sub>	IF (( U <sub>csi</sub> > 2 ) AND ( U <sub>esh</sub> > 3 )) THEN ( D <sub>result</sub> = P )。

表 10 乳腺癌诊断灵敏度、特异度和准确率

Table 10 Sensitivity, specificity and accuracy of BCWD

方法	准确率	灵敏度	特异度
kNN	0.964 8	0.941 7	0.977 5
Naive Bayes	0.961 9	0.979 2	0.952 7
SVM	0.970 7	0.975 0	0.968 5
C5.0	0.948 7	0.928 8	0.959 5
Random Forests	0.972 1	0.970 8	0.972 9
本文方法	0.962 3	0.928 6	0.941 6

4 结束语

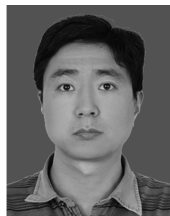
本文以决策模式信息表为研究对象,提出一种基于属性偏序关系的、不依赖于先验知识完全以数据驱动的规则提取可视化方法——属性偏序决策图。属性偏序决策图通过属性的聚类完成事物“类内紧,类间松”的聚类,并以直观图形的形式进行表示,从中发现事物之间相区别的属性,从而达到提取事物共同特征的目的。属性偏序决策图因不需要进行复杂的浮点运算,其算法运算速度快。另外,该方法在实际应用中还有若干问题有待解决,如如何根据原始属性生成决策模式信息表、如何处理决策过程中的不确定性问题、如何提供人机交互方式完成专家知识与机器学习的融合等,都有待进一步研究。

参考文献:

[1] POELMANS J, KUZNETSOV S O, IGNATOV D I, et al. Formal concept Analysis in knowledge processing: a survey on models and techniques[J]. Expert systems with applications, 2013, 40(16): 6601-6623.

- [2] 王国胤, 姚一豫, 于洪. 粗糙集理论与应用研究综述[J]. 计算机学报, 2009, 32(7): 1229-1246.  
WANG Guoyin, YAO Yiyu, YU Hong. A survey on rough set theory and applications[J]. Chinese journal of computers, 2009, 32(7): 1229-1246.
- [3] YAO Yiyu. A comparative study of fuzzy sets and rough sets[J]. Information sciences, 1998, 109(1/2/3/4): 227-242.
- [4] ZHANG Ling, ZHANG Bo. The quotient space theory of problem solving[J]. Fundamenta informaticae, 2004, 59(2/3): 287-298.
- [5] 于洪, 王国胤, 姚一豫. 决策粗糙集理论研究现状与展望[J]. 计算机学报, 2015, 38(8): 1628-1639.  
YU Hong, WANG Guoyin, YAO Yiyu. Current research and future perspectives on decision-theoretic rough sets[J]. Chinese journal of computers, 2015, 38(8): 1628-1639.
- [6] POELMANS J, IGNATOV D I, KUZNETSOV S O, et al. Formal concept analysis in knowledge processing: a survey on applications[J]. Expert systems with applications, 2013, 40(16): 6538-6560.
- [7] HONG Wenxue, LI Shaoxiong, YU Jianping, et al. A new approach of generation of structural partial-ordered attribute diagram[J]. ICIC express letters, part B: applications, 2012, 3(4): 823-830.
- [8] 洪文学, 栾景民, 张涛, 等. 基于偏序结构理论的知识发现方法[J]. 燕山大学学报, 2014, 38(5): 394-402.  
HONG Wenxue, LUAN Jingmin, ZHANG Tao, et al. A new method for knowledge discovery based on partial ordered structure theory[J]. Journal of Yanshan university, 2014, 38(5): 394-402.
- [9] 刘超男, 徐笋晶, 李赛美, 等. 基于多层次复杂概念网络表示方法的《伤寒论》方药按治法分类的知识发现[J]. 北京中医药大学学报, 2014, 37(7): 452-457.  
LIU Chaonan, XU Sunjing, LI Saimei, et al. Knowledge discovery of formula classification according to therapies in Shanghanlun based on representation method of multi-layer complex concept network[J]. Journal of Beijing university of traditional Chinese medicine, 2014, 37(7): 452-457.
- [10] YU Jianping, LI Chen, HONG Wenxue, et al. A new approach of rules extraction for word sense disambiguation by features of attributes[J]. Applied soft computing, 2015, 27: 411-419.
- [11] GALOTTI K M. 认知心理学[M]. 吴国宏, 译. 3 版. 西安: 陕西师范大学出版社, 2005.  
GALOTTI K M. Cognitive Psychology[M]. WU Guohong, Trans. 3rd ed. Xi'an: Shaanxi Normal University Press, 2005.
- [12] 李少雄, 闫恩亮, 宋佳霖, 等. 偏序结构图的一种计算机生成算法[J]. 燕山大学学报, 2014, 38(5): 403-408.  
LI Shaoxiong, YAN Enliang, SONG Jialin, et al. Computational generation algorithm of partial ordered structure diagram[J]. Journal of Yanshan university, 2014, 38(5): 403-408.
- [13] TIBSHIRANI R. Regression shrinkage and selection via the lasso: a retrospective[J]. Journal of the royal statistical society, 2011, 73(3): 273-282.

#### 作者简介:

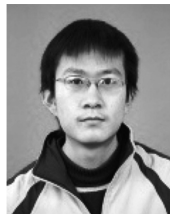


郑存芳,男,1979 年生,讲师,博士研究生,CAAI 粗糙集与软计算专委会会员、CCF 会员,主要研究方向为可视化模式识别、偏序结构理论、中医工程学等。先后参与国家自然科学基金 3 项、河北省自然科学基金 3 项。



洪文学,男,1953 年生,教授,博士生导师,燕山大学生物医学工程研究所所长,CAAI 粗糙集与软计算专委会委员,主要研究方向为大数据偏序结构理论、复杂概念网络、混合数据信息融合与模式识别和中医工程学。所带领的

学术团队近年来主持省部级科研项目 20 余项。



李少雄,男,1987 年生,博士研究生,主要研究方向为偏序结构理论、可视化模式识别、中医工程学等。参与国家自然科学基金项目 4 项。