

一种新的关联规则 Vis-Meta 图知识表示方法

陈敏^{1,2,3} 赵书良^{1,2,3} 郭晓波¹ 李晓超^{1,2,3} 柳萌萌^{1,2,3}

(河北师范大学数学与信息科学学院 石家庄 050024)¹

(河北省计算数学与应用重点实验室 石家庄 050024)² (河北师范大学移动物联网研究院 石家庄 050024)³

摘要 针对传统关联规则表示方法无法展示领域知识,缺少对规则信息一对一、一对多、多对一、多对多的多模式表示,忽略知识发现结果的共享等问题,提出了一种新的基于 Vis-Meta 图的多模式关联规则知识表示方法。首先给出了 Vis-Meta 图的相关定义与关联规则的 Vis-Meta 图表示方法,接着定义了关联规则 Vis-Meta 图知识表示中的概念关系,并在此基础上给出了关联规则概念关系知识表示算法、关联规则实例对比算法和关联规则知识表示优化算法。最后,以某省全员人口数据为基础,对关联规则信息进行可视化分析。实验结果表明,所提出的知识表示算法具有良好的展示效果与知识共享能力。

关键词 Meta 图,关联规则,知识表示,可视化

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.3.045

New Vis-Meta Graph Knowledge Representation for Association Rules

CHEN Min^{1,2,3} ZHAO Shu-liang^{1,2,3} GUO Xiao-bo¹ LI Xiao-chao^{1,2,3} LIU Meng-meng^{1,2,3}

(Mathematics and Information Science College, Hebei Normal University, Shijiazhuang 050024, China)¹

(Hebei Key Laboratory of Computational Mathematics and Applications, Shijiazhuang 050024, China)²

(Institute of Mobile Internet of Things, Hebei Normal University, Shijiazhuang 050024, China)³

Abstract Considering the problems aroused by the traditional association rules presentation formalizing approaches which are powerless to demonstrate the domain knowledge, lack of displaying multi-schema association rules of one to one, one to many, many to one, many-to-many, and especially ignoring the sharing knowledge of discovering results, this paper proposed a novel knowledge representation method for showing multi-mode association rules based on Vis-Meta graph. Firstly, it gave the relevant definitions of Vis-Meta graph and Vis-Meta graph presentation method of association rules, then introduced the conceptual relationship in Vis-Meta graph for knowledge representation, and presented association rule's conceptual relationship knowledge representation algorithm, association rule's instance compared algorithm, as well as association rule's knowledge representation optimizing algorithm. Finally, with the help of experimental data obtained from demographic data of a province, we finished the visualizing analysis for association rules information. Experimental results turn out that the knowledge representation algorithm proposed has better display effect and knowledge-sharing.

Keywords Meta graph, Association rules, Knowledge representation, Visualization

1 引言

关联规则知识表示已经成为数据挖掘领域中一个新的研究重点,其主要目的是将从数据集中发现属性间存在的相关关系进行知识表示,帮助用户从海量数据中获取更为精确的信息。然而,一般方法无法有效地对已经挖掘出的关联模式予以数据可视化。元图是图论领域的新兴研究课题,利用人类很强的视觉处理能力,对信息进行可视化展示,能够提高可

视化过程的效率^[1]。Gaur D等^[2]利用决策元图展示集合到集合之间的路径,发现其他图展示中不易发现的规则;若从初始节点到目标节点有元路径,则初始节点能够通过路径上的其他节点(决策条件)做出目的节点的决策。同时,Gaur D等^[3]在数据集中进行基于元图的频繁模式挖掘,应用一个新的词典顺序,将每个子元图映射到一个唯一的最小深度优先搜索编码,将其作为该子元图的标准标签,提出一个不用产生候选项集(candidate generation)的挖掘算法,适用于深度优先

收稿日期:2014-04-24 返修日期:2014-08-05 本文受国家自然科学基金项目(71271067),国家社会科学基金项目(13BTY011),国家社科基金重大项目(13&·ZD091),河北师范大学数学与信息科学学院硕士基金(201301)资助。

陈敏(1988-),女,硕士生,CCF 学生会员,主要研究方向为数据挖掘、智能信息处理,E-mail:cm316029906@126.com;赵书良(1967-),男,博士,教授,博士生导师,主要研究方向为智能信息处理,E-mail:zhaoshuliang@sina.com(通信作者);郭晓波(1986-),男,硕士生,主要研究方向为数据挖掘;李晓超(1986-),男,硕士生,主要研究方向为数据挖掘、智能信息处理;柳萌萌(1988-),女,硕士生,主要研究方向为数据挖掘、智能信息处理。

策略来有效挖掘频繁连通子元图,但该方法不适合对数据间的关联模式进行信息可视化表示。Hu Z等^[4]利用元图的思想来可视化细胞网络,通过语义缩放展示不同尺度的信息,展示细胞周期中蛋白质复合物的角色,以层次形式展示新陈代谢分子,整合实验交互数据与结构化词汇(如:基因本体类别)来展示同源基因组间的持久交互。Dashore P等^[5]利用模糊元图表示股票市场的3个阶段:buy, hold和sell;在基于模糊规则的系统的基础上,股票交易数据可以用于克服股票市场的不确定性、模糊性和不精确性,但是该方法的仿真结果并未对股票市场进行元图展示。另外, Dashore P等^[6]结合元图实现了一种基于规则的电子商务系统,该系统能够在复杂的形式和不精确的信息情况下,通过规则集合的形式获得知识来帮助顾客指定合适的决策。Mukherjee A等^[7]利用元图表示业务流程的分析和验证,表示元图具有良好的信息数据表示效果。郭晓波等^[8]提出一种新的基于概念图的关联规则知识表示方法,该方法将关联规则所涉及的概念属性信息及相互关系转换成概念结点和关系节点,并能够准确、简洁且充分地展示规则信息,但展示形式无规律,不便于降低认知负担。目前,一些研究人员利用元图实现规则管理^[9]、文档搜索引擎^[10]、基于规则的专家系统^[11]、基于模糊元图的知识表示^[12]等。然而,海内外研究学者研究的主要内容大都是利用元图理论进行数据处理与分析,在关联规则知识表示方面主要集中在关联规则的一般可视化方法^[13-14],大都存在以下不足:对领域知识展示不足;数据展示结果的可解释性较差;缺少对规则信息的多模式表示;尚未对最终结构进行知识组织,不便于进行知识共享。

针对上述问题,本文提出一种新的基于Vis-Meta图的多模式关联规则知识表示方法,结合元图理论给出了关联规则的Vis-Meta图表示方法,实现了一对一、一对多、多对一、多对多形式的关联规则表示。本文进一步在Vis-Meta图关联规则表示方法中加入概念关系,对关联规则进行有效合理的组织,以层次结构表现数据关系和规则关系,以水平结构表现规则对比关系,便于用户将规则映射为自然语言,实现关联规则知识共享,并从中发现新的隐含的知识。最后,本文给出了关联规则的Vis-Meta图知识表示算法,并以某省全员人口数据库为基础对算法进行实现和验证。实验表明,所提出方法具有良好的规则知识表示效果,能够帮助用户获取更精确的信息。

2 基于Vis-Meta图的关联规则表示方法

2.1 Vis-Meta图

定义1(Meta图) Meta图 S 是定义在生成集 $X = \{x_i, i=1, 2, \dots, N\}$ 上的一个序对,即 $S = \langle X, E \rangle$ 。其中 $E = \{e_j, j=1, 2, \dots, N\}$ 表示Meta图 S 中所有边组成的集合,对于任意的 $e_j \in E$ 是由生成集 X 的两子集 $V_j, W_j \subseteq X$ (其中 $V_j \cup W_j \neq \emptyset$)组成的序对,记 $e_j = \langle V_j, W_j \rangle$,分别称 V_j, W_j 为 e_j 的入点集(Invertex)和出点集(Outvertex)。

为了更好地实现基于Vis-Meta图的关联规则表示任务,本文对Meta图进行了改进,提出Vis-Meta图的定义。

定义2(Vis-Meta图) Vis-Meta图 S 是定义在生成集 $X = \{x_i, i=1, 2, \dots, I\}$ 上的一个三元组即 $S = \langle X, E, W \rangle$ 。其中 $E = \{e_k, k=1, 2, \dots, K\}$ 为边集,记 $e_k = \langle V_k, W_k \rangle$ 由 X 的子集作

为端点集,满足 $V_k, W_k \subseteq X$ 且 $V_k \cup W_k \neq \emptyset, V_k \cap W_k = \emptyset$,称 V_k, W_k 分别为 e_k 的入点集和出点集。 $W = \{i_1, i_2, i_3, \dots, i_n\}, i_r = \{Supp_r, Conf_r\}, 1 \leq r \leq n, n$ 是关联规则的数目, $Supp_r$ 表示第 r 个关联规则的支持度, $Conf_r$ 表示第 r 个关联规则的置信度。

2.2 关联规则的Vis-Meta图表示

Vis-Meta图是表示规则前件集合、规则后件集合以及相关信息的三元关系,为实现多模式的关联规则表示提供了有利条件。下面给出了Vis-Meta图关联规则表示的相关定义和图形符号表示。

定义3(关联规则挖掘背景) 关联规则挖掘背景是一个三元组 $D = \langle T, K, R \rangle$ 。 T 代表数据集中的事务集合, K 代表项集合, $R \subseteq T \times K$ 是事务和项间的二元关系。每个二元组 $(t, k) \in R$ 表示 $t \in T$ 和 $k \in K$ 之间的一个关系实例。

定义4(Meta属性) 已知项集合 K 的子集 $l = \{k_1, \dots, k_l\} \subseteq K$ 且 $k_i \in l$ 属于同一属性,则称该属性为Meta属性。给定项集合 $K = \{k_i, i=1, 2, \dots, H\}$,定义项集合 K 上的Meta属性 A ,若 $k_1, k_2, \dots, k_i \in K$ 且属于同一属性,则记Meta属性为 $A = k_1 \Delta k_2 \Delta \dots \Delta k_i$ 。

定义5(Meta值) Meta值 k_i 是关联规则的前件或后件中的某一属性值,是项集合 K 中的一个元素。

定义6(Meta边) Meta边 F 是连接一条关联规则前件和后件的边的集合。

假设某条关联规则的前件中包含 M 个Meta属性,后件中包含 N 个Meta属性,连接前件属性的 M 条边与连接后件属性的 N 条边交于一点 P ,则该点保存此关联规则的支持度 $Supp$ 和置信度 $Conf$ 。

定义7(Meta规则) 关联规则的Vis-Meta图知识表示中,由Meta值、Meta属性、Meta边表示的一条关联规则称为Meta规则。根据模式的不同,分为一对一规则、一对多规则、多对一规则、多对多规则。

给定项集合 $K = \{k_i, i=1, 2, \dots, H\}$, K 的幂集为 $P(K)$,假设存在这样的映射 $\phi: P_1(K) \rightarrow P_2(K)$ 表示关联规则前件与后件的关系。若对于任意的 $\{k_{i1}, k_{i2}, \dots, k_{if}\} \in P_1(K), \{k_{j1}, k_{j2}, \dots, k_{jn}\} \in P_2(K)$,且 $\phi(\{k_{i1}, k_{i2}, \dots, k_{if}\}) = \{k_{j1}, k_{j2}, \dots, k_{jn}\}$,其中 $\{k_{i1} \in A_{i1}, k_{i2} \in A_{i2}, \dots, k_{if} \in A_{if}\} \in P_1(K), \{k_{j1} \in A_{j1}, k_{j2} \in A_{j2}, \dots, k_{jn} \in A_{jn}\} \in P_2(K)$,并且 $A_{i1} \cap A_{i2} \cap \dots \cap A_{if} \cap A_{j1} \cap A_{j2} \cap \dots \cap A_{jn} = \emptyset, A_{i1}, \dots, A_{jn}$ 为Meta属性,则称 $\{k_{i1}, k_{i2}, \dots, k_{if}\}$ 为规则前件, $\{k_{j1}, k_{j2}, \dots, k_{jn}\}$ 为规则后件。

基于Vis-Meta图的关联规则图形符号表示如表1所列。

表1 关联规则Vis-Meta图图形符号表示

概念	符号表示
Meta属性	
Meta值	
集合	
Meta规则 (多对多关联规则)	
关系	
对象	

3 关联规则的 Vis-Meta 图知识表示方法

人类进行一切认知活动都离不开概念,概念既作为知识内容存在于人类的大脑中,又作为知识目标引导大脑进行知识构建。因此要实现真正意义上的关联规则知识表示,需要考虑概念在感知觉、问题求解、推理等认知活动过程中的重要作用。本文进一步在 Vis-Meta 图关联规则表示方法中加入概念关系,对关联规则进行有效合理的组织,以层次结构表现数据关系和规则关系,以水平结构表现对比关系,表示结果的对称性与层次性便于用户将规则映射为自然语言,实现关联规则知识共享,获得知识中的知识。下面给出了基于 Vis-Meta 图的关联规则知识表示方法。

定义 8 (Meta 属性的内涵和外延) Meta 属性中所有 Meta 值的共同性质为该 Meta 属性的内涵,用该 Meta 属性的名称表示;Meta 属性中的所有 Meta 值为该 Meta 属性的外延。我们将 Meta 属性的内涵与外延构成数据关系用 $R(C, Set(Meta_value))$ 表示,其中 C 表示概念(Meta 属性)的内涵, $Set(Meta_value)$ 表示概念(Meta 属性)的 Meta 值集合,即外延。Meta 属性“初婚年龄”的内涵和外延构成的数据关系如图 1 所示,箭头方向由内涵指向外延。

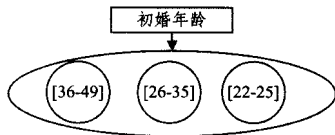


图 1 Meta 属性内涵和外延的数据关系表示

由于 Meta 边由规则前件概念外延 $Set(left)$ 指向规则后件概念外延 $Set(right)$,因此 Meta 边体现了一种规则关系 $R(Set(left), Set(right))$ 。关联规则“{文化程度,地区,户口性质} ==> {世代间隔} (Supp=20%, Conf=70%)”的规则关系表示如图 2 所示。

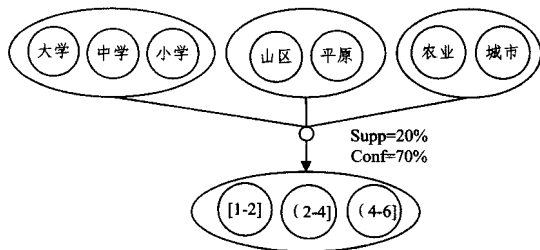


图 2 规则关系表示

定义 9 (对象概念) 对象概念 $OC \supset \{Meta_attr_1, Meta_attr_2, \dots, Meta_attr_n\}$ 是一个由多个 Meta 属性构成的实体,对象概念与 Meta 属性内涵之间构成数据关系 $R(OC, Set(Meta_attr))$ (见图 3)表示所研究的 Meta 属性{生育年龄,文化程度,地区,户口性质,间隔}来自于对象概念“育龄妇女”。

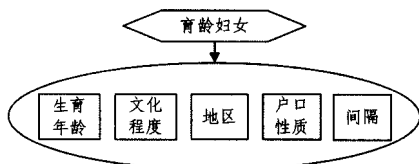


图 3 对象概念表示

为了便于用户进行规则实例对比,引入关系节点,表示两个对象概念实例之间的关系 $R(OC_Insx, OC_Insy)$ (见图 4)

表示“育龄妇女 1 是育龄妇女 2 的母亲”。



图 4 关系节点表示

4 算法实现

本文数据来源于某省全员人口数据库,为了实现人口关联规则知识表示,在 Vis-Meta 图中引入概念关系,提出关联规则概念关系知识表示算法、关联规则实例对比算法、关联规则知识表示优化算法。构建便于转换为自然语言的关联规则知识表示方法,并根据人口关联规则的特点,通过合并与优化算法,简化表示形式,便于用户进行规则实例对比,获取关联规则知识中的隐含知识,并最终能够用自然语言描述出来进行知识传递与共享。

算法思想:首先构建关联规则的 Vis-Meta 图的概念关系表示 $C_VizRule$, $C_VizRule$ 可能包含多个规则实例,接着可以通过合并操作对比两个具有相同 Meta 属性内涵的规则实例,并可以通过优化操作简化合并后的规则实例图或者简化前件后件具有相同 Meta 属性内涵的规则,最终可从中获得先前未知的知识。具体实现过程如下:

算法 1 RuleSetToMGs()//关联规则概念关系知识表示算法

输入:关联规则 Rule

输出:关联规则的 Vis-Meta 图概念关系表示 $C_VizRule$

RuleSetToMGs (Rule)

- 1) CreateDataRelation (OCLeft \cup OCRight, LeftSet (Meta_attr) \cup RightSet(Meta_attr));
- 2) FOREACH attr \in {LeftSet(Meta_attr) \cup RightSet(Meta_attr)} DO{//规则前件/后件中的每个 Meta 属性
- 3) CreateDataRelation(attr, Set(attr.value));
- 4) //end foreach
- 5) FOREACH leftattr \in LeftSet(Meta_attr) DO{ //每个前件 Meta 属性
- 6) CreateRuleRelation(Set(leftattr.value), Set(rightattr.value));
- 7) //end foreach
- 8) Return $C_VizRule$;

算法 1 第 1)步根据第 3 节的定义 9,构建规则前件(后件)对象概念 OCLeft(OCRight)与前件(后件)Meta 属性内涵的数据关系;第 2)至 4)步根据第 3 节定义 8,构建前件(后件)Meta 属性内涵与外延的数据关系;第 5)至 7)步根据第 3 节规则关系的定义,构建前件 Meta 属性的外延与后件 Meta 属性的外延之间的规则关系。

算法 2 CombineRule()//关联规则实例对比算法

输入:两个具有相同 Meta 属性内涵的规则实例 Instancex, Instancey

输出:实例的合并展示 Combine_Layout

CombineRule (Instancex, Instancey)

- 1) CreateInstanceOCRelation (Instancex.oc, RelationNode, Instancey.oc);
- 2) CreateDataRelation(Instancex.oc \cup Instancey.oc, LeftSet(Meta_attr));
- 3) CreateDataRelation(LeftSet(Meta_attr), LeftSet(attr.value));
- 4) CreateRuleRelation (Set (Instancex.left.value), Set (Instancex.right.value));
- 5) CreateRuleRelation (Set (Instancey.left.value), Set (Instancey.

```

right. value));
6) CreateDataRelation(Instancex. oc ∪ Instancey. oc, RightSet (Meta_
attr));
7) CreateDataRelation(Instancex. oc ∪ Instancey. oc, RightSet (Meta_
attr));
8) return Combine_Layout;

```

算法 2 第 1)步根据第 3 节的实例关系定义,引入关系节点 RelationNode 构建两个规则的对象概念实例的关系;第 2)步根据第 3 节的定义 9,构建两个对象概念实例与公共规则前件内涵的数据关系;第 3)步根据第 3 节的定义 8,构建前件内涵与外延间的数据关系;第 4)步和第 5)步根据第 3 节规则关系的定义,构建两个规则实例前件 Meta 属性的外延与后件 Meta 属性的外延之间的规则关系;第 6)步根据第 3 节的定义 8,构建后件内涵与外延间的数据关系;第 7)步根据第 3 节的定义 9,构建两个对象概念实例与公共规则后件内涵的数据关系。

算法 3 OptimizeLayout()//关联规则知识表示优化算法

输入:规则实例展示 K

输出:优化的展示 Optimize_Layout

OptimizeLayout (K)

```

1) IF K == Combine_Layout
   //如果 K 是两条规则实例的合并展示
2){
3) DeleteRightInstanceOCRelation (Instancex. oc, RelationNode, In-
   stancey. oc)
4) ReLinkToLeftInstance (R (Instancex. oc ∪ Instancey. oc, RightSet
   (Meta_attr)), LeftInstance. oc)
5)}
6) ELSE IF K == C_ VizRuleIns and LeftSet (Meta_attr) == Right-
   Set (Meta_attr)
7){
8) CreateInstanceOCRelation (OC. Left, RelationNode, OC. Right);
9) DeleteRuleRelationLink (LeftSet (Meta_attr). Set (left. value),
   RightSet (Meta_attr). Set (right. value));
10) DeleteDataRelationLink (RightSet (Meta_attr), Set (right. value));
11) CreateDataRelation (OC. Right, LeftSet (Meta_attr));
12) CreateDataRelation (RightSet (Meta_attr); LeftSet (Meta_
   attr), Set (right. value ); Set (left. value));

```

```
13) CreateLink (OC. Left, OC. Right, Supp, Conf)
```

```
14)}
```

```
15) return Optimize_Layout;
```

算法 3 第 1)至 5)步处理规则实例展示 K 是两条规则实例的合并展示的情况,第 3)步删掉后件对象概念实例关系的图形展示,以避免展示结果中存在重复的图形元素,第 4)步将后件 Meta 属性内涵与后件概念对象实例之间的数据关系重连接到前件概念实例。

算法 3 第 6)至 14)步处理规则展示 K 是一条规则前件内涵与规则后件内涵相同的实例的情况;第 8)步引入关系节点 RelationNode 构建前后件对象概念实例的关系;第 9)步删掉规则前件后件外延的规则关系连接线,但不删掉规则关系;第 10)步删掉后件 Meta 属性的内涵与外延的数据关系连接线;第 11)步构建规则后件对象概念实例与规则前件内涵的数据关系;由于前件和后件共享 Meta 属性内涵,第 12)步根据后件 Meta 属性的内涵与外延的数据关系,在前件 Meta 属性内涵与外延上重新进行后件数据关系连接;第 13)步连接前件对象和后件对象,并标注支持度和置信度信息。

5 实验结果与分析

本文以某省全员人口数据库为数据源对算法进行了实际测试和验证。选取 5 个以育龄妇女为核心的属性:文化程度、生育年龄、地区、户口性质、间隔(育龄妇女生育世代间隔差),对关联规则挖掘结果进行知识表示。

表 2 列出了育龄妇女信息记录, ID 代表育龄妇女人口编码。表 3 列出了由表 2 的 Meta 属性集合所挖掘出的关联规则结果, AR_ID 代表关联规则序号及与该规则相关的实例编号, ObjectLHS 代表规则前件对象, LHS 代表规则前件, ObjectRHS 代表规则后件对象, RHS 代表规则后件, Domain Knowledge 代表与规则前件和后件对象相关的其他 Meta 属性, Supp 代表支持度, Conf 代表置信度。

表 2 育龄妇女信息记录表

ID	姓名	生育年龄	文化程度	户口性质	地区	世代间隔	母亲
0001	刘*丹	23	中学	农业	山区	4	0009
0002	杨*红	25	大学	城市	平原	1	0007
0003	吴*丽	27	大学	城市	平原	1	0001
...

表 3 关联规则挖掘结果

AR_ID	ObjectLHS	LHS	ObjectRHS	RHS	Domain Knowledge	Supp(%)	Conf(%)
001	育龄妇女	{文化程度,地区,户口性质}	育龄妇女	{世代间隔}	生育年龄	20	70
001_ins1	育龄妇女: 0001	{文化程度:中学,地区:山区,户口性质:农业}	育龄妇女: 0001	{世代间隔:4}	生育年龄:23		
001_ins2	育龄妇女: 0002	{文化程度:大学,地区:平原,户口性质:城市}	育龄妇女: 0002	{世代间隔:1}	生育年龄:25		
...		
002	育龄妇女 (母亲)	{生育年龄,户口性质}	育龄妇女 (女儿)	{生育年龄,户口性质}	地区,文化程度,世代间隔	66.8	15.8
002_ins1	育龄妇女: 0001	{生育年龄:23,户口性质:农业}	育龄妇女: 0003	{生育年龄:27,户口性质:城市}	{育龄妇女:0001,地区:山区,文化程度:中学,世代间隔:4} {育龄妇女:0003,地区:平原,文化程度:大学,世代间隔:1}		
...

依据算法 1 构建规则 001{文化程度,地区,户口性质}

==>{世代间隔}(Supp=20%, Conf=70%)的 Vis-Meta 图

概念关系表示 C_VizRule,如图 5 所示。

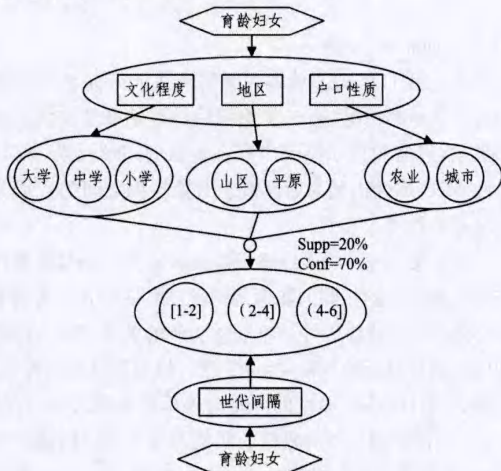


图 5 关联规则概念关系表示

其中包含的数据关系有：

(1)前件对象概念与前件 Meta 属性之间的数据关系：

$R(\text{育龄妇女}, \{\text{文化程度}, \text{地区}, \text{户口性质}\})$

(2)后件对象概念与后件 Meta 属性之间的数据关系：

$R(\text{育龄妇女}, \{\text{世代间隔}\})$

(3)前件 Meta 属性的内涵与外延构成数据关系：

$R(\text{文化程度}, \{\text{大学}, \text{中学}, \text{小学}\})$

$R(\text{地区}, \{\text{山区}, \text{平原}\})$

$R(\text{户口性质}, \{\text{农业}, \text{城市}\})$

(4)后件 Meta 属性的内涵与外延构成数据关系 $R(\text{世代间隔}, \{[1-2], (2-4), (4-6)\})$

规则关系为：

$R(\{\{\text{大学}, \text{中学}, \text{小学}\}, \{\text{山区}, \text{平原}\}, \{\text{农业}, \text{城市}\}\}, \{[1-2], (2-4), (4-6)\})$

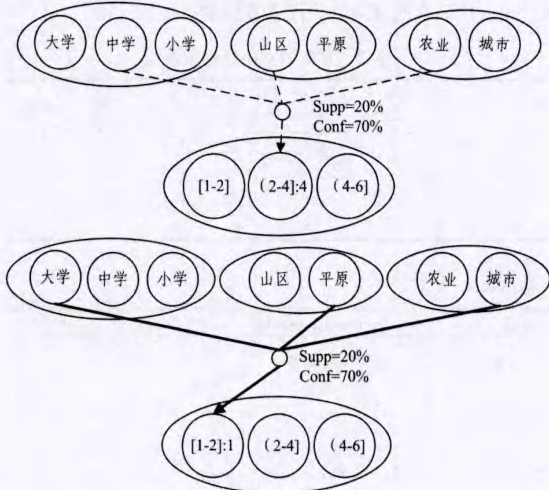


图 6 规则实例关系

1)对比两个规则实例:001_ins1{育龄妇女:刘*丹,文化程度:中学,地区:山区,户口性质:农业}==>{世代间隔:4};001_ins2{育龄妇女:杨*红,文化程度:大学,地区:平原,户口性质:城市}==>{世代间隔:1}。001_ins1 和 001_ins2 具有相同的 Meta 属性内涵{文化程度,地区,户口性质,世代间隔},首先根据 2.2 节的关联规则表示方法构建规则实例关系,如图 6 所示。接着引入概念关系根据算法 2 将两个规则

实例进行合并,共享 Meta 属性内涵,对于离散化的连续属性,以“:值”的形式注明,如世代间隔为 1 表示为“[1-2]:1”,为了保持展示效果的对称性与层次性,根据算法 3 来简化知识表示形式,为了提高对比的视觉效果,用虚线表示规则 001_ins1 的实例化对象关系、规则关系和 Meta 属性内涵,用粗实线表示规则 001_ins2 的对应内容,其余内容以细实线表示。从图 7 中的对比我们能够发现以下隐含知识:育龄妇女的文化程度、所处地区和户口性质对本人的生育世代间隔差存在一定的影响,并且文化程度、所属地区与生育世代间隔差成反比关系。

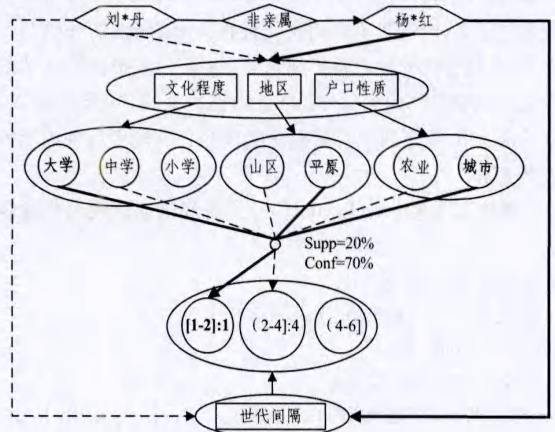


图 7 关联规则实例对比

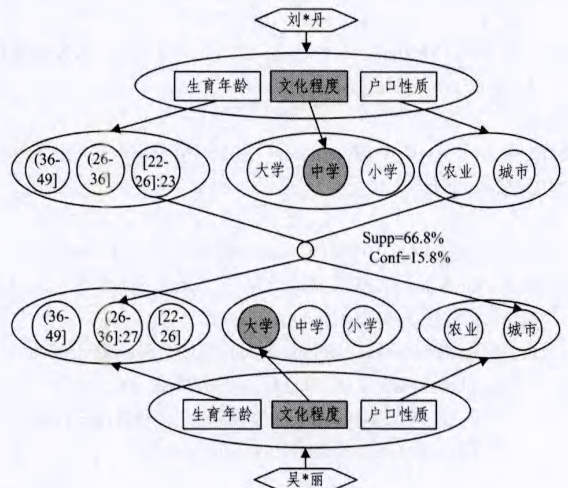


图 8 规则的实例化概念关系

2)将人口关联规则实例{母亲:刘*丹,生育年龄:23,户口性质:农业}==>{女儿:吴*丽,生育年龄:27,户口性质:城市}(Sup=66.80%,Conf=15.8%)进行知识表示,首先构建实例化概念关系,如图 8 所示,其中灰色节点代表没有出现在规则中的领域知识,领域知识数据关系实例为 $R(\text{刘*丹}, \text{文化程度}, \{\text{大学}, \text{中学}, \text{小学}\}; \text{中学})$, $R(\text{杨*红}, \text{文化程度}, \{\text{大学}, \text{中学}, \text{小学}\}; \text{大学})$ 。根据算法 3 对规则表示形式进行简化,规则前件和后件共享 Meta 属性内涵,图 9 为了提高对比的视觉效果,用虚线表示规则前件的实例化对象关系、数据关系,用粗实线表示规则后件的对应内容,其余内容以细实线色表示。我们可以得到以下隐含知识:女儿的生育年龄一般比母亲的生育年龄大,二者相差 4 岁;母女生育世代间隔差为 4 年,表明文化程度对生育年龄、生育世代间隔差有一定的影

响,并且文化程度与生育年龄、生育世代间隔差成反比关系。

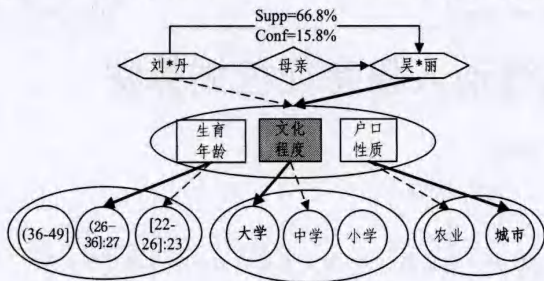


图9 关联规则实例简化表示

本文所提方法能够有效地实现基于 Vis-Meta 图的关联规则知识表示,并且规则信息表示形式简洁明确。与基于表、基于矩阵、基于 TwoKey 图、基于概念图以及基于平行坐标的方法^[8,13,14]相比具有以下优点:能够实现一对一、一对多、多对一以及多对多的多模式关联规则实例化表示,表现力强;能充分展示规则的属性信息和相关领域知识,帮助用户获取更为精确的规则信息以及隐含的知识;表示结果布局清晰,具有较强的可解释性,便于普通用户理解。

结束语 本文给出了 Vis-Meta 图的概念,引入概念关系实现了 Vis-Meta 图应用于关联规则知识表示,所设计的方法能够明确展示领域知识与规则项,不仅能够进行多模式规则实例的对比,而且能够通过展示领域知识辅助规则分析,层次对称结构便于将表示符号转换为自然语言,降低普通用户的感知负担,在一定程度上实现了知识共享。通过将所提算法应用于某省全员人口数据规则,表明该方法具有良好的展示效果,用户在规则知识的基础上发现了新的潜在的知识,为决策提供了有力支撑。

参 考 文 献

[1] Basu A, Blanning R. Metagraphs and Their Applications [M]. Berlin: Springer-Verlag, 2006; 1-11, 77-115

[2] Gaur D. Metagraph a New Hierarchical Data Structured As a Decision Tree[J]. The Journal of Computer Science and Information Technology, 2007, 6(1): 1-5

[3] Gaur D, Shastri A, Biswas R. Metagraph-Based Substructure

(上接第 213 页)

[8] ACE (Automatic Content Extraction) Chinese Annotation Guidelines for Events[R]. National Institute of Standards and Technology, 2005

[9] 刘茂福,李文捷,姬东鸿. 基于事件项语义图聚类多文档摘要方法[J]. 中文信息学报, 2010, 24(5): 77-84

[10] 韩永峰. 基于事件抽取的网络新闻多文档自动摘要[J]. 中文信息学报, 2012, 26(1): 58-66

[11] Ge Shu-zhi, Zhang Zheng-chen, He Hong-shen. Weighted Graph Model Based Sentence Clustering and Ranking for Document Summarization[C]//4th International Conference on Interaction Sciences (ICIS). 2011; 90-95

[12] Thwaites P. Causal identifiability via Chain Event Graphs[J]. Artificial Intelligence, 2013(195): 291-315

[13] Zhong Zhao-man, Liu Zong-tian. Ranking Events Based on E-

Pattern mining [C] // International Conference on Advanced Computer Theory and Engineering, 2008. (ICACTE'08). IEEE, 2008; 865-869

[4] Hu Zen-jun, Mellor J, Wu Jie, et al. Towards zoomable multidimensional maps of the cell[J]. Nature biotechnology, 2007, 25(5): 547-554

[5] Dashore P, Jain S, Dashore S R. Fuzzy Metagraph and Rule Based System for Decision Making in Share Market[J]. International Journal of Computer Applications, 2010, 6(2): 10-13

[6] Dashore P, Jain S. Fuzzy Rule Based Expert System to Represent Uncertain Knowledge of E-commerce [J]. International Journal of Computer Theory and Engineering, 2010, 2: 882-886

[7] Mukherjee A, Sen A K, Bagchi A. The representation, analysis and verification of business processes: a metagraph-based approach[J]. Information Technology and Management, 2007, 8(1): 65-81

[8] 郭晓波,赵书良,刘军丹,等. 基于概念图的关联规则知识表示[J]. 计算机科学, 2013, 40(8): 261-265

[9] 谭政华,胡光锐,任晓林. 模糊元图及其特性分析[J]. 计算机研究与发展, 2000(3): 272-277

[10] Velazquez-Garcia E, Lopez-Arevalo I, Sosa-Sosa V. Distributed Computing and Artificial Intelligence[M]. Springer Berlin Heidelberg, 2012; 469-476

[11] Jain P D S K. Fuzzy rule based system and metagraph for risk management in electronic banking activities [J]. International Journal of Engineering and Technology, 2009, 1(1): 1793-8236

[12] Tan Z H. Fuzzy metagraph and its combination with the indexing approach in rule-based systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(6): 829-841

[13] Bruzzese D, Davino C. Visual mining of association rules[C]// Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, LNAI 6208. Berlin: Springer-Verlag, 2008; 103-122

[14] Liu Gui-mei, Suchitra A, Zhang Hao-jun, et al. AssocExplorer: an association rule visualization system for exploratory data analysis[C]// Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2012; 1536-1539

vent Relation Graph for a Single Document [J]. Information Technology Journal, 2010, 9(1): 174-178

[14] 刘宗田,黄美丽,周文,等. 面向事件的本体研究[J]. 计算机科学, 2009, 36(11): 189-192, 199

[15] Page L, Brin S, Motwani R, et al. The Pagerank citation ranking: Bringing order to the Web, Technical report [J]. Stanford University, 1998

[16] 蒋效宇. 基于关键词抽取的自动文摘算法[J]. 计算机工程, 2012, 38(3): 183-186

[17] 葛斌,李芳芳,李卓,等. 基于无向图构建策略的主题句抽取[J]. 计算机科学, 2011, 38(5): 181-185

[18] Jaruskulchai C, Kruengkrai C. Generic text summarization using local and global properties of sentences[C]// Proceedings of the IEEE/WIC International Conference on Web Intelligence. Piscataway, USA: IEEE Press, 2003; 201-206