

基于 RBFCM 的启发型协调器算法的研究

彭 珍^{1,2} 杨炳儒¹ 谢永红¹

(北京科技大学信息工程学院 北京 100083)¹ (华北科技学院计算机系 北京 101601)²

摘要 用于知识发现的启发型协调器能够模拟认知心理学中的“创建意向”,提高机器的认知自主性。为了进一步提高启发型协调器算法的认知特性和运算效果,构造了一种基于规则模糊认知图的启发型协调器算法,该算法用规则模糊认知图表示知识和进行有效推理,发现知识库中的不关联态,进而对海量数据库进行“定向挖掘”。实验证明该方法与有向超图方法相比,缩小了知识发现算法的搜索空间,增强了智能化程度。

关键词 知识发现,双库协同机制,启发型协调器,规则模糊认知图,可达矩阵,定向挖掘

Research on RBFCM-based Heuristic Coordination Algorithm

PENG Zhen^{1,2} YANG Bing-ru¹ XIE Yong-hong¹

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(Department of Computer, North China of Science and Technology, Beijing 101601, China)²

Abstract Heuristic coordinator for knowledge discovery could simulate “creating intent” of cognitive psychology feature and enhanced the ability of self-cognition. With the aim to improve the cognitive feature and the performance of heuristic coordinator, the paper proposed one new heuristic coordinator algorithm, which used rule based fuzzy cognitive map to represent knowledge and to be effective inference in order to get non-association state of knowledge base for directional mining shortage knowledge in massive database. And the experiment demonstrates that the approach effectively reduces the searching space and increases the intelligence of knowledge discovery compared with the directed hypergraph based method.

Keywords Knowledge discovery in database(KDD), Double bases cooperating mechanism, Heuristic coordinator, Rule based fuzzy cognitive map(RBFCM), Accessible matrix, Directional mining

1 引言

目前,社会各行各业中存储的数据量呈飞速增长趋势,但知识匮乏。KDD^[1]为解决这种“数据丰富而知识贫乏”的困境提供了一种有效途径。但是目前国际上 KDD 的研究无法从根本上明显提高现有的知识发现过程的性能^[2]。而基于双库协同机制^[3,4]的 KDD 从体系结构上对知识发现模型进行了创造性的改革,将知识发现过程视为认知过程,通过模拟认知心理的“创建意向”和“心理特征修复”两个重要特征来提高 KDD 系统的智能性和工作效率。其中,启发型协调器^[5,6]实现了前者,它能够从知识库中自主发现知识短缺,激发数据库中的相应子结构,启动挖掘进程使机器自主聚焦,进而形成数据库中的定向挖掘。

文献[6]中基于有向超图的启发型协调器在发现知识短缺过程中存在以下缺陷:只考虑单个后件规则,只包含了在知识库中出现的合结点以及有限的可达矩阵推理等。基于此,本文提出了一种基于规则模糊认知图^[7,8]的启发型协调器算法。规则模糊认知图是一种软计算工具,具有更强的推理能

力。这种启发型协调器算法使用规则模糊认知图表示知识,通过模糊认知图的推理得到模糊认知图的可达矩阵,从中发现知识短缺,进而启动定向挖掘。实验证明,这种方法与基于有向超图的方法相比,具备了更有效的推理规则,缩小了搜索空间,降低了算法的复杂度;同时,它的智能程度更高,更符合知识发现的认知规律。

2 RBFCM 知识表示和推理机制

2.1 RBFCM 的知识表示方法

模糊认知图 FCM^[9,10]是一种软计算工具,它的概念及概念间的关系是模糊变量,知识存储在概念结点及概念结点间的关系中。但是,FCM 中每个概念及概念间的关系只具有一个模糊成员函数,只适应概念间的简单模糊关系,不能处理概念间的“and”关系。而规则模糊认知图(Rule Based Fuzzy Cognitive Maps, RBFCM)能够解决这些问题,它把模糊认知图中的模糊值用模糊规则来代替,概念结点的模糊变量用多个模糊成员函数来代替,它的知识表示和推理能力更强。本文提出了一种用于启发型协调器的 RBFCM,定义如下:

到稿日期:2009-04-03 返修日期:2009-06-15 本文受国家自然科学基金资助课题(编号:60875029,60675030),横向课题(调解脾胃法治疗冠心病个体)资助。

彭 珍(1981—),女,博士生,讲师,CCF 会员,主要研究方向为数据挖掘,E-mail:yx_dpzc@yahoo.com.cn;杨炳儒(1946—),男,博士生导师,主要研究方向为知识发现与智能系统;谢永红(1970—),女,博士,副教授,主要研究方向为数据挖掘。

定义 1 RBFCM 的拓扑结构 U 是一个四元组即 $U = (C, E, W, S)$, 其中 $C = \{C_1, C_2, \dots, C_n\}$ 表示 RBFCM 的所有素概念结点以及关系结点所涉及到的合概念结点(包括指向关系结点的所有概念结点的与、关系结点指向的所有概念结点的与); $E = \{\langle C_i, C_j \rangle \mid C_i, C_j \in C\}$ 表示概念结点之间的有向弧, 即关系结点涉及到的有向弧; $W = \{W_{ij} \mid W_{ij}$ 是有向弧 $\langle C_i, C_j \rangle$ 的权值}; $S = \{S_i \mid S_i$ 是 C_i 的状态值}。

如图 1 所示, 每个 RBFCM 中每个素概念结点代表的是数据库中的每个属性, 这些结点存放于 Nodes 集合中(包括所有的素结点和已存在的合结点)。Nodes 对应一个 S , 其中的 S_i 是概念结点 C_i 对应的状态值, 它等于 $\sigma(C_i)/N$, 其中 $\sigma(C_i)$ 表示在数据集中 C_i 为真的记录数, N 是数据记录的总数。

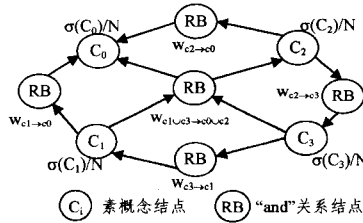


图 1 一个规则模糊认知图模型

定理 1 任意结点的父结点的状态值都小于等于其子结点的状态值, 且该结点的状态值处于其子结点中的最小状态值与其父结点中的最大状态值之间。

证明: 假设存在任意结点 C_j , 那么对于其任意子结点 $C_i \subset C_j$, 其任意父结点 $C_k \supset C_j$, 所以 $\delta(C_i) \geq \delta(C_j)$ 与 $\delta(C_j) \geq \delta(C_k)$ 成立, 那么任意父结点的状态值 $\delta(C_k)/N$ 必定小于等于任意子结点的状态值 $\delta(C_i)/N$, 且 C_j 结点的状态值必定处于父结点中最大状态值与子结点中最大状态值之间。

RBFCM 中每个关系结点都对应一个边权值 w , 表示为结点 C_i 与 C_j 之间存在着“if C_i then C_j ”的概率关系规则, 其中 C_i 为指向该关系结点的各个概念结点的“与”, C_j 为该关系结点指向的各个概念结点的“与”, 关系结点的个数即为规则的个数。

每个 RBFCM 都对应一个 W 关联矩阵, 其中的规则 $C_i \rightarrow C_j$ 永真, 并且无需参与推理计算。其余的每个权值 w 表示为 $\{\text{flag}, \text{sup}(C_i \rightarrow C_j)\}$, 其中 flag 是该规则的标识位, 其值为 1 或者 0, 分别表明规则“if C_i then C_j ”为知识性规则、非知识性规则。 $\text{sup}(C_i \rightarrow C_j)$ 表示的是规则 $C_i \rightarrow C_j$ 的支持度, 说明了规则“if C_i then C_j ”成立的模糊程度。

定理 2 任意规则 $C_i \rightarrow C_j$ 的支持度处于结点 $C_i \cup C_j$ 的所有子结点的最小状态值与所有父结点最大状态值之间。

证明: 因为规则 $C_i \rightarrow C_j$ 的支持度等于 $\sigma(C_i \cup C_j)/N$, 设存在 $C_{k1} \subseteq C_i \cup C_j \subseteq C_{k2}$ 并且在所有子结点中 C_{k1} 的状态值最小和所有父结点中 C_{k2} 的状态值最大。根据定理 1, $C_i \rightarrow C_j$ 的支持度处于 $C_i \cup C_j$ 的 C_{k1} 的状态值与 C_{k2} 的状态值之间。

定理 3 任意规则 $C_i \rightarrow C_j$ 的支持度为 $\text{sup}(C_i \rightarrow C_j)$, 其可信度则为 $\text{sup}(C_i \rightarrow C_j)/S_i$ 。

证明: 因为对于任意规则 $C_i \rightarrow C_j$ 都存在 $\text{sup}(C_i \rightarrow C_j) = \sigma(C_i \cup C_j)/N$, $\text{con}(C_i \rightarrow C_j) = \sigma(C_i \cup C_j)/\sigma(C_i)$, 所以 $\text{sup}(C_i \rightarrow C_j) * N/\sigma(C_i) = \text{con}(C_i \rightarrow C_j)$, 其中 $\sigma(C_i) = S_i * N$, 故 $\text{sup}(C_i \rightarrow C_j)/S_i = \text{con}(C_i \rightarrow C_j)$ 。

所以规则的模糊程度可只用支持度来代表。如果支持度 $\text{sup}(C_i \rightarrow C_j)$ 和可信度 $\text{sup}(C_i \rightarrow C_j)/S_i$ 都大于等于其阈值, 规则 $C_i \rightarrow C_j$ 为知识性规则, flag 则记为 1; 否则规则 $C_i \rightarrow C_j$ 为非知识性规则, 标识 flag 记为 0。

2.2 RBFCM 的推理机制

知识推理是启发式协调器算法的基础, 它的过程可根据规则模糊认知图的邻接矩阵 W 已有的信息推理出 W 的可达矩阵, 从而缩小短缺知识的范围。推理主要包括以下推理 1—推理 5, 而推理 6 与推理 7 是不确定性推理。

推理 1 若规则 $A \rightarrow B$ 成立, 其中后件 B 是合结点, 那么对任意 $B_1 \subset B$, 规则 $A \rightarrow B_1$ 以及规则 $A \cup B_1 \rightarrow (B - B_1)$ 都必定成立。

证明: 已知规则 $A \rightarrow B$ 成立, 所以 $\sigma(A \cup B)/N \geq \text{sup_th}$, $\sigma(A \cup B)/\sigma(A) \geq \text{con_th}$ 必定成立, 其中 sup_th , con_th 分别为支持度阈值和可信度阈值。

对规则 $A \rightarrow B_1$ 而言, 对任意 $B_1 \subset B$, 其支持度 $\sigma(A \cup B_1)/N \geq \sigma(A \cup B)/N$, 可信度 $\sigma(A \cup B_1)/\sigma(A) \geq \sigma(A \cup B)/\sigma(A)$ 。所以 $\sigma(A \cup B_1)/N \geq \text{sup_th}$ 与 $\sigma(A \cup B_1)/\sigma(A) \geq \text{con_th}$ 必然成立, 则 $A \rightarrow B_1$ 必定成立。

对规则 $A \cup B_1 \rightarrow (B - B_1)$ 而言, 对任意 $B_1 \subset B$, 其支持度 $\sigma(A \cup B_1 \cup (B - B_1))/N = \sigma(A \cup B)/N \geq \text{sup_th}$, 可信度 $\sigma(A \cup B)/\sigma(A \cup B_1) \geq \sigma(A \cup B)/\sigma(A) \geq \text{con_th}$, 故规则 $A \cup B_1 \rightarrow (B - B_1)$ 必定成立。

推理 2 若规则 $A \rightarrow B$ 成立, 其中后件 B 是合结点, 那么对已存在的 $B_1 \subseteq B$, 可以确定规则 $B_1 \rightarrow A \cup (B - B_1)$ 是否成立。

证明: 规则 $B_1 \rightarrow A \cup (B - B_1)$ 的支持度 $\sigma(B_1 \cup A \cup (B - B_1))/N = \sigma(A \cup B)/N \geq \text{sup_th}$; 规则 $B_1 \rightarrow A \cup (B - B_1)$ 的可信度 $\sigma(A \cup B)/\sigma(B_1) = \text{sup}(A \rightarrow B)/S(\text{id}_{B_1})$, 因为 $B_1 \subset B$ 且存在, 取 id_{B_1} 为结点 B_1 的编号, 故可以根据 B_1 结点的状态值判定规则的可信度是否大于等于可信度阈值, 从而可以确定规则 $B_1 \rightarrow A \cup (B - B_1)$ 是否成立。

推理 3 若规则 $A \rightarrow B$ 成立, 其中前件 A 是合结点, 那么对已存在的 $A_1 \subset A$, 可以确定规则 $A_1 \rightarrow B \cup (A - A_1)$ 是否成立。

证明: 略(证明思路同上)。

推理 4 对已存在的 $A_1 \subset A$ (A 为合结点), 若 A 的状态值大于等于支持度阈值, 并且大于等于 A_1 的状态值与可信度阈值的乘积, 则规则 $A_1 \rightarrow (A - A_1)$ 必定成立, 否则规则 $A_1 \rightarrow (A - A_1)$ 必定不成立。

证明: 任意合结点 A 都可看作 $A_1 \cup (A - A_1)$, 对 $A_1 \rightarrow (A - A_1)$ 而言, 其支持度 $\sigma(A_1 \cup (A - A_1))/N = \sigma(A)/N \geq \text{sup_th}$, 可信度 $\sigma(A_1 \cup (A - A_1))/\sigma(A_1) \geq \text{con_th}$ 。因此, 规则 $A_1 \rightarrow (A - A_1)$ 必定成立; 否则, 支持度或者可信度不成立, 则 $A_1 \rightarrow (A - A_1)$ 必定不成立。

推理 5 若规则 $A \rightarrow B$ 不成立, 那么对任意 C 且 $B \subset C$, 规则 $A \rightarrow C$ 必定不成立。

证明: 因为规则 $A \rightarrow B$ 不成立, 所以 $\sigma(A \cup B)/N < \text{sup_th}$ 或 $\sigma(A \cup B)/\sigma(A) < \text{con_th}$ 成立。对规则 $A \rightarrow C$ 而言, 其支持度 $\sigma(A \cup C)/N \leq \sigma(A \cup B)/N$, 可信度 $\sigma(A \cup C)/\sigma(A) \leq \sigma(A \cup B)/\sigma(A)$, 所以 $\sigma(A \cup C)/N < \text{sup_th}$ 与 $\sigma(A \cup C)/\sigma(A) < \text{con_th}$ 都成立, 因此, 规则 $A \rightarrow C$ 不成立。

推理 6 若规则 $A \rightarrow B$ 不成立,那么对任意 $A_1 \subset A$,规则 $A_1 \rightarrow B$ 不一定成立。

证明:已知 $\sigma(A \cup B)/N < \text{sup_th}$ 或 $\sigma(A \cup B)/\sigma(A) < \text{con_th}$,规则 $A_1 \rightarrow B$ 的支持度 $\sigma(A_1 \cup B)/N \geq \sigma(A \cup B)/N$,可信度 $\sigma(A_1 \cup B)/\sigma(A_1)$ 都无法与其阈值进行比较,故规则 $A_1 \rightarrow B$ 不一定成立。

推理 7 若规则 $A \rightarrow B$ 和 $B \rightarrow C$ 成立,则规则 $A \rightarrow C$ 不一定成立。

证明:因为每一规则都有一定的模糊度(支持度和可信度),不能通过规则 $A \rightarrow B$ 和 $B \rightarrow C$ 的支持度和可信度推得规则 $A \rightarrow C$ 的支持度与可信度值一定成立,故无法断定规则 $A \rightarrow C$ 是否成立。

3 启发型协调器算法

基于认知心理特征的启发型协调器的实现中,最关键的部分就是可达矩阵的推理算法,它能够缩小知识结点不关联态的范围,从而启发与激活真实数据库中相应的数据子类结构,产生“定向挖掘”以发现短缺知识。

3.1 可达矩阵的推理算法

可达矩阵推理主要来源于 5 个推理(推理 1—推理 5),分别用以下 5 个函数来实现。其中,Nodes 和 S 分别是数据库中获取的相应的概念结点队列和概念状态队列的全局对象。Nodes 是按照先素结点、后合结点的顺序进行排列的,它包含所有的素结点(对应属性),合结点可通过计算而增加,其中的每个结点都有一个编号 id。而 S 始终与 Nodes 存在一一对应的关系。

```
void W::Reference1(int x,int y)
1 for each  $C_k \subset C_y$ 
2 {  $k = \text{Nodes.GetID}(C_k)$ ;
3  $\text{State} = (\text{S.MaxState}(x,k) + \text{S.MinState}(x,k))/2$ ;
4  $\text{W.SetValue}(x,k,1,\text{state})$ ;
5  $\text{id1} = \text{Nodes.GetID}(C_x \cup C_k)$ ;
6  $\text{id2} = \text{Nodes.GetID}(C_y - C_k)$ ;
7  $\text{W.CopyValue}(\text{id1},\text{id2},x,y)$ ;
void W::Reference2(int x,int y)
1 for each  $C_k \subset C_y$ 
2 if( $\text{Nodes.IsNode}(C_k) \neq 0$ )
3 {  $\text{id} = \text{Nodes.GetID}(C_x \cup (C_y - C_k))$ ;
4  $\text{W.CopyValue}(k,\text{id},x,y)$ 
5 if( $\text{W.Sup}(x,y) < \text{con\_th} * \text{S.State}(k)$ )
6  $\text{W.Set\_Flag}(k,\text{id},0)$ ;
void W::Reference3(int x,int y)
1 for each  $C_k \subset C_x$ 
2 if( $\text{Nodes.IsNode}(C_k) \neq 0$ )
3 {  $\text{id} = \text{Nodes.GetID}(C_y \cup (C_x - C_k))$ ;
4  $\text{W.CopyValue}(k,\text{id},x,y)$ ;
7 if( $\text{W.Sup}(x,y) < \text{con\_th} * \text{S.State}(k)$ )
5  $\text{W.Set\_Flag}(k,\text{id},0)$ ;
void W::Reference4(int x)
1 for each  $C_k \subset C_x$ 
2 if( $(\text{Nodes.IsNode}(C_k)) \neq 0$ )
3 {  $\text{id} = \text{Nodes.GetID}(C_x - C_k)$ ;
4  $\text{W.SetValue}(k,\text{id},0,\text{S.State}(x))$ ;
5 if( $\text{S.State}(x) \geq \text{sup\_th} \&\&$ 
```

```
 $\text{S.State}(x) \geq \text{con\_th} * \text{S.State}(k)$ )
6  $\text{W.SetFlag}(k,\text{id},1)$ ;
void W::Reference5(int x,int y)
1 for each  $C_y \subset C_k$ 
2 { if( $\text{Nodes.IsNode}(C_k) \neq 0$ )
3  $\text{state} = (\text{S.MinState}(x,k) + \text{S.MaxState}(x,k))/2$ ;
 $\text{W.SetValue}(x,k,0,\text{state})$ ;}
```

函数 GetID 实现的是返回参数结点的标号。但如果该结点在 Nodes 队列不存在,则还需创建新结点,同时调用 CreateState 函数向 S 队列中加入相应结点的状态值。未知结点的(必定是合结点)状态值,即是 S 队列中该结点的所有子结点的最小状态值与所有父结点的最大状态值的平均值,可根据定理 1 按照取中原则获得。同样,未知规则($C_x \rightarrow C_k$)支持度的计算,根据定理 2,是取规则($C_x \rightarrow C_y$ 且 $C_k \subset C_y$)支持度与包括在 C_k 和 C_y 内的所有子结点的最小状态值与所有父结点的最大状态值的平均值。函数 IsNode 用于判定 Nodes 队列中是否存在参数结点,若存在返回 1,否则返回 0。

算法 1 可达矩阵 W 的推理算法 Ref_AcsMatrix

输入:关联矩阵 W,规则前件标识 r,规则后件标识 l
输出:规则 $C_r \rightarrow C_l$ 的可达矩阵 W
Ref_AcsMatrix(W,r,l)
1 if($l > \text{Nodes.sLength}()$)
2 { if($f = \text{W.GetFlag}(r,l) = 1$)
3 { $\text{W.Reference1}(r,l)$;
4 $\text{W.Reference2}(r,l)$;
5 $\text{W.Reference4}(l)$;
6 if($r > \text{Nodes.sLength}()$)
7 { if($f = 1$)
8 $\text{W.Reference3}(r,l)$;
9 $\text{W.Reference4}(r)$;
10 if($f = 0$)
11 $\text{W.Reference5}(r,l)$;

对规则 $C_r \rightarrow C_l$ 可达矩阵的计算是根据具体条件通过调用 Reference1—Reference5 实现的。而对规则前件结点 C_r 、后件结点 C_l 是否为合结点的判断依赖于 Nodes.sLength(),它是用来获取结点队列中素结点个数的函数,因为结点队列中是按照先素结点、后合结点的顺序排列的,所以以此可确定结点是否为合结点。

3.2 启发型协调器算法

可达矩阵 W 中权值仍为空的规则即为不关联态,它可能是短缺的知识,也可能是非知识性规则。短缺的知识是指在知识库中尚未出现且无法通过已有知识推理出符合一定支持度和可信度的关联规则。反之,若不满足支持度或可信度,则为非知识性规则。启发型协调器算法 RBFCM_HC 针对不关联态启发与激活真实数据库中相应的数据子类结构,进行“定向挖掘”。

算法 2 启发型协调器算法 RBFCM_HC

输入:数据库对象 DB,初始知识库 K,初始关联矩阵 W
输出:可达矩阵 W,更新的知识库 K
RBFCM_HC(DB,K,W)
1 if($K.\text{RuleNum}() \neq 0$)
2 for(i,j) != Null;
3 { $\text{W.SetValue}(i,j,1,K.\text{RuleSup}(i,j))$;
4 $\text{Ref_AcrMatrix}(W,i,j)$;

```

5 for(; W.FindBest(r,l)==1;)
6 { sup= DB.MineSup(r,l);
7   con= DB.MineCon(r,l);
8   r=Nodes.GetID(Cr);
9   l=Nodes.GetID(Cl);
10  W.SetValue(r,l,0,sup);
11  if (sup>=sup_th && con>=con_th)
12  { K.AddRule(Cr→Cl,sup);
13    W.SetFlag(r,l,1);}
14  Ref_AcrMatrix(W,r,l);
15 }

```

初始关联矩阵是根据初始知识库中的知识构建的。如果初始知识库 K 中存在知识 ($K.RuleNum() \neq 0$)，那么通过算法前 4 步的双层循环计算目前所有结点的可达矩阵；否则算法实质上是从关联矩阵确定最需挖掘的结点 (FindBest) 开始的 (第 5 步)，确定的原则是在当前所有的不关联态中它能够产生最大范围的可达状态；而后根据从数据库中进行“定向挖掘”来更新关联矩阵 W 和知识库 K ；无论得到的是知识规则还是非知识规则，都要调用 Ref_AcrMatrix 来计算该规则的可达矩阵。这样再一次去确定最需挖掘结点时，已通过可达矩阵的推理缩小了搜索的范围，从而减少了挖掘的消耗，改善了知识发现的过程模型，提高了知识发现的工作效率。

4 实验与结果

本文实验的数据来源于某医学科学院提供的胸痹病例。在对原始数据预处理的基础上，获得了基本症状的 4 个素结点 (胸 C、舌 T、脉 P、头 H)。根据专家经验构造了含有 4 个知识结点的基础知识库，所对应的初始关联矩阵如图 2 所示。其中包括 3 个合结点 (CUP, TUH, PUH)，并设置了最小支持度 0.3 和最小可信度 0.6。

	C	T	P	H	CUP	TUH	PUH
C	Null	Null	Null	Null	Null	Null	Null
T	{1,0.45}	Null	Null	Null	Null	Null	Null
P	Null	Null	Null	{1,0.4}	Null	Null	Null
H	Null	Null	Null	Null	Null	Null	Null
CUP	Null	Null	Null	Null	Null	Null	Null
TUH	Null	Null	Null	Null	{1,0.32}	Null	Null
PUH	Null	{1,0.35}	Null	Null	Null	Null	Null

图 2 胸痹症状的初始关联矩阵 W

在主频为 1.60GHz 的 Window xp 环境下，采用 VC++ 6.0 开发环境分别就 RBFCM 的和有向超图的启发型协调算法进行了实验。关于 RBFCM 方法与有向超图方法在可达矩阵方面的结果如表 1 所列。

表 1 RBFCM 方法与有向超图方法的比较

可达矩阵	RBFCM	有向超图
矩阵阶数	14×14	7×4
知识性规则	14	5
非知识性规则	71	0

RBFCM 方法与有向超图方法的实验运行耗时与占用主存大小的结果如图 3 和图 4 所示。为了保证结果的准确性，每一次都运行了 100 次并取其平均值。

从实验结果可以看出，两种方法的时间消耗方面相差不大，而 RBFCM 方法要比有向超图方法在内存空间占有量稍大。原因有以下两点：一是 RBFCM 中每个权值都是由权值

标识和支持度两部分组成的；二是 RBFCM 方法中还引入了 Nodes 和 S 两个队列空间。但是，在计算效果方面 (发现的规则数量)，RBFCM 方法要远远优于有向超图方法，这足以弥补在空间中的消耗。

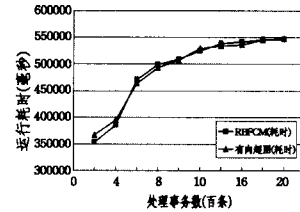


图 3 实验运行耗时的结果

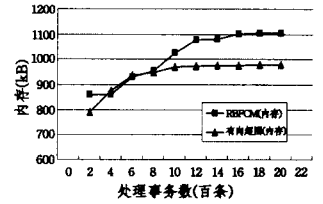


图 4 实验内存占有大小的结果

结束语 启发型协调器使得知识发现系统自身可以实现自我聚焦功能，从而充分体现了知识发现系统的认知自主性与智能性，这将对 KDD 的主流发展起到重要的推动作用。本文针对基于有向超图方法的缺陷提出了一种新的启发型协调器算法，它建立在 RBFCM 的知识表示和推理机制的基础上，与基于超图的启发型协调器算法相比，提供了更多的推理机制，增多了原有知识库中未出现的合结点，而且能够在可达矩阵中确定非知识性结点。实验证明它能够更有效地缩小搜索空间，提高启发协调算法的效率及其智能化程度。

参考文献

- [1] Mannila H. Theoretical frameworks for data mining [J]. SIGKDD Explorations, 2000, 1(2): 30-32
- [2] Piatetsky-Shapiro G. Knowledge Discovery in Database: 10 Years After [J]. SIGKDD Explorations, 2000, 1(2): 59-61
- [3] 杨炳儒, 宋威, 徐章艳. 基于知识发现创新技术的专家系统新构造[J]. 中国科学, 2007, 37(7): 738-747
- [4] 杨炳儒, 王建新, 孙海洪. KDD 中双库协同机制的研究 (II) [J]. 中国工程科学, 2002, 4(5): 34-44
- [5] Yang Bingru. Knowledge Discovery Based on Inner Mechanism: Construction, Realization and Application [M]. USA: Elliott&Fitzpatrick Inc., 2004
- [6] Yang Bingru, Zhang Taohong, Song Wei. Coordinations based on cognitive psychology features and the cooresponding KDD process model[J]. Journal of University of Science and Technology of China, 2007, 37(2): 212-216
- [7] Carvalho J P, Tomé J A B. Rule-based Fuzzy Cognitive Maps and Fuzzy Cognitive Maps-A Comparative Study [C]// Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society. New York, USA, 1999: 115-119
- [8] Carvalho J P, Tomé J A. Qualitative modelling of an economic system using rule-based fuzzy cognitive maps [C]// IEEE International Conference on Fuzzy Systems. Budapest, 2004: 659-663
- [9] Aguilar J. A Survey about Fuzzy Cognitive Maps Papers (Invited Paper) [J]. International Journal of Computational Cognition, 2005, 3(2): 27-33
- [10] Stylios C D, Gmumpca P P. Fuzzy Cognitive Maps: A Soft Computing Technique for Intelligent Control [C]// Proceedings of the IEEE International Symposium on Intelligent Control. Patras, 2000: 97-102