



# (12) 发明专利申请

(10) 申请公布号 CN 105306438 A

(43) 申请公布日 2016. 02. 03

(21) 申请号 201510593297. 1

(22) 申请日 2015. 09. 17

(71) 申请人 杭州安恒信息技术有限公司

地址 310051 浙江省杭州市滨江区通和路浙  
江中财大厦 15 层

(72) 发明人 范渊 王吉文 苏华仕

(74) 专利代理机构 杭州中成专利事务所有限公  
司 33212

代理人 周世骏

(51) Int. Cl.

H04L 29/06(2006. 01)

权利要求书3页 说明书7页 附图1页

## (54) 发明名称

基于模糊粗糙集的网络安全态势评估方法

## (57) 摘要

本发明涉及网络安全态势评估方法,旨在提供基于模糊粗糙集的网络安全态势评估方法。该基于模糊粗糙集的网络安全态势评估方法包括步骤:获取态势因子与态势等级,构成态势等级关于态势因子的决策表;对决策表中决策规则的态势因子进行属性约简,得到约简后的决策表,然后利用约简后的决策表,可通过 KNN 分类器对当前网络需要判断的数据进行决策,得到态势等级。本发明解决了粗糙集方法中实数连续属性需要离散化的问题,另一方面,采用规则聚合的方式,降低了模糊粗糙集方法的计算复杂度,同时,能提供良好的决策结果。



1. 基于模糊粗糙集的网络安全态势评估方法,其特征在于,具体包括下述步骤:

(1) 获取态势因子与态势等级,构成态势等级关于态势因子的决策表;决策表的每一行为一条决策规则,每条决策规则包括各个态势因子的取值和态势等级的取值;

所述态势因子是指能引起网络态势变化的因素,由网络安全信息抽象得到,在决策表中,态势因子为条件属性,取值类型包括离散型和连续型;

所述态势等级是对网络态势的衡量,在决策表中态势等级为决策属性,取值类型为离散型;

(2) 对步骤(1)得到的决策表中,具有相同离散型条件属性值且有相同决策属性值的决策规则进行聚合,即将这些决策规则中的连续实数值聚合成区间值;

(3) 步骤(2)聚合后得到包含离散属性和区间值属性的决策表,对其中不同类型属性值间的相似程度进行定义,构成模糊相似关系;

(4) 将步骤(3)通过相似性定义得到的模糊相似关系,表示为模糊相似矩阵,并定义在模糊关系上模糊粗糙集的上近似、下近似;然后利用定义的模糊粗糙集的上下近似,定义模糊正域和基于模糊正域的依赖函数;

再基于模糊正域的方法对决策表中的态势因子进行属性约简,即当一个属性子集下的依赖函数的取值与属性全集上的依赖函数的取值相同,按照模糊粗糙集的定义,将该属性子集看做是属性全集的一个约简;

(5) 利用步骤(4)得到约简后的决策表,并通过KNN分类器对当前网络需要判断的数据进行决策,得到态势等级。

2. 根据权利要求1所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤(1)中的决策表是一个四元组  $DT = \langle U, C \cup D, V, f \rangle$ ;

其中,  $U$  是一个非空有限集;  $C$  是一个非空有限条件属性集;  $D$  是决策属性;  $V$  是所有属性值域的并,即  $V = \bigcup_{a \in C \cup D} V_a$ ,  $V_a$  是在属性  $a$  上的  $V$  的所有可能取值; 函数  $f: U \times C \cup D \rightarrow V$  是从元素  $x \in U$  和属性  $\forall a \in C \cup D$  到  $V_a$  的一个映射,即  $f(a, x) \in V_a$ ,  $f(a, x)$  是元素  $x$  在属性  $a$  上的取值。

3. 根据权利要求1所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤(2)中,对相同离散型条件属性值且有相同决策属性值的决策规则中,连续实数值进行聚合,具体方式为:  $interval_i(a) = [\min(V_i(a)), \max(V_i(a))]$ ;

其中,  $a$  代表某个连续实数属性,  $i$  代表在  $a$  上离散属性值都相同的某个集合的序号,  $V_i$  代表该集合上某个连续属性的取值的集合,  $\min(V_i(a))$  是集合上属性  $a$  的最小值,  $\max(V_i(a))$  是集合上属性  $a$  的最大值,  $interval_i(a)$  表示聚合后的区间。

4. 根据权利要求1所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤(3)中,对决策表中不同类型的属性值间的相似程度定义如下:

当属性值  $a(x)$ 、 $a(y)$  是布尔值时,属性值的距离定义为

$$db(a(x), a(y)) = \begin{cases} 0, & a(x) = a(y) \\ 1, & a(x) \neq a(y) \end{cases}, \quad \text{即属性值相同时为0,不同时为1,则属性值的相似性定}$$

$$\text{义为 } S(a(x), b(x)) = \frac{1}{1 + db(a(x), a(y))};$$

当属性值  $a(x)$ 、 $a(y)$  是类别值时,属性值的距离定义为

$$dm(a(x), a(y)) = \frac{1}{|U/D|} \sum_{d_i \in U/D} \left( \frac{|a(x)_1 d_i|}{|a(x)|} - \frac{|a(y)_1 d_i|}{|a(y)|} \right)^2, \quad \text{属性值}$$

$$\text{的相似性定义为 } S(a(x), b(x)) = \frac{1}{1 + dm(a(x), a(y))};$$

其中,  $U/D$  表示全集  $U$  在属性集  $D$  上的划分,  $a(x)_1 d_i$  表示属性  $a$  上与元素  $x$  属性值相同的元素的集合与决策属性集划分的一个分块的交集,  $|*|$  代表集合中元素的个数;

当属性值是区间值时,对于两个区间值  $A = [a^-, a^+]$  和  $B = [b^-, b^+]$ ,  $A$  大于  $B$  的概率定义为  $P_{(A \geq B)} = \min(1, \max(\frac{a^+ - b^-}{a^+ - a^- + b^+ - b^-}, 0))$ , 区间值  $A$  和  $B$  的相似性定义为  $S_{AB} = 1 - |P_{(A \geq B)} - P_{(B \geq A)}|$ 。

5. 根据权利要求 1 所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤 (4) 中,模糊粗糙集的上下近似利用  $T$  算子和  $I$  算子进行定义,其定义形式如下:

$$\text{下近似: } \underline{RA}(x) = \inf_{y \in U} I(R(x, y), A(y));$$

$$\text{上近似: } \overline{RA}(x) = \sup_{y \in U} T(R(x, y), A(y));$$

其中,  $R$  是模糊关系,  $A$  是模糊集,  $x$  和  $y$  都是全集  $U$  上的元素;  $T$  是三角模算子,  $I$  是模糊蕴含算子。

6. 根据权利要求 1 所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤 (4) 中,模糊正域定义为:

$$\mu_{POS_C(D)}(x) = \mu_{CX}(x);$$

$\begin{matrix} X \in U/D \\ x \in X \end{matrix}$

其中,  $C$  是条件属性集合,  $D$  是决策属性,  $pos$  表示正域,  $x$  表示任意元素,  $U/D$  表示所有元素在决策属性  $D$  上的划分,  $\mu_{CX}(x)$  表示划分中的一个集合  $X$  在条件属性集合  $C$  下的下近似当中元素为  $x$  时的取值;

基于模糊正域的依赖函数定义为:

$$\gamma_C(D) = \frac{\sum_{x \in U} \mu_{POS_C(D)}(x)}{|U|};$$

其中,  $C$  是条件属性集合,  $D$  是决策属性,  $|U|$  表示所有元素的个数,  $\sum_{x \in U} \mu_{POS_C(D)}(x)$  友示所有元素止域的和。

7. 根据权利要求 1 所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤 (4) 中,对属性全集计算属性约简采用启发式方法:从空集的属性子集开始,不断向属性子集中添加属性全集中的剩余属性,每添加一个属性使属性子集的依赖函数值递增,直至属性子集的依赖函数值与属性全集的依赖函数值相同,得到的属性子集即为该属性全

集的约简。

8. 根据权利要求 1 所述的基于模糊粗糙集的网络安全态势评估方法,其特征在于,所述步骤 (5) 中 KNN 分类器的分类方法具体如下:

选择满足属性条件最多的规则,当有多条满足全部属性条件的规则时,若决策属性一致,则选择该决策属性进行决策,若决策属性不一致,选择其中频数最大的决策属性作为最终决策;

所述满足属性条件定义为:对于离散属性,相同即为满足条件;对于区间值属性,属性值落在区间内即为满足条件。

## 基于模糊粗糙集的网络安全态势评估方法

### 技术领域

[0001] 本发明是关于网络安全态势评估方法,特别涉及基于模糊粗糙集的网络安全态势评估方法。

### 背景技术

[0002] 随着网络规模的不断扩大和趋于复杂,网络的安全威胁也日益加剧。各类日志与报警信息形式多样、格式不一,且冗余较大、误警率高,传统的处理方式已难以应对。在此背景下,为应对大规模网络和多源安全信息,从传统的安全评估衍生出了网络安全态势感知(network security situation awareness,NSSA)。网络安全态势感知包含态势理解、态势评估、态势预测及态势可视化 4 个环节。

[0003] 态势评估是网络安全态势感知的核心环节,也是数据融合领域的研究重点。网络态势评估方法大致可分为 3 类:基于数学模型的方法、基于知识推理的方法和基于模式识别的方法。其中,基于数学模型的方法可细分为层次分析法、集对分析法等;基于知识推理的方法可分为基于图模型的方法、基于证据理论的方法等;基于模式识别的方法可分为灰关联分析方法、粗糙集合方法和神经网络方法等。网络态势评估即在融合各安全信息并进行简单处理的基础上,通过一些数学方法或者数学模型,经过分析,得到一个对当前网络安全状态的整体描述。简言之,该过程即态势因子集合到态势集合的映射。态势因子是指可以引起网络态势变化的因素,由网络安全信息抽象得到。

[0004] 对于网络态势评估方法中的粗糙集方法,由于粗糙集方法不需要先验知识,适合进行网络态势评估的属性约简,但在处理实数型属性值上存在需要离散化的问题,离散化将损失精度,同时,不同的离散化方法或者同一离散化方法设定不同的参数可能导致不同的结果。为解决实数连续属性需要离散化的问题,模糊粗糙集方法作为粗糙集方法的一种推广,是一种可行的方法,但计算复杂度过高。

### 发明内容

[0005] 本发明的主要目的在于克服现有技术中的不足,提供计算时间复杂度大大降低,并能提供良好决策结果的基于模糊粗糙集的网络安全态势评估方法。为解决上述技术问题,本发明的解决方案是:

[0006] 提供基于模糊粗糙集的网络安全态势评估方法,具体包括下述步骤:

[0007] (1) 获取态势因子与态势等级,构成态势等级关于态势因子的决策表;决策表的每一行为一条决策规则,每条决策规则包括各个态势因子的取值和态势等级的取值;

[0008] 所述态势因子是指能引起网络态势变化的因素,由网络安全信息抽象得到(比如将监测数据、日志等作为态势因子),在决策表中,态势因子为条件属性,取值类型包括离散型和连续型;

[0009] 所述态势等级是对网络态势的衡量(如正常、紧急、高危等),在决策表中态势等级为决策属性,取值类型为离散型;

[0010] (2) 对步骤 (1) 得到的决策表中, 具有相同离散型条件属性值且有相同决策属性值的决策规则进行聚合, 即将这些决策规则中的连续实数值聚合成区间值;

[0011] (3) 步骤 (2) 聚合后得到包含离散属性和区间值属性的决策表, 对其中不同类型属性值间 (包括离散属性值和区间属性值) 的相似程度进行定义, 构成模糊相似关系;

[0012] (4) 将步骤 (3) 通过相似性定义得到的模糊相似关系, 表示为模糊相似矩阵, 并定义在模糊关系上模糊粗糙集的上近似、下近似; 然后利用定义的模糊粗糙集的上下近似, 定义模糊正域和基于模糊正域的依赖函数;

[0013] 再基于模糊正域的方法对决策表中的态势因子进行属性约简 (属性约简只是对态势因子进行约简, 即获得一个态势因子的子集, 态势因子的取值类型经过聚合后只有区间型和离散型), 即当一个属性子集下的依赖函数的取值与属性全集上的依赖函数的取值相同, 按照模糊粗糙集的定义, 将该属性子集看做是属性全集的一个约简;

[0014] (5) 利用步骤 (4) 得到约简后的决策表, 并通过 KNN 分类器对当前网络需要判断的数据进行决策, 得到态势等级。

[0015] 在本发明中, 所述步骤 (1) 中的决策表是一个四元组  $DT = \langle U, C \cup D, V, f \rangle$ ;

[0016] 其中,  $U$  是一个非空有限集;  $C$  是一个非空有限条件属性集;  $D$  是决策属性;  $V$  是所有属性值域的并, 即  $V = \bigcup_{a \in C \cup D} V_a$ ,  $V_a$  是在属性  $a$  上的  $V$  的所有可能取值; 函数  $f: U \times C \cup D \rightarrow V$  是从元素  $x \in U$  和属性  $\forall a \in C \cup D$  到  $V_a$  的一个映射, 即  $f(a, x) \in V_a$ ,  $f(a, x)$  是元素  $x$  在属性  $a$  上的取值。

[0017] 在本发明中, 所述步骤 (2) 中, 对相同离散型条件属性值且有相同决策属性值的决策规则中, 连续实数值进行聚合, 具体方式为:  $\text{interval}_i(a) = [\min(V_i(a)), \max(V_i(a))]$ ;

[0018] 其中,  $a$  代表某个连续实数属性,  $i$  代表在  $a$  上离散属性值都相同的某个集合的序号,  $V_i$  代表该集合上某个连续属性的取值的集合,  $\min(V_i(a))$  是集合上属性  $a$  的最小值,  $\max(V_i(a))$  是集合上属性  $a$  的最大值,  $\text{interval}_i(a)$  表示聚合后的区间。

[0019] 在本发明中, 所述步骤 (3) 中, 对决策表中不同类型的属性值间的相似程度定义如下:

[0020] 当属性值  $a(x)$ 、 $a(y)$  是布尔值时, 属性值的距离定义为  $db(a(x), a(y)) = \begin{cases} 0, & a(x) = a(y) \\ 1, & a(x) \neq a(y) \end{cases}$ , 即

属性值相同时为 0, 不同时为 1, 则属性值的相似性定义为  $S(a(x), b(x)) = \frac{1}{1 + db(a(x), a(y))}$ ;

[0021] 当属性值  $a(x)$ 、 $a(y)$  是类别值时, 属性值的距离定义为

$dm(a(x), a(y)) = \frac{1}{|U/D|} \sum_{d \in U/D} \left( \frac{|a(x)_1 d_i|}{|a(x)|} - \frac{|a(y)_1 d_i|}{|a(y)|} \right)^2$  属性值的相似性定义为  $S(a(x), b(x)) = \frac{1}{1 + dm(a(x), a(y))}$ ;

[0022] 其中,  $U/D$  表示全集  $U$  在属性集  $D$  上的划分,  $a(x)_1 d_i$  表示属性  $a$  上与元素  $x$  属性值相同的元素的集合与决策属性集划分的一个分块的交集,  $|*|$  代表集合中元素的个数;

[0023] 当属性值是区间值时, 对于两个区间值  $A = [a^-, a^+]$  和  $B = [b^-, b^+]$ ,  $A$  大于  $B$  的

概率定义为  $P_{(A \geq B)} = \min(1, \max(\frac{a^+ - b^-}{a^+ - a^- + b^+ - b^-}, 0))$ , 区间值  $A$  和  $B$  的相似性定义为  $S_{AB} =$



$$1 - |P_{(A \geq B)} - P_{(B \geq A)}|。$$

[0024] 在本发明中,所述步骤(4)中,模糊粗糙集的上下近似利用T算子(t-norm)和I算子(模糊蕴含算子)进行定义,其定义形式如下:

$$[0025] \quad \text{下近似: } \underline{RA}(x) = \inf_{y \in U} I(R(x, y), A(y)) ;$$

$$[0026] \quad \text{上近似: } \overline{RA}(x) = \sup_{y \in U} T(R(x, y), A(y)) ;$$

[0027] 其中,R是模糊关系,A是模糊集,x和y都是全集U上的元素;T是三角模算子,I是模糊蕴含算子(在实际操作中,需要指定具体的算子,比如T算子用最小值算子 $T_M(a, b) = \min(a, b)$ ,I算子用最大值算子 $I_{S_M}(a, b) = \max(1 - a, b)$ )。

[0028] 在本发明中,所述步骤(4)中,模糊正域定义为:

$$[0029] \quad \mu_{POS_C(D)}(x) = \mu_{\substack{CX \\ X \in U/D \\ x \in X}}(x) ;$$

[0030] 其中,C是条件属性集合,D是决策属性,pos表示正域,x表示任意元素,U/D表示所有元素在决策属性D上的划分, $\mu_{\substack{CX \\ X \in U/D \\ x \in X}}(x)$ 表示划分中的一个集合X在条件属性集合C下的

下近似当中元素为x时的取值;

[0031] 基于模糊正域的依赖函数定义为:

$$[0032] \quad \gamma_C(D) = \frac{\sum_{x \in U} \mu_{POS_C(D)}(x)}{|U|} ;$$

[0033] 其中,C是条件属性集合,D是决策属性,|U|表示所有元素的个数, $\sum_{x \in U} \mu_{POS_C(D)}(x)$ 表示所有元素正域的和。

[0034] 在本发明中,所述步骤(4)中,对属性全集计算属性约简采用启发式方法:从空集的属性子集开始,不断向属性子集中添加属性全集中的剩余属性,每添加一个属性使属性子集的依赖函数值递增,直至属性子集的依赖函数值与属性全集的依赖函数值相同,得到的属性子集即为该属性全集的约简。约简步骤举例如下:

[0035]

```

 $R \leftarrow \{\}, \gamma_{cur} \leftarrow 0, \gamma_C(D)$  //初始化, R 为约简集合
do
 $T \leftarrow R$  //R 暂存到 T
 $\forall x \in (C - R)$  //在剩余属性中选择一个属性
if  $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$  //采用贪心策略, 每一步都加入可使依赖值增大的属性
 $T = R \cup \{x\}$  //加入到约简中
 $\gamma_{cur} \leftarrow \gamma_T(D)$  //更新当前依赖值
 $R \leftarrow T$  //T 赋给 R
until  $\gamma_{cur} = \gamma_C(D)$  //逐个加入属性直到依赖值等于所有条件属性的依赖值
return R //得到约简

```

[0036] 在本发明中,所述步骤(5)中 KNN 分类器的分类方法具体如下:

[0037] 选择满足属性条件最多的规则,当有多条满足全部属性条件的规则时,若决策属性一致,则选择该决策属性进行决策,若决策属性不一致,选择其中频数最大的决策属性作为最终决策;

[0038] 所述满足属性条件定义为:对于离散属性,相同即为满足条件;对于区间值属性,属性值落在区间内即为满足条件。

[0039] 与现有技术相比,本发明的有益效果是:

[0040] 本发明解决了粗糙集方法中实数连续属性需要离散化的问题,另一方面,采用规则聚合的方式,降低了模糊粗糙集方法的计算复杂度,同时,能提供良好的决策结果。

## 附图说明

[0041] 图1为本发明的工作流程图。

## 具体实施方式

[0042] 下面结合附图与具体实施方式对本发明作进一步详细描述:

[0043] 如图1所示的基于模糊粗糙集的网络安全态势感知方法,具体实现方式为:

[0044] 收集态势因子和态势等级得到一张决策表,一张决策表是一个四元组  $DT = \langle U, C \cup D, V, f \rangle$ 。其中,U是一个非空有限集;C是一个非空有限条件属性集;D是决策属性;V是所有属性值域的并,即  $V = \bigcup_{a \in C \cup D} V_a$ ,  $V_a$ 是在属性a上的V的所有可能取值。函数  $f: U \times C \cup D \rightarrow V$  是从元素  $x \in U$  和属性  $\forall a \in C \cup D$  到  $V_a$  的一个映射,即  $f(a, x) \in V_a$ ,  $f(a, x)$  是元素x在属性a上的取值。

[0045] 决策表举例如下:

[0046]



对象	a1	a2	a3	d
x1	A	3.7	yes	1
x2	A	2.5	yes	1
x3	B	1.2	no	0

[0047] 在获得决策表之后,对离散属性值及决策属性值相同的决策规则进行聚合,通过观察可知,上述决策表中对象 x1 和 x2 在 a1,a3,d 上的属性值都相同,而 a2 为实数型属性,对于离散类型属性值相同的规则进行聚合,实数型属性值聚合成区间,分别用这些规则该属性上的最小值和最大值表示左右区间,聚合后的决策表表示如下:

[0048]

对象	a1	a2	a3	d
x1'(x1,x2)	A	[2.5,3.7]	yes	1
x3	B	1.2	no	0

[0049] 此时, x1 与 x2 合并为一个对象,属性 a2 也从原来的实数值类型属性变成了现在的区间值类型属性。

[0050] 聚合后我们得到的是包含离散属性和区间值属性的决策表,为了导出模糊相似关系,需要定义属性值间的相似性,由于属性类型的不同,不同类型需要不同的定义方法:

[0051] 当属性值是布尔值时,属性值的距离定义为  $db(a(x), a(y)) = \begin{cases} 0, & a(x) = a(y) \\ 1, & a(x) \neq a(y) \end{cases}$  属性

值的相似性定义为  $S(a(x), b(x)) = \frac{1}{1 + db(a(x), a(y))}$  以上述决策表举例,  $db(a3(x_1), a3(x_2)) = 0$ ,  $db(a3(x_1), a3(x_3)) = 1$ 。

[0052] 当属性值是类别值时,属性值的距离定义为  $dm(a(x), a(y)) = \frac{1}{|U/D|} \sum_{d_i \in D} (\frac{a(x)_i d_i}{|a(x)|} - \frac{a(y)_i d_i}{|a(y)|})^2$ ,

U/D 表示 U 在决策属性集 D 上的划分。属性值的相似性定义为  $S(a(x), a(y)) = \frac{1}{1 + dm(a(x), a(y))}$ ,

以上述决策表举例,

$$dm(a1(x_1), a1(x_3)) = \frac{1}{2} \left( \left( \frac{|\{x_1, x_2\}|}{|\{x_1, x_2\}|} - \frac{|\{x_3\}|}{|\{x_3\}|} \right)^2 + \left( \frac{|\{x_1, x_2\}|}{|\{x_1, x_2\}|} - \frac{|\{x_3\}|}{|\{x_3\}|} \right)^2 \right) = 1$$

$$S(a(x), a(y)) = \frac{1}{1 + dm(a(x), a(y))} = 0.5$$

[0053] 当属性值是区间值时,对于两个区间值  $A = [a^-, a^+]$  和  $B = [b^-, b^+]$ , A 大于 B

的概率定义为  $P_{(A \geq B)} = \min(1, \max(\frac{a^+ - b^-}{a^+ - a^- + b^+ - b^-}, 0))$  区间值 A 和 B 的相似性定义为  $S_{AB}$

$= 1 - |P_{(A \geq B)} - P_{(B \geq A)}|$ 。下面举例说明,给定两个区间值分别为  $A = [3, 5]$  和  $B = [2, 4]$ ,

$$P_{(A \geq B)} = \min(1, \max(\frac{5-2}{5-3+4-2}, 0)) = 0.75, \quad P_{(B \geq A)} = \min(1, \max(\frac{4-3}{4-2+5-3}, 0)) = 0.25, \quad S_{AB} =$$

$$1 - |P_{(A \geq B)} - P_{(B \geq A)}| = 0.5。$$

[0054] 通过相似性的定义,将得到一个模糊相似矩阵,继而定义在模糊关系上的模糊粗糙集,模糊粗糙集的上下近似定义为  $\overline{RA}(x) = \sup_{y \in U} T(R(x, y), A(y))$  和  $\underline{RA}(x) = \inf_{y \in U} I(R(x, y), A(y))$ ,

其中 T 是 T 算子 (三角模算子), I 是模糊蕴含算子。在实际操作中,需要指定具体的算子,比如 T 算子可用最小值算子  $T_M(a, b) = \min(a, b)$ , I 算子可用最大值算子  $I_{S_M}(a, b) = \max(1 - a, b)$ 。

[0055] 举例说明,假设一个模糊相似矩阵为  $R = \begin{pmatrix} 1 & 0.6 & 0.5 \\ 0.6 & 1 & 0.8 \\ 0.5 & 0.8 & 1 \end{pmatrix}$  决策属性 d 导出的划分为

为  $\{x_1, x_2\}$  和  $\{x_3\}$ ,表示成模糊集为  $d_1 = \{1, 1, 0\}$  和  $d_2 = \{0, 0, 1\}$ 。T 算子和 I 算子分别以最小值算子和最大值算子为例,  $x_1$  的模糊上下近似为

[0056]

$$\overline{Rd}_1(x_1) = \sup_{y \in U} T(R(x_1, y), d_1(y)) = \sup(\min(1, 1), \min(0.6, 1), \min(0.5, 0)) = 1$$

$$\underline{Rd}_1(x_1) = \inf_{y \in U} I(R(x_1, y), d_1(y)) = \inf(\max(1 - 1, 1), \max(1 - 0.6, 1), \max(1 - 0.5, 0)) = 0.5$$

[0057] 对于获得的上下近似,可以定义正域,以及依赖函数,正域的定义为

$\mu_{POS_c(D)}(x) = \mu_{CX}(x)$  举例  $\mu_{POS_c(D)}(x_1) = \mu_{CX}(x_1) = 0.5$  也即下近似,依赖函数的定义为

$$\gamma_c(D) = \frac{|\mu_{POS_c(D)}(x)|}{|U|} = \frac{\sum_{x \in U} \mu_{POS_c(D)}(x)}{|U|}$$
 也即每个元素的下近似的和比上元素个数。

[0058] 基于正域的属性约简方法的原理在于,假如一个属性子集的依赖函数值与属性全集的依赖函数值相同,该属性子集可以看做是一个约简。计算属性约简的过程采用启发式的方法。从空集开始,不断添加属性,每一步使依赖函数递增,直至依赖函数值与在所有条件属性上的依赖函数值相同,算法停止。

[0059] 对于得到的约简后的决策表,采用 KNN 分类器进行分类决策。一个可能的决策表结果表示如下:

[0060]

对象	a1	a2	d
x1	A	[1.2,3.7]	1
x2	A	[1.5,2.5]	1

[0061]

x3	B	[4.2,5.5]	0
x4	A	[2.3,]	0

[0062] 则对于一条需要判断的测试数据,如,

[0063]

对象	a1	a2	d
x5	A	2.4	?

[0064] 显然符合对象 x5 的规则有 x1, x2 和 x4, 最终的决策值为 1、1、0。选取频数最大的决策值作为最后的决策值, 即最后的决策为  $d = 1$ 。

[0065] 最后, 需要注意的是, 以上列举的仅是本发明的具体实施例。显然, 本发明不限于以上实施例, 还可以有很多变形。本领域的普通技术人员能从本发明公开的内容中直接导出或联想到的所有变形, 均应认为是本发明的保护范围。



图 1