第41卷 第11期
2014年11月

计 算 机 科 学
Computer Science

Vol. 41 No. 11
Nov 2014

一种基于区分矩阵的属性组合权重构造方法

叶 军 王 磊

(南昌工程学院信息工程学院 南昌 330099)

摘 要 详细分析了以基于Pawlak的属性重要度来构造属性权重的方法存在的问题，给出了一种基于区分矩阵的属性重要度的定义方法，并且得到了该定义方法的一些重要性质。在此基础上提出了一种新的信息系统属性权重构造方法，新方法是以条件属性在整个信息系统中的贡献度来确定属性权重，不仅反映了属性自身区分对象的能力，而且体现了各属性在整个条件属性中的分类能力。通过算例分析表明，该方法得到的属性权重更加贴近事实，因此它能提高属性权重的准确度。

关键词 粗糙集，区分矩阵，属性重要度，权重

中图法分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.11.053

Approach of Ascertaining Combinatorial Attribute Weight Based on Discernibility Matrix

YE Jun WANG Lei

(School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China)

Abstract The problems existed in the Pawlak attribute importance based method by which the attribute weight is constructed were analyzed in detail firstly, then a discernibility matrix based definition of attribute importance was given. On this basis, a novel approach for constructing the combinatorial attribute weights of information systems was proposed. In this approach, the attribute weights are ascertained according to the condition attribute's contribution in the whole information system. The proposed approach not only reflects the ability of condition attributes to distinguish the object, but also reflects the classification capability of each attribute in the whole condition attributes. The numerical example demonstrates that the attribute weights gained by the proposed approach are more closer to the facts, so the proposed approach can improve the accuracy of the attribute weights.

Keywords Rough sets, Dscernibility matrix, Attribute significance, Weight

1 引言

在处理决策问题时，人们通常是用多个属性来描述系统的特征，并以属性作为评判条件来对系统进行评价或决策，而不同属性在刻画系统特征时所起的作用是不一样的，有的属性对系统很重要，有的属性对系统一般重要或不重要。因此，能否精确地度量属性对系统重要性的权重直接影响决策的准确性。目前常用的属性权重确定方法主要分几大类，一类是主观确定法，如有专家直接赋分、模糊统计、二元对比排序、AHP和Delphi等。如文献[1]采用区间直觉判断矩阵来确定权重；文献[2]提出了一种基于近似函数依赖的属性权重评估方法，该方法基于一致集的概念导出最大集，生成最小非平凡函数依赖集，从而找出属性之间的近似函数依赖关系；文献[3]采用区间数群决策矩阵来确定专家权重。主观法的主要缺点是仅凭经验得出的结果往往具有较大程度的随意性，有时可能会导致评价结果的准确度不高。为了克服主观法的不足，一些学者提出了客观确定法或主观与客观相结合的方法等。在客观法中，最典型的是用粗糙集理论[4]中的属性重要程度来确定属性权重。如文献[5,6]采用粗糙集理论中的基于Pawlak的属性重要程度来确定属性权重；文献[7-10]采用基于信息熵的重要程度来确定权重；文献[11,12]采用Pawlak和信息熵相结合的方法来确定属性权重。这些方法在一定程度上克服了主观性带来的不足，有利于提高决策精度。但事实上，无论是采用基于Pawlak的属性重要程度来确定属性权重，还是采用基于信息熵的属性重要度确定权重，两者都有一定的局限性。基于Pawlak的属性重要度定义侧重属性的定性分析，它定义的是核属性的重要度，非核属性的重要度都为零，无法确定非核属性重要度所占权重；而信息熵的属性重要程度定义侧重属性的定量分析，虽然避免了约简集中的非核属性都为零的情况，但有时会出现非核属性的重要度大于核属性的重要度的情况[13]，这与基于Pawlak表示的属性重要度相矛盾。为此，本文在前人研究的基础上，给出了一种新的粗糙集理论下基于区分矩阵的属性重要度定义方法，然后提出了一种以新的属性重要度和传统的属性重要度相结合来构造属性权重的新方法，并以此进行评价和决策，为人们提高决策精度提供一种新思路。

到稿日期：2013-06-25 返修日期：2013-08-16 本文受国家自然基金地区科学基金项目(61363047)，国家自然基金(61461032)，江西省自然基金(2011ZBAB201005)，南昌工程学院青年基金项目(2010KJ019)资助。

叶 军(1968－)，男，硕士，副教授，主要研究领域为粗糙集理论、粒计算等，E-mail：1208561815@qq.com；王 磊(1968－)，男，博士，副教授，主要研究领域为粗糙集理论、粒计算、智能控制等。

2 原有代数表示的属性重要度分析

基于 Pawlak 的属性重要度[14]的方法是从决策表中删除某个属性，然后考察决策表的分类变化情况。在决策表中删除某个属性后，决策表的分类能力变化越大，该属性相对决策就越重要；反之，变化越小，说明该属性相对于决策就越不重要。其定义如下：

定义 1[15-18] 决策信息系统 $S=(U,C\cup D,V,f)$。两个属性集 C 与 D 之间的依赖程度 $r(C,D)$ 定义为：$r(C,D)=\frac{|POS_C(D)|}{|U|}$。则对于 $\forall b\in C$ 关于 D 的重要度为：$sig(b,C,D)=r(C,D)-r(C-\{b\},D)$。其中 $|U|$ 表示整个集合对象的个数，$r(C-\{b\},D)$ 表示在 C 中删除属性 b 后，对象分类的变化情况。

该属性重要度定义方法被大量文献引用，用它来确定属性重要度所占权重的定义为：

定义 2[15-18] 决策信息系统 $S=(U,C\cup D,V,f)$。$\forall b\in C$ 关于 D 的重要度所占权重为：$W(b)=\frac{sig(b)}{\sum_{a\in C}sig(a)}$。

根据定义 1 可知，在属性重要度定义的公式中，它是以所有不同条件属性 C 为前提来考察正区域（下近似）的变化情况，而引起正区域变化的只有核属性。也就是说在决策表中，对于所有不同条件属性，如果删除的属性是核属性，则正区域一定会发生变化，即核属性重要度不为零；如果删除的是非核属性，正区域不会发生变化，因此，非核属性的重要度都为零，即非核属性重要性所占权重都为零。

在 Pawlak 的粗糙集理论中，有以下定义：

定义 3[15-18] 决策信息系统 $S=(U,C\cup D,V,f)$。设 $U/D=\{Y_1,Y_2,\cdots,Y_n\}$ 是 D 对论域 U 的划分，$U/C=\{X_1,X_2,\cdots,X_m\}$ 是 C 对论域 U 的划分。C 的 D 正区域记为 $POS_C(D)=\bigcup_{Y_i\in U/D}C\_(Y_i)$ 为 $\forall a\in b\subseteq C$。对 $\forall a\in B\subseteq C$：

(1) $POS_{C-\{a\}}(D)\neq POS_C D$，则属性 a 为 C 的 D 必要的属性，否则 a 为 C 的 D 不必要的属性。所有必要属性组成的集合为核集，记为 $CORE(C)$。

(2) $B\subseteq C$ 是 S 的一个约简当且仅当 $POS_B(D)=POS_C(D)$，$POS_{B-\{a\}}(D)\neq POS_B(D)$。所有约简集记为 $RED(S)$，则有 $\bigcap RED(S)=CORE(C)$。由 $RED(S)-CORE(C)$ 组成的集合为相对必要集，记为 N。

(3) 记所有不必要属性组成的集合为 I，则有 $I=C-RED(S)$。

对定义 3 进行分析可知，在决策信息系统中，核集 $CORE(C)$ 中的属性最重要，它是必不可少的，去掉会影响分类的变化，所以它所占的权重最大；相对必要集 N 中的属性，它是约简集中的属性，没有核属性重要，它的重要度所占权重比核属性小；不必要集 I 中的属性最不重要，其重要度所占权重比约简集中的属性要小，即有 $W_{sig_{CORE(C)}}>W_{sig_N}>W_{sig_I}>0$。但是，根据定义 1 和定义 2 得到的各属性所占权重为：$W_{sig_{CORE(C)}}>W_{sig_N}=W_{sig_I}=0$，这显然与定义 3 相矛盾，而且也不符合客观情况。

事实上，人们用属性来刻画事物或系统的特征时，所有描述该事物的属性都有它的作用，不同属性尽管作用大小不一样，但是不能说没有作用，也就是说，对属性的作用大小进行量化时，它们所占的权重不能为零。为了能够客观真实地反映各属性在刻画事物特征时所起的作用，下面给出一种基于区分矩阵的属性重要度定义方法。

3 基于区分矩阵的属性重要度定义

用区分（差别）矩阵来表示知识是由波兰科学家 A. Skowron[18]首次提出的。用区分矩阵来表示决策信息系统中的核属性集、约简属性集以及其它相关概念，简单明了，而且计算方便。下面简要介绍决策表的区分矩阵的定义及相关定理。

定义 4[19] 决策信息系统 $S=(U,C\cup D,V,f)$，其中 C 为条件属性，D 为决策属性，论域是对象的一个非空有限集合 $U=\{x_1,x_2,x_3,\cdots,x_n\}$，$|U|=n$。则定义如下：

$$M_{n\times n}=(c_{ij})_{n\times n}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

$$=\begin{bmatrix} c_{11} & * & \cdots & * \\ c_{21} & c_{22} & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ * & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & c_{nn} \end{bmatrix}$$

为决策信息系统的区分矩阵，其中，$i,j=1,2,\cdots,n$。

$$c_{ij}=\begin{cases}\{a\,|\,(a\in C\wedge(f(x_i)\neq f(y_j))\}, & f_D(x_i\neq f_D(x_i)\\ \varnothing, & f_D(x_i)=f_D(x_i)\end{cases}$$

由于有 $c_{ij}=c_{ji}(i,j=1,2,3,\cdots,n)$，即区分矩阵是关于主对角线对称的矩阵，因此在实际应用中我们通常采用上三角或下三角表示区分矩阵。对于区分矩阵，有以下两个定理。

定理 1[20,21] 在一个决策表中，决策表的相对 D 核等于决策表中区分矩阵中所有单个属性元素组成的集合，即：

$$CORE_C(D)=\{a\,|\,(a\in C)\wedge(\exists c_{ij},((c_{ij}\in M)\wedge(c_{ij}=\{a\})))\}$$

定理 2[20,21] 在一个决策表中，$\forall B\subseteq C$，若满足：

(1) $\forall c_{ij}\in M_{n\times n}$，当 $c_i\neq\varnothing$ 时，都有 $B\cap c_{ij}\neq\varnothing$；

(2) B 是相对 D 独立的。

那么 B 是决策表的一个相对约简。

从定义 4 可知，当论域中的两个对象的决策值相同，即对象 x_i 和 x_j 不能区分时，说明它们落入决策属性的同一个等价类中，我们就不需要考虑它们的差异，因此 $c_{ij}=\varnothing$。当论域中的两个对象 x_i 和 x_j 能够区分时，c_{ij} 是由所有能区分对象 x_i 和 x_j 的条件属性组成的集合，这些属性可能是核集 $CORE(C)$ 中的属性，也可能是相对必要集 N 中的属性，或者是不必要集 I 中的属性，但是，不管这些属性属于哪个集合，所有这些属性对区分对象 x_i 和 x_j 都是有贡献的，也就是说这些属性都是重要的，只是重要度不一样。为了能够真实反映这些属性的重要性，我们可以用这些属性在 c_{ij} 中出现次数的多少以及在每次中所占的比例来定义这些属性的重要性。当某个属性出现的次数最多且每次占的比例最大时，表示它能区分的对象最多，贡献度最大；相应的重要度也最大，反之，重要度最小。新的重要度定义如下：

定义 5 设 $S=(U,C\cup D,V,f)$ 是一个相容决策信息系统，其中 C 为条件属性，D 为决策属性。c_{ij} 为区分矩阵 M 中的非空元素，K 为区分矩阵中所有非空 c_{ij} 的总个数，对 $\forall b\in C$，属性 b 关于 D 的重要程度定义为：

$$Nsig(b,C,D)=\frac{\sum_{i,j=1,2,\cdots,n}r_b\cap c_{ij}}{K}$$

其中，$r_b \cap c_{ij} = \begin{cases} \frac{1}{|c_{ij}|}, & b \cap c_{ij} \neq \emptyset \\ 0M, & b \cap c_{ij} = \emptyset \end{cases}$，$i,j=1,2,\cdots,n$，$|c_{ij}|$表示非空元素 c_{ij} 中含有的属性个数。

由上述定义，我们可以得到如下几个性质：

性质 1 在相容决策信息系统 $S=(U,C\cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall b\in C$ 有 $0\leqslant r_b\cap c_{ij}\leqslant 1$。

证明：

对 $\forall b\in C$ 且 $b\notin c_{ij}$，即 $b\cap c_{ij}=\emptyset$，由定义 5 可得 $r_b\cap c_{ij}=0$。

对 $\forall b\in C$ 且 $b\in c_{ij}$，若 $b\in CORE(C)$，即 b 为核属性，则根据定理 1，在决策表中，核属性为区分矩阵中所有单个属性元素组成的集合，即 $|c_{ij}|=1$，因此有 $r_b\cap c_{ij}=\frac{1}{|c_{ij}|}=1$。

对 $\forall b\in C$ 且 $b\in c_{ij}$，若 $b\notin CORE(C)$，则必有 $b\in N$，或 $b\in I$，根据定理 1，元素 c_{ij} 的属性个数一定不是单个属性，至少为两个或两个以上，即有 $|c_{ij}|>1$，可得 $0<\frac{1}{|c_{ij}|}<1$，因此有 $0<r_b\cap c_{ij}=\frac{1}{|c_{ij}|}<1$。

综上所述，有：$0\leqslant r_b\cap c_{ij}\leqslant 1$。

在性质 1 中，当 $r_b\cap c_{ij}=1$ 时，反映了单个核属性区分对象的能力最强，它与基于 Pawlak 的属性重要度定义的概念完全一致。

性质 2 在相容决策信息系统 $S=(U,C\cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall b\in C$，则有 $0\leqslant Nsig(b,C,D)\leqslant 1$。

证明：对 $\forall b\in C$ 且 $b\notin c_{ij}$，即 $b\cap c_{ij}=\emptyset$，由定义 5 可知 $Nsig(b,C,D)=\frac{\sum r_b\cap c_{ij}}{K}=\frac{0}{K}=0$，即属性 b 没有出现在区分矩阵中，它对区分对象不起任何作用。

对 $\forall b\in C$ 且 $b\in c_{ij}$，若区分矩阵 M 中所有的非空元素 c_{ij} 都是单属性，根据定理 1 可知这些属性全为核属性，由定义 5 可知 $|c_{ij}|=1$，即 $r_b\cap c_{ij}=1$。因此有 $Nsig(b,C,D)=\frac{\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}}{K}=\frac{1+1+\cdots+1}{K}=\frac{K}{K}=1$。

对 $\forall b\in C$ 且 $b\in c_{ij}$，当区分矩阵 M 中的 c_{ij} 元素全为多个属性组成时，则必有 $b\notin CORE(C)$，且 $b\in N$ 或 $b\in I$，由性质 1 有：$0<r_b\cap c_{ij}<1$，因此有 $0<Nsig(b,C,D)=\frac{\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}}{K}<\frac{1+1+\cdots+1}{K}=\frac{K}{K}=1$。

对 $\forall b\in C$ 且 $b\in c_{ij}$，若区分矩阵 M 中既有单个属性组成的 c_{ij} 元素，也有多个属性组成的 c_{ij} 元素时，一定存在 $b\notin CORE(C)$，且 $b\in N$ 或 $b\in I$，由性质 1 有 $0<r_b\cap c_{ij}<1$，可得 $0<\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}<K$，因此有 $0<Nsig(b,C,D)=\frac{\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}}{K}<1$。

综上所述，证得：$0\leqslant Nsig(b,C,D)\leqslant 1$。

在性质 2 中，对 $\forall b\in c_{ij}$，只要属性 b 出现在区分矩阵的元素 c_{ij} 中，不管是核属性还是非核属性都有它的重要性，从而避免了非核属性的重要度全为零的情况。

性质 3 在相容决策信息系统 $S=(U,C\cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall a,b,c\in c_{ij}$，若有 $a\in CORE(C)$，$b\in N$，$c\in I$，则有：$Nsig_a>Nsig_b>Nsig_c>0$。

证明：因为有 $a\in CORE(C)$，$b\in N$，$c\in I$，由定义 3 可知，一定有 $a\in RED(S)$，$b\in RED(S)$ 且 $b\notin CORE(C)$，$c\notin CORE(C)$ 且 $c\notin RED(S)$，即 a 在区分矩阵素元中出现的次比 b 多，b 比 c 多，即有

$$\sum_{i,j=1,2,\cdots,n} r_a\cap c_{ij}>\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}>\sum_{i,j=1,2,\cdots,n} r_c\cap c_{ij}>0$$

从而可得：

$$\frac{\sum_{i,j=1,2,\cdots,n} r_a\cap c_{ij}}{K}>\frac{\sum_{i,j=1,2,\cdots,n} r_b\cap c_{ij}}{K}>\frac{\sum_{i,j=1,2,\cdots,n} r_c\cap c_{ij}}{K}>0$$

因此有 $Nsig_a>Nsig_b>Nsig_c>0$。

在性质 3 中，核属性最重要，重要度最大；相对必要集 N 中的属性其次，重要度比核属性小；不必要集 I 中的属性最不重要，其重要度最小。结论与定义 3 中属性相关概念完全一致。该性质的逆性质不一定成立。

根据定义 5，我们可以得到一种基于区分矩阵的属性重要度所占权重定义方法，定义如下：

定义 6 相容决策信息系统 $S=(U,C\cup D,V,f)$。c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall b\in C$，属性 b 关于 D 的重要度所占权重为：

$$W_N(b)=\frac{Nsig(b)}{\sum_{a\in C} Nsig(a)}$$

对于定义 6，同样有以下几个性质：

性质 4 在决策信息系统 $S=(U,C\cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall b\in c_{ij}\subseteq C$，则有 $0\leqslant W_N(b)\leqslant 1$。

证明：根据部分小于全体的原理，由性质 1 可直接得证。

在性质 4 中，对 $\forall b\in c_{ij}$，即只要属性 b 出现在区分矩阵的元素 c_{ij} 中，各属性重要度所占权重都不为零，否则为零。

性质 5 在相容决策信息系统 $S=(U,C\cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall a,b,c\in c_{ij}$，若有 $a\in CORE(C)$，$b\in N$，$c\in I$，则有 $W_N(a)>W_N(b)>W_N(c)>0$。

证明：根据部分小于全体的原理，由性质 3 可直接得证。

性质 5 说明了核属性重要度所占权重比非核属性所占权重大，从而避免了非核属性权重比核属性的权重还要高的情况。

4 基于区分矩阵的组合权重确定方法

由定义 5 和定义 6 以及得到的相关性质表明，由区分矩阵定义的属性重要度所占权重真实地反映了各属性在区分其它对象时所起的作用，在一定程度上克服了定义 1 和定义 2 的不足。但是，在区分矩阵中属性反映的是各属性自身区分其它对象的能力，它没有完全体现各属性在整个条件属性中的分类能力。而基于 Pawlak 属性重要度定义反映的是各属性在整个条件属性中的分类能力。因此，我们在考察属性重要度权重的大小时，既要考虑各属性自身作用，也要综合考虑各属性在系统中的整体作用，因此，我们将两种定义相结合，采用组合方法来确定属性权重，这种方式更为科学合理。

定义 7 相容决策信息系统 $S=(U,C\cup D,V,f)$，对 $\forall b\in C$，则属性 b 在整个信息系统中的组合权重为：

$$W_{new}(b)=\frac{W_N(b)+W(b)}{\sum_{a\in C}(W_N(a)+W(a))}$$

其中，$W_N(b)$ 是区分矩阵下的属性重要度所占权重，$W(b)$ 是在 Pawlak 下的属性重要度所占权重。

性质 6 相容决策信息系统 $S=(U,C\cup D,V,f)$，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall b\in C$，则有 $0\leqslant W_{new}(b)\leqslant 1$。

证明：由定义 1 可知：$0\leqslant sig(b)\leqslant 1$，则有

$$0\leqslant\frac{sig(b)}{\sum_{a\in C} sig(a)}\leqslant 1$$

即有 $0 \leqslant W(b) \leqslant 1$。

由性质 4 有 $0 \leqslant W_N(b) \leqslant 1$，从而得到

$$0 \leqslant \frac{W_N(b)+W(b)}{\sum\limits_{a \in C}(W_N(a)+W(a))} \leqslant 1$$

因此有 $0 \leqslant W_{new}(b) \leqslant 1$。

在性质 6 中，对 $\forall b \in c_{ij}$，只要属性 b 出现在区分矩阵的元素 c_{ij} 中，各属性重要度组合权重都不为零，而没有出现在区分矩阵的属性组合权重都为零。

性质 7 在相容决策信息系统 $S=(U,C \cup D,V,f)$ 中，c_{ij} 为区分矩阵 M 中的非空元素，对 $\forall a,b,c \in c_{ij}$，若有 $a \in CORE(C)$，$b \in N$，$c \in I$，则有 $W_{new}(a)>W_{new}(b)>W_{new}(c)>0$。

证明：因为 $a \in CORE(C)$，$b \in N$，$c \in I$，由定义 1 可知：$sig(a)>0$，$sig(b)=0$，$sig(c)=0$，即有 $sig(a)>sig(b)=sig(c)$。由定义 2 可知：

$$W(a)=\frac{sig(a)}{\sum\limits_{a \in C} sig(a)}>0, W(b)=\frac{sig(b)}{\sum\limits_{a \in C} sig(a)}=0,$$

$$W(c)=\frac{sig(c)}{\sum\limits_{a \in C} sig(a)}=0$$

即有 $W(a)>W(b)=W(c)=0$。

由性质 5 有 $W_N(a)>W_N(b)>W_N(c)>0$。从而得到：

$$\frac{W_N(a)+W(a)}{\sum\limits_{a \in C}(W_N(a)+W(a))}>\frac{W_N(b)+W(b)}{\sum\limits_{a \in C}(W_N(a)+W(a))}>\frac{W_N(c)+W(c)}{\sum\limits_{a \in C}(W_N(a)+W(a))}>0$$

所以有 $W_{new}(a)>W_{new}(b)>W_{new}(c)>0$。

性质 7 表明，核属性的组合权重最大；相对必要集 N 中的属性其次，重要度组合权重比核属性小；不必要集 I 中的属性最不重要，其重要度组合权重最小。这与定义 3 中所描述的各属性所起作用的概念完全一致。该性质的逆性质不一定成立。

5 实例分析

为了验证本文提出的确定权重方法的合理性，下面分别用两种方法来计算权重，并对结果进行分析。

设 $S=(U,C \cup D,V,f)$ 是一个相容决策表，其中 $C=\{a_1,a_2,a_3,a_4\}$ 为条件属性，$D=\{d\}$ 为决策属性，论域是对象的一个非空有限集合 $U=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$，$|U|=7$，如表 1 所列。

表 1 相容决策表 S

U	a_1	a_2	a_3	a_4	d
u_1	2	2	0	1	1
u_2	1	2	0	1	0
u_3	1	1	2	1	1
u_4	0	0	0	1	0
u_5	1	0	1	1	0
u_6	2	0	1	1	1
u_7	2	1	0	0	0

对相容决策表 1 进行运算：

$POS_C(D)=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$。

$POS_{C-a_1}(D)=\{u_3,u_4,u_7\} \neq POS_C(D)$。

$POS_{C-a_2}(D)=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}=POS_C(D)$。

$POS_{C-a_3}(D)=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}=POS_C(D)$。

$POS_{C-a_4}(D)=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}=POS_C(D)$。

由粗糙集理论中约简的相关知识得到决策表的核集为：$CORE(C)=\{a_1\}$，两个约简集分别为 $RED(C)=\{\{a_1,a_2\},\{a_1,a_3\}\}$。

根据定义 3 分别计算得到以下各属性集合：

核集：$CORE(C)=\{a_1\}$。

相对必要集：$N=RED(C)-CORE(C)=\{a_2,a_3\}$。

不必要集：$I=C-RED(C)=\{a_4\}$。

首先，我们采用 Pawlak 属性重要度来确定权重。

根据定义 1，我们可分别计算得到各属性的重要度为：

$$sig(a_1)=\frac{|POS_C(D)|-|POS_{C-\{a_1\}}(D)|}{|U|}=\frac{4}{7} \approx 0.57$$

$$sig(a_2)=\frac{|POS_C(D)|-|POS_{C-\{a_2\}}(D)|}{|U|}=\frac{0}{7}=0$$

同理有：$sig(a_3)=sig(a_4)=0$。

因此，各属性重要性所占的权重分别为：

$$W(a_1)=\frac{a_1}{\sum\limits_{i=1,2,3,4} sig(a_i)}=\frac{0.57}{0.57+0+0+0}=1$$

$$W(a_2)=\frac{a_2}{\sum\limits_{i=1,2,3,4} sig(a_i)}=\frac{0}{0.57+0+0+0}=0$$

同理有 $W(a_3)=W(a_4)=0$。

从以上计算结果我们知道，核属性 a_1 的权重是 1，反映了其分类能力最强，这与事实符合。但是，其它属性的权重都为 0，特别是相对必要集中的属性 a_2 和 a_3，它们所占权重也都是 0，这明显与它们所起的作用不相符。a_2 和 a_3 是约简集中的成员，如果从每个最小约简集中去掉它们，分类和正区域一定会发生变化。例如，对于约简集 $\{a_1,a_2\}$ 来说，由其产生的原始分类为 $\{\{u_1\},\{u_2\},\{u_3\},\{u_4\},\{u_5\},\{u_6\},\{u_7\}\}$，正区域为 $\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$，而从 $\{a_1,a_2\}$ 中删除属性 a_2 后，分类为 $\{\{u_1,u_6,u_7\},\{u_2,u_3,u_5\},\{u_4\}\}$，正区域为 $\{u_4\}$。同样，对于约简集 $\{a_1,a_3\}$，由其产生的原始分类为 $\{\{u_1\},\{u_2\},\{u_3\},\{u_4\},\{u_5\},\{u_6\},\{u_7\}\}$，正区域为 $\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$，从 $\{a_1,a_3\}$ 中删除属性 a_3 后，分类变为 $\{\{u_1,u_6,u_7\},\{u_2,u_3,u_5\},\{u_4\}\}$，正区域变为 $\{u_4\}$。这说明约简集删除相对必要属性后，无论是分类还是正区域都发生了变化，但是，从上述计算结果中得到它们的属性重要度的权重却为 0，也就是说它们没有起作用，有违客观事实。

下面我们采用本文提出的方法来计算各属性所占的权重。根据定义 4，由决策表 1 得到的区分矩阵为：

$$\begin{bmatrix} \varnothing & & & & & & \\ a_1 & \varnothing & & & & & \\ \varnothing & a_1a_2 & \varnothing & & & & \\ a_1a_2 & \varnothing & a_1a_2a_3 & \varnothing & & & \\ a_1a_2a_3 & \varnothing & a_2a_3 & \varnothing & \varnothing & & \\ \varnothing & a_1a_2a_3 & \varnothing & a_1a_3 & a_1 & \varnothing & \\ a_2a_4 & \varnothing & a_1a_3a_4 & \varnothing & \varnothing & a_2a_3a_4 & \varnothing \end{bmatrix}$$

我们先计算各属性在区分矩阵下的重要度。

根据定义 5，可得区分矩阵中非空元素个数为 $K=12$。这里以属性 a_1 为例，它在区分矩阵的非空元素中共出现了 9 次，第一次是以单个元素 a_1 出现，则有 $r_b \cap c_{ij}=\frac{1}{|c_{ij}|}=\frac{1}{|\{a_1\}|}=\frac{1}{1}=1$；第二次是以 $\{a_1,a_2\}$ 出现，即 $r_b \cap c_{ij}=\frac{1}{|c_{ij}|}=\frac{1}{|\{a_1,a_2\}|}=\frac{1}{2}$；第三次也是以 $\{a_1,a_2\}$ 出现，即 $r_b \cap c_{ij}=\frac{1}{2}$；第四次是以 $\{a_1,a_2,a_3\}$ 出现，即 $r_b \cap c_{ij}=\frac{1}{|\{a_1,a_2,a_3\}|}=\frac{1}{3}$，以此类推，得到属性 a_1 重要度为：

$$Nsig(a_1)=\frac{\sum\limits_{i,j=1,2,\cdots,n} r_b \cap c_{ij}}{K}$$

$$=\frac{1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{3}}{12}$$

$$\approx 0.402$$

同理可得：

$$Nsig(a_2)=\frac{\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\frac{1}{2}+\frac{1}{3}+\frac{1}{2}+\frac{1}{3}}{12}$$

$$\approx 0.278$$

$$Nsig(a_3)=\frac{\frac{1}{3}+\frac{1}{3}+\frac{1}{2}+\frac{1}{3}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}}{12}\approx 0.222$$

$$Nsig(a_4)=\frac{\frac{1}{2}+\frac{1}{3}+\frac{1}{3}}{12}\approx 0.097$$

各属性的重要度大小关系为：$Nsig(a_1)>Nsig(a_2)>Nsig(a_3)>Nsig(a_4)>0$，结果与性质 3 得到的结论完成一致。

因此，在区分矩阵下各属性所占的权重分别为：

$$W_N(a_1)=\frac{0.402}{0.402+0.278+0.222+0.097}\approx 0.402$$

$$W_N(a_2)=\frac{0.278}{0.402+0.278+0.222+0.097}\approx 0.278$$

$$W_N(a_3)=\frac{0.222}{0.402+0.278+0.222+0.097}\approx 0.222$$

$$W_N(a_4)=\frac{0.278}{0.402+0.278+0.222+0.097}\approx 0.097$$

各属性所占权重的大小关系为：$W_N(a_1)>W_N(a_2)>W_N(a_3)>W_N(a_4)>0$，结果与性质 5 得到的结论完成一致。

最后，我们来计算各属性重要度所占的组合权重。

由定义 7 可得：

$$W_{new}(a_1)=\frac{W_N(a_1)+W(a_1)}{\sum\limits_{a\in C}(W_N(a)+W(a))}$$

$$=\frac{0.402+1}{0.402+0.278+0.222+0.097+1+0+0+0}$$

$$\approx 0.701$$

同理可得：

$$W_{new}(a_2)=\frac{0.278+0}{0.402+0.278+0.222+0.097+1+0+0+0}$$

$$\approx 0.139$$

$$W_{new}(a_3)=\frac{0.222+0}{0.402+0.278+0.222+0.097+1+0+0+0}$$

$$\approx 0.111$$

$$W_{new}(a_4)=\frac{0.097+0}{0.402+0.278+0.222+0.097+0.57+0+0+0}$$

$$\approx 0.049$$

上述计算结果表明，无论是各属性的重要度大小，还是各属性所占权重大小，都与实际情况相吻合。核属性 a_1 最重要，它所占权重也最大，$a_1=0.701$；相对必要属性 a_2 和 a_3 的重要性其次，它们所占权重 $a_2=0.139$，$a_3=0.111$，比核属性小；不必要属性 a_4 最不重要，所占的权重最小，$a_4=0.049$。即：$W_{new}(a_1)>W_{new}(a_2)>W_{new}(a_3)>W_{new}(a_4)>0$，结果与性质 7 得到的结论完全一致。

结束语 由粗糙集代数表示的属性重要度定义侧重属性的定性分析，它定义的是核属性的重要度，非核属性的重要度都为零，无法确定非核属性重要度所占权重；而信息熵的属性重要程度定义侧重属性的定量分析，但有时会出现非核属性的重要度大于核属性的重要度的情况，这与代数表示的粗糙集属性重要度相矛盾。本文给出的基于区分矩阵属性重要度的定义，不仅反映了核属性重要度，而且体现了非核属性的重要度，在此基础上提出了属性权重构造方法，在一定程度上克服了代数表示法和信息熵表示法的不足。该方法既能对属性定量刻画，又能对属性定性分析，确定的属性权重更加科学，更贴近实际，有一定的实用价值，为人们提高决策精度提供了一种新思路。

参考文献

[1] 徐泽水，陈剑. 一种基于区间直觉判断矩阵的群决策方法[J]. 系统工程理论与实践，2007，4(4)：126-133

[2] 张霄雁，孟祥福，马宗民，等. 基于近似函数依赖的关系数据属性权重评估方法[J]. 计算机科学，2013，40(2)：172-176

[3] 陈晓红，刘益凡. 基于区间数群决策矩阵的专家权重确定方法及其算法实现[J]. 系统工程与电子技术，2010，10(32)：2129-2131

[4] Pawlak Z. Rough sets：Theoretical Aspects of Reasoning about-Data[M]. Kluwer Academic Boston：Publishers，1991

[5] 曹秀英，梁静国. 基于粗集理论的属性权重确定方法[J]. 中国管理科学，2002，10(5)：98-100

[6] 刘盾，胡培. 一种基于粗糙集理论的属性权重构造方法[J]. 系统工程与电子技术，2008，30(8)：1231-1233

[7] 王洪凯，姚炳学，胡海清. 基于粗集理论的权重确定方法[J]. 计算机工程与应用，2003，39(36)：20-21

[8] 孙斌，王立杰. 基于粗糙集理论的权重确定方法研究[J]. 计算机工程与应用，2006，42(29)：216-217

[9] 鲍新中，刘澄. 一种基于粗糙集的权重确定方法[J]. 管理学报，2009，6(6)：729-732

[10] 万俊，邢焕革，张晓晖. 基于熵理论的多属性群决策专家权重的调整算法[J]. 控制与决策，2010，6(25)：907-910

[11] 吴坚，梁昌勇，李文年. 基于主观与客观集成的属性权重求解方法[J]. 系统工程与电子技术，2007，3(29)：384-386

[12] 谭宗凤，徐章艳，王帅. 一种改进的粗糙集权重方法[J]. 计算机工程与应用，2012，48(18)：115-118

[13] 朱红灿，陈能华. 粗糙集条件信息熵权重确定方法的改进[J]. 统计与决策，2011，8(322)：154-156

[14] Pawlak Z. Rough set theory and its applications todata analysis [J]. Cybernetics and Systems：An International Journal，1998，29(1)：661-688

[15] 张文修，吴伟志. 粗糙集理论与方法[M]. 北京：科学出版社，2001

[16] 刘清. Rough 集及 Rough 推理[M]. 北京：科学出版社，2001

[17] 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版，2001

[18] 苗夺谦，李道国. 粗糙集理论、算法与应用[M]. 北京：清华大学出版社，2008：163-190

[19] Skowron A，Swiniarski R，Synak P. Approximation Spaces and Information Granulation[J]. Transaction on Rough Sets III，2005，3400：175-189

[20] Hu X H，Cercone N. Learning in Relational Databases：A Rough Set Approach[J]. Computational Intelligence，1995，11(2)：323-337

[21] 叶东毅，陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报，2002，30(7)：1086-1088