ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 5,Sep. 2025,pp. 1207－1221
DOI:10. 16337/j. 1004-9037. 2025. 05. 008

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于单值中智优势条件熵的增量式属性约简算法

骆公志，王　聪

(南京邮电大学管理学院,南京 210003)

摘　要:在大数据环境下,序决策信息系统中数据的持续增长导致对象间的优势关系动态变化,高效计算属性约简成为亟待解决的关键问题。为此,提出一种增量单值中智优势条件熵,并由此构建了新的增量式属性约简算法。首先,在单值中智序决策信息系统下给出单值中智优势条件熵;随后,针对4种不同类型的新增对象,深入研究了单值中智优势条件熵的增量更新机制,进而根据该更新机制设计了增量式属性约简算法;最后,选取6个具有优势关系的UCI数据集对增量算法与非增量算法的有效性和高效性进行了对比分析。实验结果表明,新给出的增量属性约简算法在保持相同分类精度的条件下,可以显著提升数据处理的计算效率。

关键词:优势条件熵;单值中智粗糙集;增量学习;序决策信息系统;属性约简

中图分类号:TP181　　**文献标志码**:A

Incremental Attribute Reduction Algorithm Based on Single-Valued Medium-Intelligence Dominance Conditional Entropy

LUO Gongzhi, WANG Cong

(School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In the big data environment, the continuous growth of data in the ordered decision information system leads to the dynamic change of the dominance relationship between objects. Efficient calculation of attribute reduction has become a key problem to be solved urgently. Therefore, an incremental single-valued medium-intelligence dominance conditional entropy is proposed, and a new incremental attribute reduction algorithm is constructed accordingly. Firstly, the single-valued medium-intelligence dominance conditional entropy is given under the single-valued medium-intelligence ordered decision information system. Subsequently, for four different types of new objects, the incremental update mechanism of single-valued medium-intelligence dominance conditional entropy is deeply studied, and then an incremental attribute reduction algorithm is designed according to this update mechanism. Finally, six UCI datasets with dominance relations are selected to conduct a comparative experimental analysis on the effectiveness and efficiency of the incremental algorithm and the non-incremental algorithm. Experimental results show that the newly given incremental attribute reduction algorithm can significantly improve the computational efficiency of data processing while maintaining the same classification accuracy.

Key words: entropy of dominant condition; single-valued medium-intelligence rough set; incremental learning; order decision information system; attribute reduction

---

基金项目:国家自然科学基金(72171124);江苏高校哲学社会科学研究重大项目(2021SJZDA129);江苏省教育科学“十四五”规划重点课题(GK-202103);江苏省研究生科研创新计划项目(KYCX23_0936)。
收稿日期:2025-04-08;**修订日期**:2025-07-08

引　言

属性约简，也称特征选择，是机器学习和数据挖掘领域的重要数据预处理技术之一[1]，其主要任务是去除冗余和不相关属性，同时维持或提升数据分类性能。粗糙集理论无需先验知识即可处理不准确和不一致信息，是属性约简的理论基础[2]。近年来，众多基于各种拓展粗糙集模型的属性约简方法被用于解决实际问题，其中模糊粗糙集[3-4]和邻域粗糙集[5-6]应用相对广泛，然而这两类粗糙集模型并不适合处理有序关系的数据集。为此，Greco等[7]提出了基于优势关系的粗糙集模型，以处理具有偏序关系的数据集。随后，王雅辉等[8]将有序数据集用优势集表示，设计出基于模糊优势互补互信息的有序决策树算法。Yang等[9]通过构造分层有序信息粒设计了一种优势关系下的属性约简算法。

在大数据环境下，数据快速变化，传统属性约简算法处理这类动态数据时因迭代运用约简算法，故存在计算耗时久、内存占用大以及难以满足数据持续增长需求等问题，致使计算效率不高。为解决动态数据集下属性约简的可用性难题，众多学者引入了增量式属性约简算法[10-12]，该算法的优势在于仅需关注新变化数据对原有关系的影响，无需重复评估原始数据间的稳定关系，从而有效避免了对原有稳定关系的重复计算。具体而言，增量属性约简聚焦于数据变化给原有关系带来的改变，如确定新数据中样本与原始数据的关系应在新数据添加后进行，减少了对原有数据之间关系的重复计算，提升了属性约简的效率。因此，增量式属性约简算法因其高效利用已有知识加速获取新知识的能力而备受学者关注。闫振超等[13]基于信息粒度的增量更新机制提出两种面向部分标记混合数据的增量式属性约简算法，以加速动态混合型数据集的属性约简。王锋等[14]利用信息融合机制，通过对粒空间及粒结构的变化更新进行分析，提出了面向动态混合数据的多粒度增量特征选择算法。Ding等[15]提出了一种利用属性树的渐进加速方法来减少动态数据集的属性约简时间。Zhang等[16]针对不完整动态数据集提出知识粒度更新的增量属性约简方法。Ni等[17]针对模糊动态数据集提出基于模糊粗糙集的增量特征选择。

然而，现有方法忽略了现实中有序数据随时间动态变化的情况，对于动态有序数据集，当有新对象加入时，原有对象间的优势关系会动态调整，进而原有对象的优势集发生相应变化，故这些增量算法不再适用。为此，Luo等[18]将知识粒度引入序集值决策信息系统，针对系统中属性增加和删除的情况，基于矩阵提出了一种增量更新近似集的方法。Zhang等[19]针对数值型数据的有序信息系统多维变化中动态更新近似集的问题，提出了模糊知识粒度并研究了分别向系统中添加对象和对象集的增量机制和属性约简算法。此外，Sang等[20]从信息观的角度出发，利用邻域优势条件熵的变化设计了序决策信息系统下的增量属性约简算法。随后，模糊优势邻域粗糙集模型[21]被提出，用以解决区间值序决策信息系统的增量属性约简问题。

鉴于一维属性值在表达模糊信息方面的局限性，难以充分应对广泛存在的不确定性问题，学者们进一步提出单值中智集的概念，通过3个独立的隶属度来评估信息系统不一致和不准确的信息，随后研究者将单值中智集引入粗糙集模型，建立了单值中智粗糙集模型以拓展粗糙集理论在处理不确定性问题方面的应用。Wang等[22]基于单值中智数的排序方法，构建了多尺度主导的单值中智粗糙集模型，以更好地处理多尺度下的模糊且不一致的单值中智序决策信息系统。Zheng等[23]进一步将单值中智集拓展为时变单值中智集，以有效描述模糊、不确定且随时间变化的应急事件信息，并基于时变单值中智决策信息系统构建了应急决策模型，更精准地反映实际需求，为决策者提供有力支持。尽管在上述的决策信息系统研究领域利用粗糙集模型处理增量决策信息系统取得了丰硕成果，然而单值中智序决策信息系统中的增量式属性约简问题却仍未得到妥善解决。

鉴于此，本文在单值中智序决策信息系统下提出一种基于单值中智优势条件熵的增量属性约简算

法。首先,在单值中智序决策信息系统下提出单值中智优势条件熵的概念。接着,对新增对象的4种不同类型进行分析,基于这4种不同类型探讨单值中智优势条件熵的更新原理,并据此设计一种新的增量属性约简算法。最后,在6个具有优势关系的UCI数据集上的对比实验表明所提算法能够在保证分类精度的同时提高计算效率。

1 基本知识

1.1 单值中智集

定义1[24] 设$U=\{x_1,x_2,\cdots,x_n\}$为一个非空有限论域,U上的一个单值中智集M可定义为如下形式

$$M=\{\langle x,T_M(x),I_M(x),F_M(x)\rangle\ |x\in U\} \tag{1}$$

式中:$\forall x\in U$,$T_M(x)$、$I_M(x)$、$F_M(x)\in[0,1]$,$T_M(x)$、$I_M(x)$和$F_M(x)$分别为正确隶属度、不确定隶属度和错误隶属度,且$0\leqslant T_M(x)+I_M(x)+F_M(x)\leqslant 3$。对$\forall x\in U$,称$M(x)=(T_M(x),I_M(x),F_M(x))$是集合$M$下的单值中智数。$U$上的所有单值中智数的集合记作SVNS($U$)。

定义2[25] 设$U=\{x_1,x_2,\cdots,x_n\}$为一个非空有限论域,U上的一个单值中智集$M=\{\langle x,T_M(x),I_M(x),F_M(x)\rangle\ |x\in U\}$,该单值中智数的得分函数定义为$Sc(p)=\dfrac{T_M(p)+w_1I_M(p)+(1-w_2I_M(p))+(1-F_M(p))}{4}$,其中$p\in U$为该单值中智数的任一具体对象,$w_1=\dfrac{T_M(p)}{T_M(p)+F_M(p)}$、$w_2=\dfrac{F_M(p)}{T_M(p)+F_M(p)}$为不确定隶属度的权重,$Sc(p)\in[0,1]$,若$Sc(p)$的值越大,则相应的单值中智数$M(p)=(T_M(p),I_M(p),F_M(p))$越大。

1.2 基于得分函数的优势关系

定义3 设决策信息系统为$S=(U,C\cup D,F,G)$,其中$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域,$C=\{a_1,a_2,\cdots,a_r\}$为条件属性集,$D=\{d_1,d_2,\cdots,d_s\}$为决策属性集,$F=\{f_k:U\to V_k,k\leqslant r\}$为$U$与$C$的关系集,其中$V_k$为条件属性$a_k$的有限值域,$G=\{g_f:U\to V_f,f\leqslant s\}$为$U$与$D$的关系集,$V_f$为决策属性$d_f$的有限值域。如果条件属性值为单值中智数$M(x)=(T_M(x),I_M(x),F_M(x))$,那么称该信息系统为单值中智决策信息系统,记为SVNDS$=(U,C\cup D,F,G)$。

在决策信息系统中,当条件属性的取值呈现递增或递减的偏序关系时,该条件属性被视为系统的条件准则。同样地,如果决策属性的值域也呈现递增或递减的偏序关系,那么该决策属性被视为系统的决策准则。若决策信息系统中的条件属性和决策属性均为准则,则该系统称为序决策信息系统。由于递增和递减的偏序关系仅由考察的准则意义决定,并不影响实际研究,因此将重点放在基于递增偏序关系所获得的优势关系的研究上,通过得分函数对单值中智数之间的比较,可以建立单值中智决策信息系统的序关系。

定义4 给定单值中智决策信息系统SVNDS$=(U,C\cup D,F,G)$,其中$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域,$C=\{a_1,a_2,\cdots,a_r\}$为条件属性集,$D=\{d_1,d_2,\cdots,d_s\}$为决策属性集,在SVNDS中根据条件属性子集$A\subseteq C$和决策属性D所定义的序关系为

$$R_A^{\geqslant}=\{(x,y)\in U\times U\,|\,x\geqslant y,\forall a\in A\}=\{(x,y)\in U\times U\,|\,\mathcal{S}c_a(x)\geqslant\mathcal{S}c_a(y),\forall a\in A\} \tag{2}$$

$$R_d^{\geqslant}=\{(x,y)\in U\times U\,|\,g(x,d)\geqslant g(y,d),\forall d\in D\} \tag{3}$$

式中：$\mathcal{S}c_a(x)$表示对象x在属性a上的得分，$g(x,d)$表示对象x的决策属性值，此时称该决策信息系统为单值中智序决策信息系统，记为SVNDS$^{\succcurlyeq}$。$R_A^{\succcurlyeq}$和$R_d^{\succcurlyeq}$称为单值中智序关系也称为单值中智优势关系。

基于单值中智优势关系$R_A^{\succcurlyeq}$和$R_d^{\succcurlyeq}$可以诱导出条件属性子集以及决策属性的单值中智优势类，即

$$[x]_A^{\succcurlyeq}=\{y\in U|(y,x)\in R_A^{\succcurlyeq},\forall a\in A\}=\{y\in U|\mathcal{S}c_a(y)\succcurlyeq\mathcal{S}c_a(x),\forall a\in A\} \tag{4}$$

$$[x]_d^{\succcurlyeq}=\{y\in U|g(y,d)\succcurlyeq g(x,d),\forall d\in D\} \tag{5}$$

性质1[25] 设SVNDS$^{\succcurlyeq}=(U,C\cup D,F,G)$为单值中智序决策信息系统，$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域；$C\cup D\neq\varnothing$，$C$为条件属性集，$D$为决策属性集；$B_1$、$B_2$为条件属性$C$的属性子集；$R_C^{\succcurlyeq}$为属性集$C$下的单值中智优势关系，$[x_i]_C^{\succcurlyeq}$为由单值中智优势关系$R_C^{\succcurlyeq}$诱导出的单值中智优势类，对$x_i$、$x_j\in U$，$1\leqslant i\leqslant n$，$1\leqslant j\leqslant n$，且$i\neq j$，有如下性质：(1)若$B_1\subseteq B_2\subseteq C$，则有$R_C^{\succcurlyeq}\subseteq R_{B_2}^{\succcurlyeq}\subseteq R_{B_1}^{\succcurlyeq}$；(2)若$B_1\subseteq B_2\subseteq C$，则有$[x_i]_A^{\succcurlyeq}\subseteq[x_i]_{B_2}^{\succcurlyeq}\subseteq[x_i]_{B_1}^{\succcurlyeq}$；(3)对于$\forall x_j\in[x_i]_C^{\succcurlyeq}$，有$[x_j]_C^{\succcurlyeq}\subseteq[x_i]_C^{\succcurlyeq}$。

1.3 优势单值中智粗糙集模型

定义5 设SVNDS$^{\succcurlyeq}=(U,C\cup D,F,G)$为单值中智序决策信息系统，$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域；$C\cup D\neq\varnothing$，$C$为条件属性集，$D$为决策属性集；对$\forall A\subseteq C$，$R_A^{\succcurlyeq}$为条件属性子集$A$下的单值中智优势关系；对$\forall X\in U$，关于单值中智优势关系$R_A^{\succcurlyeq}$的下近似和上近似定义为

$$\underline{R_A^{\succcurlyeq}}(X)=\{x_i\in U|[x_i]_A^{\succcurlyeq}\subseteq X,1\leqslant i\leqslant n\} \tag{6}$$

$$\overline{R_A^{\succcurlyeq}}(X)=\{x_i\in U|[x_i]_A^{\succcurlyeq}\cap X\neq\varnothing,1\leqslant i\leqslant n\} \tag{7}$$

式中：$[x_i]_A^{\succcurlyeq}$为单值中智优势关系$R_A^{\succcurlyeq}$诱导出的单值中智优势类，如果$\overline{R_A^{\succcurlyeq_{(\alpha,\beta)}}}(X)=\underline{R_A^{\succcurlyeq_{(\alpha,\beta)}}}(X)$，那么称$X$是可定义的，否则，$X$是不可定义的，即粗糙的，称$[\overline{R_A^{\succcurlyeq}}(X)=\underline{R_A^{\succcurlyeq}}(X)]$为优势单值中智粗糙集。

2 单值中智序决策信息系统下属性约简

文献[26]提出了优势条件熵模型，用以解决信息熵模型在衡量具有偏序关系的数据时的不足。受此启发本文提出了单值中智序决策信息系统下的单值中智优势熵、单值中智优势联合熵以及单值中智优势条件熵的概念，并将单值中智优势条件熵作为属性重要度的度量标准。

定义6(单值中智优势熵) 给定单值中智序决策信息系统SVNDS$^{\succcurlyeq}=(U,C\cup D,F,G)$，$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域；$C\cup D\neq\varnothing$，$C$为条件属性集，$D$为决策属性集；对$\forall A\subseteq C$，条件属性将论域划分为$U/A=\{[x_1],[x_2],\cdots,[x_n]\}$，定义条件属性子集$A$在$U$下的单值中智优势熵为

$$\mathrm{SMDE}_A^{\succcurlyeq}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]_A^{\succcurlyeq}|}{n} \tag{8}$$

式中$|\cdot|$为集合的基数。

定义7(单值中智优势联合熵) 给定单值中智序决策信息系统SVNDS$^{\succcurlyeq}=(U,C\cup D,F,G)$，$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域；$C\cup D\neq\varnothing$，$C$为条件属性集，$D$为决策属性集；对$\forall A,B\subseteq C$，定义条件属性子集$A$和$B$在$U$下的单值中智优势联合熵为

$$\mathrm{SMDE}_{A\cup B}^{\succcurlyeq}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]_A^{\succcurlyeq}\cap[x_i]_B^{\succcurlyeq}|}{n} \tag{9}$$

定义8(单值中智优势条件熵) 给定单值中智序决策信息系统SVNDS$^{\succcurlyeq}=(U,C\cup D,F,G)$，

$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,C为条件属性集,D为决策属性集;对$\forall A$、$B\subseteq C$,定义条件属性子集B关于A在U下的单值中智优势条件熵为

$$\mathrm{SMDE}^{\geqslant}_{B|A}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]^{\geqslant}_A\cap[x_i]^{\geqslant}_B|}{|[x_i]^{\geqslant}_A|}\tag{10}$$

单值中智优势熵衡量了单个变量自身的不确定性程度。单值中智优势联合熵则反映了多个变量共同作用下的不确定性度量。而单值中智优势条件熵则表示在已知某一条件下,变量的不确定性度量。

定理1 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,对于条件属性子集$A,B\subseteq C$,$D$为决策属性集,$\forall d\in D$,则有

$$\mathrm{SMDE}^{\geqslant}_{B|A}(U)=\mathrm{SMDE}^{\geqslant}_{A\cup B}(U)-\mathrm{SMDE}^{\geqslant}_{A}(U)\tag{11}$$

证明 根据定义6可得$\mathrm{SMDE}^{\geqslant}_{A}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]^{\geqslant}_A|}{n}$,根据定义7可得$\mathrm{SMDE}^{\geqslant}_{A\cup B}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]^{\geqslant}_A\cap[x_i]^{\geqslant}_B|}{n}$,进而可得

$$\mathrm{SMDE}^{\geqslant}_{A\cup B}(U)-\mathrm{SMDE}^{\geqslant}_{A}(U)=-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]^{\geqslant}_A\cap[x_i]^{\geqslant}_B|}{n}-\left(-\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[x_i]^{\geqslant}_A|}{n}\right)=$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left(\log_2\frac{|[x_i]^{\geqslant}_A\cap[x_i]^{\geqslant}_B|}{n}-\log_2\frac{|[x_i]^{\geqslant}_A|}{n}\right)=-\frac{1}{n}\sum_{i=1}^{m}\log_2\frac{|[x_i]^{\geqslant}_A\cap[x_i]^{\geqslant}_B|}{|[x_i]^{\geqslant}_A|}=\mathrm{SMDE}^{\geqslant}_{B|A}(U)$$

证毕。

定理1阐明了单值中智优势联合熵、单值中智优势熵与单值中智优势条件熵之间的重要关系。通过定理1可以清晰认识到,单值中智优势条件熵实际上是单值中智优势联合熵与单值中智优势熵之间的差值。

定义9 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$D$为决策属性集;对$\forall A\subseteq C$,定义$\forall a\subseteq A$在$A$下的内部属性重要度为

$$\mathrm{SIG}^{U}_{\mathrm{in}}(a,A,D)=\mathrm{SMDE}^{\geqslant}_{D|A-\{a\}}(U)-\mathrm{SMDE}^{\geqslant}_{D|A}(U)\tag{12}$$

定义10 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$D$为决策属性集;对$\forall A\subseteq C$,定义$\forall a\subseteq A$在$A$下的外部属性重要度为

$$\mathrm{SIG}^{U}_{\mathrm{out}}(a,A,D)=\mathrm{SMDE}^{\geqslant}_{D|A}(U)-\mathrm{SMDE}^{\geqslant}_{D|A\cup\{b\}}(U)\tag{13}$$

定义9和定义10表明内部重要度越高,则该属性在已选属性集里的重要度越高;与内部重要度类似,外部重要度越高,则该属性在已选属性集之外的重要度越高,即可以从已选条件属性集之外来选择重要度高的条件属性。

定义11 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;条件属性子集$R\subseteq C$,若以下两个条件同时满足,则称$R$为该单值中智序决策信息系统的约简集:(1)$\mathrm{SMDE}^{\geqslant}_{D|R}(U)=\mathrm{SMDE}^{\geqslant}_{D|C}(U)$;(2)$\forall a\in R,\mathrm{SMDE}^{\geqslant}_{D|R-\{a\}}(U)>\mathrm{SMDE}^{\geqslant}_{D|C}(U)$。

根据定义9~11,利用单值中智优势条件熵对单值中智序决策信息系统进行属性约简的实现如算法1所示。

算法 1 单值中智优势条件熵属性约简算法(Single-valued medium-intelligence dominant conditional entropy atribute reduction algorithm, SMADA)

输入:单值中智序决策信息系统 $\mathrm{SVNDS}^{\succcurlyeq}=(U, C\cup D, F, G)$

输出:t时刻的单值中智序决策信息系统的约简集 Red_U^t

(1) 初始化约简集 $\mathrm{Red}_U^t\leftarrow\varnothing$,$\mathrm{Red}\leftarrow\varnothing$进入步骤(2);

(2) 根据得分函数$Sc(x)$计算系统所有对象的所有条件属性的得分,进入步骤(3);

(3) 遍历单值中智序决策信息系统的所有对象的条件属性和决策属性,对于每一个对象,根据定义4,将满足条件:$\mathcal{S}c_a(x_j)\geqslant\mathcal{S}c_a(x_i),\forall a\in A$ 的对象加入到该对象的优势类中,求得所有对象的优势类 $[x_i^t]_C^{\succcurlyeq}$,同理求得 $[x_i^t]_D^{\succcurlyeq}$ 和 $[x_i^t]_A^{\succcurlyeq}\cap[x_i^t]_D^{\succcurlyeq}$,进入步骤(4);

(4) 根据定理1求得 $\mathrm{SMDE}_{D|C}^{\succcurlyeq}(U)$,进入步骤(5);

(5) 对于 $\forall a_k\in C$,根据定理1计算 $\mathrm{SMDE}_{D|C-\{a_k\}}^{\succcurlyeq}(U)$,进入步骤(6);

(6) 对于 $\forall a_k\in C$,根据定义9计算所有 $\mathrm{SIG}_{\mathrm{in}}^U(a_k,\mathrm{Red},D)$,如果 $\mathrm{SIG}_{\mathrm{in}}^U(a_k,\mathrm{Red},D)>0$,那么将 a_k 加入 Red,进入步骤(7);

(7) 计算 $\mathrm{SMDE}_{D|\mathrm{Red}}^{\succcurlyeq}(U)$,若 $\mathrm{SMDE}_{D|\mathrm{Red}}^{\succcurlyeq}(U)=\mathrm{SMDE}_{D|C}^{\succcurlyeq}(U)$,进入步骤(9),否则进入步骤(8);

(8) 对于 $\forall a_k\in C-\mathrm{Red}$,根据定义10计算 $\mathrm{SIG}_{\mathrm{out}}^U(a_k,\mathrm{Red},D)$,将使得 $\mathrm{SIG}_{\mathrm{out}}^U(a_k,\mathrm{Red},D)$最大的 a_k 加入 Red,再转入步骤(7);

(9) 将 Red 赋值给 Red_U^t,进入步骤(10);

(10) t时刻的单值中智序决策信息系统的约简集 Red_U^t。

例 1 给定单值中智序决策信息系统 $\mathrm{SVNDS}^{\succcurlyeq}=(U, C\cup D, F, G)$,如表1所示,其中论域 $U=\{x_1,x_2,x_3,x_4,x_5\}$ 分别表示5个对象;条件属性集 $C=\{a_1,a_2,a_3,a_4\}$ 分别表示4个条件属性;$D=\{0,1,2\}$表示3个等级,且优势关系为由低到高排序。

表 1 单值中智序决策信息系统

Table 1 Single-valued medium-intelligence order decision information system

U	a_1	a_2	a_3	a_4	d
x_1	0.3,0.2,0.7	0.3,0.0,0.5	0.3,0.2,0.7	0.3,0.2,0.7	0
x_2	0.3,0.0,0.5	0.0,0.2,0.9	0.3,0.2,0.7	0.0,0.2,0.9	2
x_3	0.0,0.2,0.9	0.3,0.2,0.7	0.3,0.0,0.5	0.3,0.2,0.7	2
x_4	0.6,0.2,0.4	0.3,0.0,0.5	0.6,0.2,0.4	0.3,0.2,0.7	1
x_5	0.3,0.0,0.5	0.0,0.2,0.9	0.3,0.2,0.7	0.0,0.2,0.9	2

首先通过算法1计算约简,步骤如下:

步骤 1 根据得分函数$Sc(x)$的定义计算系统所有对象的所有条件属性的得分,如表2所示。

步骤 2 根据定理1求得 $\mathrm{SMDE}_{D|C}^{\succcurlyeq}(U)=0.434\,0$;

步骤3 根据定理1计算 $\mathrm{SMDE}_{D|C-\{a_1\}}^{\succcurlyeq}(U)=0.494\,8$,$\mathrm{SMDE}_{D|C-\{a_2\}}^{\succcurlyeq}(U)=0.434\,0$,$\mathrm{SMDE}_{D|C-\{a_3\}}^{\succcurlyeq}(U)=0.551\,0$,$\mathrm{SMDE}_{D|C-\{a_4\}}^{\succcurlyeq}(U)=0.434\,0$;

步骤4 根据定义9计算所有 $\mathrm{SIG}_{\mathrm{in}}^U(a_1,\mathrm{Red},D)=$

表 2 单值中智决策信息系统得分表

Table2 Score table of single-valued medium-intelligence decision information system

U	a_1	a_2	a_3	a_4	d
x_1	0.38	0.45	0.38	0.38	0
x_2	0.45	0.23	0.38	0.23	2
x_3	0.23	0.38	0.45	0.38	2
x_4	0.56	0.45	0.56	0.38	1
x_5	0.45	0.23	0.38	0.23	2

$0.060\,8$,$\mathrm{SIG}_{\mathrm{in}}^{U}(a_2,\mathrm{Red},D)=0$,$\mathrm{SIG}_{\mathrm{in}}^{U}(a_3,\mathrm{Red},D)=0.117\,0$,$\mathrm{SIG}_{\mathrm{in}}^{U}(a_4,\mathrm{Red},D)=0$,因此$\mathrm{Red}=\{a_1,a_3\}$;

步骤5 计算$\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U)=\mathrm{SMDE}^{\geqslant}_{D|C}(U)=0.434\,0$;

步骤6 输出约简集$\mathrm{Red}=\{a_1,a_3\}$。

3 单值中智序决策信息系统下增量式属性约简

在单值中智序决策信息系统中,当有新的对象流入时,对象的条件优势集和决策优势集都会随之改变,原有的属性约简结果不再有效,而重新计算新单值中智序决策信息系统的属性约简时间复杂度较高,因此本文设计了一种增量属性约简算法。

定理2 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$\forall a\in A\subseteq C$,$D$为决策属性集,$\forall d\in D$;在$t+1$时,单值中智序决策信息系统中增加一个新的对象$x^+$,即$U_{t+1}=U_t\cup\{x^+\}$,则增加后在$t$时对象$x_i$的条件优势集和决策优势集在$t+1$时分别为

$$[x_i^{t+1}]_A^{\geqslant}=\begin{cases}[x_i^t]_A^{\geqslant} & f(x_i^t,a)>f(x^+,a)\\ [x_i^t]_A^{\geqslant}\cup\{x^+\} & f(x_i^t,a)\leqslant f(x^+,a)\end{cases}\tag{14}$$

$$[x_i^{t+1}]_A^{\geqslant}\cap[x_i^{t+1}]_D^{\geqslant}=\begin{cases}[x_i^t]_A^{\geqslant}\cap[x_i^t]_D^{\geqslant} & [f(x_i^t,a)>f(x^+,a)]\wedge f(x_i^t,d)>f(x^+,d)\\ [x_i^t]_A^{\geqslant}\cap[x_i^t]_D^{\geqslant}\cup\{x^+\} & f(x_i^t,a)\leqslant f(x^+,a)\wedge f(x_i^t,d)>f(x^+,d)\end{cases}\tag{15}$$

式中$[x_i^t]_a^{\geqslant}$表示在t时对象x_i相对于属性a的优势集。

证明 首先证明$[x_i^{t+1}]_A^{\geqslant}$:

(1)对$\forall a\in A\subseteq C$,当$f(x_i^t,a)>f(x^+,a)$时,因为随着时间变化,原对象的属性值未发生变化,即$f(x_i^t,a)=f(x_i^{t+1},a)$,所以$f(x_i^{t+1},a)>f(x^+,a)$,进而由定义4可知$x^+\notin[x_i^{t+1}]_A^{\geqslant}$,故$[x_i^{t+1}]_A^{\geqslant}=[x_i^t]_A^{\geqslant}$;

(2)对$\forall a\in A\subseteq C$,当$f(x_i^t,a)\leqslant f(x^+,a)$时,因为随着时间变化,原对象的属性值未发生变化,即$f(x_i^t,a)=f(x_i^{t+1},a)$,所以$f(x_i^{t+1},a)\leqslant f(x^+,a)$,进而由定义4可知$x^+\in[x_i^{t+1}]_A^{\geqslant}$,故$[x_i^{t+1}]_A^{\geqslant}=[x_i^t]_A^{\geqslant}\cup\{x^+\}$。

$[x_i^{t+1}]_A^{\geqslant}\cap[x_i^{t+1}]_D^{\geqslant}$同理可证,证毕。

定理2描述了在新增对象之后,对象x_i的条件优势集、条件优势集与决策优势集的交集在$t\sim t+1$时变化情况,即定理2说明了如何通过对新增对象的影响来更新$t+1$时的条件优势集和决策优势集结构。

在$t+1$时刻,向单值中智序决策信息系统中加入一个新对象后,系统中原有对象的条件优势集与决策优势集的数量变动情形可归纳为以下4种类型:

(1)若新加入对象x^+在条件属性子集$A\subseteq C$下的得分和决策值均低于论域中原有对象的得分和决策值,则原有对象的条件优势集和决策优势集均保持不变,即此时对象在$A\subseteq C$下的单值中智优势条件熵保持不变;

(2)若新加入对象x^+在条件属性子集$A\subseteq C$下的得分和决策值均高于论域中原有对象的得分和决策值,则原有对象的条件优势集数目和决策优势集均需改变,因此需要调整论域中所有对象的条件优势集和决策优势集并计算新增加对象的优势集以更新单值中智优势条件熵;

(3)若新加入对象x^+在条件属性子集$A\subseteq C$下的得分高于论域中原有对象的得分,而决策值低于论域中原有对象的决策值,则原有对象的条件优势集发生变化,因此需要调整此部分对象的条件优势集并计算新增加对象的条件优势集以更新单值中智优势条件熵,而决策优势集保持不变;

(4)若新加入对象x^+在条件属性子集$A\subseteq C$下的得分低于论域中原有对象的得分,而决策值高于论域中原有对象的决策值,则原有对象的决策优势集发生变化,因此需要调整此部分对象的决策优势集并计算新增加对象的决策优势集以更新单值中智优势条件熵,而条件优势集保持不变。

定理3 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$\forall a\in A\subseteq C$,$D$为决策属性集,$\forall d\in D$;假设$t+1$时,单值中智序决策信息系统中增加一个新的对象$x^+$,即$U_{t+1}=U_t\cup\{x^+\}$,此时增量单值中智优势条件熵为

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=-\frac{1}{n+1}\left(\sum_{i=1}^{p}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}+\sum_{j=1}^{q}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}+\right.$$
$$\left.\sum_{k=1}^{r}\log_2\frac{|[x_k^t]^{\geqslant}_A\cap[x_k^t]^{\geqslant}_D|}{|[x_k^t]^{\geqslant}_A|+1}+\sum_{l=1}^{s}\log_2\frac{|[x_l^t]^{\geqslant}_A\cap[x_l^t]^{\geqslant}_D|}{|[x_l^t]^{\geqslant}_A|}+\log_2\frac{|[x^+]^{\geqslant}_A\cap[x^+]^{\geqslant}_D|}{|[x^t]^{\geqslant}_A|}\right)\tag{16}$$

式中:$[x_i^t]^{\geqslant}_a$表示在t时对象x_i相对于属性a的优势集,p表示条件属性的得分和决策值都高于新加入的x^+的对象个数,q表示条件属性的得分和决策值都低于新加入的x^+的对象个数,r表示条件属性的得分低于而决策值高于新加入的x^+的对象个数,s表示条件属性的得分高于而决策值低于新加入的x^+的对象个数。

证明 假设$t+1$时,单值中智序决策信息系统中增加的新对象x^+,位于最优属性值和最劣属性值之间,则论域中原有的n个对象与对象x^+存在以下4种关系,分别是条件属性和决策属性都优于x^+的数目,条件属性和决策属性都劣于x^+的数目,条件属性劣于而决策属性优于x^+的数目,条件属性优于而决策属性劣于x^+的数目,分别用p、q、r、s表示,$p+q+r+s=n$。

(1)对于p个条件属性的得分和决策值都高于x^+的原有对象,条件优势集和决策优势集没有发生改变,进而由定理1可知该部分的增量单值中智优势条件熵为

$$\frac{1}{p}\sum_{i=1}^{p}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}\tag{17}$$

(2)对于q个条件属性的得分和决策值都低于x^+的原有对象,条件优势集和决策优势集中都会增加对象x^+,进而由定理1可知该部分的增量单值中智优势条件熵为

$$\frac{1}{q}\sum_{j=1}^{q}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}\tag{18}$$

(3)对于r个条件属性的得分低于而决策值高于x^+的原有对象,条件优势集中会增加对象x^+而决策优势集没有发生改变,进而由定理1可知该部分的单值中智优势条件熵为

$$\frac{1}{r}\sum_{k=1}^{r}\log_2\frac{|[x_k^t]^{\geqslant}_A\cap[x_k^t]^{\geqslant}_D|}{|[x_k^t]^{\geqslant}_A|+1}\tag{19}$$

(4)对于s个条件属性的得分高于而决策值低于x^+的原有对象,条件优势集没有发生改变而决策优势集中会增加对象x^+,进而由定理1可知对于此部分的增量单值中智优势条件熵为

$$\frac{1}{s}\sum_{l=1}^{s}\log_2\frac{|[x_l^t]^{\geqslant}_A\cap\{[x_l^t]^{\geqslant}_D+1\}|}{|[x_l^t]^{\geqslant}_A|}\tag{20}$$

进而根据定义6和定理2可以计算此时增量单值中智优势条件熵为

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=-\frac{1}{p+q+r+s+1}\left(\sum_{i=1}^{p}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}+\sum_{j=1}^{q}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}+\right.$$
$$\sum_{k=1}^{r}\log_2\frac{|[x_k^t]^{\geqslant}_A\cap[x_k^t]^{\geqslant}_D|}{|[x_k^t]^{\geqslant}_A|+1}+\sum_{l=1}^{s}\log_2\frac{\left|[x_l^t]^{\geqslant}_A\cap\{[x_l^t]^{\geqslant}_D+1\}\right|}{|[x_l^t]^{\geqslant}_A|}+$$
$$\left.\log_2\frac{|[x^+]^{\geqslant}_A\cap[x^+]^{\geqslant}_D|}{|[x^t]^{\geqslant}_A|}\right)=-\frac{1}{n+1}\left(\sum_{i=1}^{p}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}+\right.$$
$$\sum_{j=1}^{q}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}+\sum_{k=1}^{r}\log_2\frac{|[x_k^t]^{\geqslant}_A\cap[x_k^t]^{\geqslant}_D|}{|[x_k^t]^{\geqslant}_A|+1}+$$
$$\left.\sum_{l=1}^{s}\log_2\frac{\left|[x_l^t]^{\geqslant}_A\cap\{[x_l^t]^{\geqslant}_D+1\}\right|}{|[x_l^t]^{\geqslant}_A|}+\log_2\frac{|[x^+]^{\geqslant}_A\cap[x^+]^{\geqslant}_D|}{|[x^t]^{\geqslant}_A|}\right)$$

证毕。

定理3给出了增量单值中智优势条件熵的具体表达式,表明了增量条件下优势熵的变化规律,为增量式属性约简算法的设计提供了理论基础。下面基于两种特殊情况给出两个推论。

推论1 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$\forall a\in A\subseteq C$,$D$为决策属性集,$\forall d\in D$;当$t+1$时,在单值中智序决策信息系统中加入一个新的对象$x^+$,且原有的所有对象的条件属性的得分和决策值都高于$x^+$的条件属性的得分和决策值,此时增量单值中智优势条件熵为

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=\frac{n}{n+1}\mathrm{SMDE}^{\geqslant}_A(U)\tag{21}$$

证明 因为原有的所有对象的条件属性的得分和决策值都高于新增对象的条件属性的得分和决策值,所以由定理3可知,对于n个条件属性的得分和决策值都高于x^+的原有对象的条件属性的得分和决策值,条件优势集和决策优势集均保持不变,即$p=n,q,r,s=0$,进而得到

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=-\frac{1}{n+1}\left(\sum_{i=1}^{n}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}+\log_2\frac{|[x^+]^{\geqslant}_A\cap[x^+]^{\geqslant}_D|}{|[x^t]^{\geqslant}_A|}\right)=\frac{n}{n+1}\times\left(-\frac{1}{n}\right)\times$$
$$\left(\sum_{i=1}^{n}\log_2\frac{|[x_i^t]^{\geqslant}_A\cap[x_i^t]^{\geqslant}_D|}{|[x_i^t]^{\geqslant}_A|}+\log_2 1\right)=\frac{n}{n+1}\mathrm{SMDE}^{\geqslant}_A(U)$$

证毕。

推论2 给定单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$U=\{x_1,x_2,\cdots,x_n\}$为非空有限论域;$C\cup D\neq\varnothing$,$C$为条件属性集,$\forall a\in A\subseteq C$,$D$为决策属性集,$\forall d\in D$;当$t+1$时,在单值中智序决策信息系统中加入一个新的对象$x^+$,且原有的所有对象的条件属性的得分和决策值都低于$x^+$的条件属性的得分和决策值,此时增量单值中智优势条件熵为

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=-\frac{1}{n+1}\sum_{j=1}^{q}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}\tag{22}$$

证明 因为原有的所有对象的条件属性的得分和决策值都低于新增对象的条件属性的得分和决策值,所以由定理3可知,对于q个条件属性的得分和决策值都低于x^+的原有对象条件属性的得分和决

策值,条件优势集和决策优势集中都会增加对象x^+,即$q=n,p,r,s=0$,进而得到

$$DH^{\geqslant}_{D|A}(U\cup\{x^+\})_{t+1}=-\frac{1}{n+1}\left(\sum_{j=1}^{n}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}+\log_2\frac{|[x^+]^{\geqslant}_A\cap[x^+]^{\geqslant}_D|}{|[x^t]^{\geqslant}_A|}\right)=$$

$$-\frac{1}{n+1}\left(\sum_{j=1}^{n}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}+\log_2 1\right)=-\frac{1}{n+1}\sum_{j=1}^{n}\log_2\frac{|[x_j^t]^{\geqslant}_A\cap[x_j^t]^{\geqslant}_D|+1}{|[x_j^t]^{\geqslant}_A|+1}$$

证毕。

算法2 单值中智优势条件熵增量属性约简算法(Incremental SMADA, ISMADA)

输入:单值中智序决策信息系统$\mathrm{SVNDS}^{\geqslant}=(U,C\cup D,F,G)$,$\mathrm{Red}_U^t$,新增对象$x^+$

输出:$t+1$时刻的单值中智序决策信息系统的约简集$\mathrm{Red}_{U\cup\{x^+\}}^{t+1}$

(1) 初始化约简集$\mathrm{Red}\leftarrow\mathrm{Red}_U^t$进入步骤(2);

(2) 同算法1的步骤(3,4)计算$t+1$时刻的$\mathrm{SMDE}^{\geqslant}_{D|C}(U\cup x^+)$和$\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)$,进入步骤(3);

(3) 判断$\mathrm{SMDE}^{\geqslant}_{D|C}(U\cup x^+)$和$\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)$是否相等,如果相等进入步骤(5),否则进入步骤(4);

(4) 对于$\forall a_k\in C-\mathrm{Red}$,根据定义10计算$\mathrm{SIG}^U_{\mathrm{out}}(a_k,\mathrm{Red},D)$,将使得$\mathrm{SIG}^U_{\mathrm{out}}(a_k,\mathrm{Red},D)$最大的$a_k$加入Red,再转回步骤(3);

(5) 将Red赋值给$\mathrm{Red}_{U\cup\{x^+\}}^{t+1}$。进入步骤(6);

(6) 输出$t+1$时刻的单值中智序决策信息系统的约简集$\mathrm{Red}_{U\cup\{x^+\}}^{t+1}$。

算法时间复杂度分析:在算法1中,步骤(2)计算所有对象的得分的时间复杂度为$O(|U||C|)$,步骤(3)求解条件优势类和决策优势类的时间复杂度都为$O\left(|U|^2\right)$,步骤(4~10)计算约简集的时间复杂度为$O\left(|U|^2|C|\right)$,因此算法1的时间复杂度为$O\left(|U|^2|C|\right)$。而算法2是基于算法1提出的,其只需关注新加入对象对论域中原有对象产生的影响来更新约简结果,故时间复杂度为$O(|U||C|)$。与算法1相比,计算的时间复杂度明显降低,进而显著提升属性约简的计算效率。

算法空间复杂度分析:在算法1中,原始数据存储需存储论域U中对象的条件属性与决策属性值,空间复杂度为$O(n\times(|C|+1))$,每次加入对象都需要重新计算所有对象的优势类,所以优势类存储空间复杂度为$O(n^2)$,因此算法1的总空间复杂度为$O(n\times(|C|+1)+n^2)$。而算法2是通过增量更新机制,仅需存储新增对象与受影响原有对象的优势类变化,实际动态空间复杂度为$O(kn)$,其中k为新增对象且$k\leqslant n$,故总空间复杂度为$O(n\times(|C|+1)+kn)$。与算法1相比,计算的空间复杂度明显降低,在数据持续增长场景中内存占用更低。

例2 在例1基础上,新加入对象,继续通过算法2计算约简。

步骤1 接算法1计算$t+1$时刻的$\mathrm{SMDE}^{\geqslant}_{D|C}(U\cup x^+)=0.459\,1$和$\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)=0.415\,3$,并判断$\mathrm{SMDE}^{\geqslant}_{D|C}(U\cup x^+)\neq\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)$;

步骤2 根据定义10计算$\mathrm{SIG}^U_{\mathrm{out}}(a_2,\mathrm{Red},D)=-0.043\,8$,$\mathrm{SIG}^U_{\mathrm{out}}(a_4,\mathrm{Red},D)=-0.043\,8$,将$a_2$加入Red;

步骤3 计算$\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)$并判断$\mathrm{SMDE}^{\geqslant}_{D|C}(U\cup x^+)=\mathrm{SMDE}^{\geqslant}_{D|\mathrm{Red}}(U\cup x^+)$;

步骤4 输出约简集$\mathrm{Red}_{U\cup\{x^+\}}^{t+1}=\mathrm{Red}=\{a_1,a_2,a_3\}$。

4 实验与结果分析

本节通过将单值中智优势条件熵属性约简算法SMADA与单值中智优势条件熵增量属性约简算法ISMADA进行对比实验,以验证ISMADA算法的可行性和有效性。实验硬件环境配置为CPU Intel(R) Core(TM) i7-8750H 2.20 GHz,16.0 GB内存,操作系统为Window10(64位)的个人计算机,算法运行的软件环境为Matlab R2020a。

4.1 实验数据集

从UCI机器学习数据库选取6个具有优势关系的数据集进行算法的性能测试,数据集如表3所示。由于数据集都是单一属性值的决策信息系统,因此实验前需要对原始数据进行预处理,以构造单值中智序决策信息系统,实验采用的隶属度计算方法参考了文献[27]提出的多尺度单值中智系统构建框架。首先,计算各数据集中各属性值与该属性列的最大值、平均值和最小值之间的欧式距离。其次,对得到的数据进行归一化处理。最后,将处理后的结果均减去一,以此作为单值中智序决策信息系统中各条件属性的正确隶属度、不确定隶属度和错误隶属度。这种处理方法的合理性在于:从几何意义来看,属性值与最大值的距离可反映其"优越性",对应正确隶属度,与最小值的距离体现"劣势性",对应错误隶属度,而与平均值的距离则表征"偏离度",对应犹豫度;在统计特性方面,最大值、平均值和最小值构成了属性列的数据分布骨架,能够有效捕捉数据的整体特征;从归一化处理角度,通过距离比值消除量纲影响,既确保了$0\leqslant T_M(x)+I_M(x)+F_M(x)\leqslant 3$,又符合单值中智数的数学定义。

表3 实验数据集

Table 3 Experimental dataset

序号	数据集	对象数	属性数	类别数
1	Postoperative	87	8	3
2	Iris	150	4	3
3	Wine	178	13	3
4	Dermatology	358	34	6
5	Avila	20 867	10	12
6	Car	1 728	6	4

4.2 约简结果比较

将各数据集按对象数随机分为5等份,第1份作为基础数据集,随后逐步添加剩余4个数据集。在此过程中,分别使用SMADA算法与ISMADA算法来进行属性约简,比较此过程中对数据集属性约简结果的分类精度及算法运行时间。针对所获取的属性约简结果,采用十折交叉验证策略,在支持向量机(Support vector machine, SVM)与K近邻(K-nearest neighbor, KNN)分类器上分别评估分类性能,两种算法得到的结果如表4、5所示。

表4 两种算法在6个数据集属性约简结果比较

Table 4 Comparison of attribute reduction results of two algorithms in six datasets

数据集	SMADA	ISMADA
Postoperative	3,4,6,8	3,4,6,8
Iris	1,3,4	1,3,4
Wine	1,2,3,4,5,6,7,10,12,13	1,2,3,5,6,7,10,12,13
Dermatology	1,5,7,9,14,16,17,20,21,22,25,26,28,29,31,33	1,5,7,9,14,16, 20,21,22,26,28,29,31,33
Avila	1,2,4,5,7,8,10	1,2,4,5,7,10
Car	1,2,4,5,6	1,2,4,5,6

表5 分类精度比较

Table 5 Comparison of classification accuracy %

数据集	原始数据集		SMADA		ISMADA	
	SVM	KNN	SVM	KNN	SVM	KNN
Postoperative	82.6	83.3	86.4	82.3	86.4	82.3
Iris	89.3	88.1	91.3	92.2	91.3	92.2
Wine	79.6	81.2	88.5	90.2	**89.3**	**91.2**
Dermatology	72.3	73.8	78.3	77.9	**82.4**	**81.5**
Avila	88.9	86.5	90.2	89.1	**90.6**	**90.3**
Car	87.5	84.9	91.4	88.6	91.4	88.6

由表4中的属性约简结果可知,在所有数据集下,ISMADA算法和SMADA算法的约简结果极为相近,二者均能够从原始属性中去除不必要的属性,表明ISMADA算法能有效地实现属性约简。从表5分类精度方面来看,在SVM与KNN分类器下,ISMADA算法约简后的属性集的平均分类精度与SMADA算法相比,呈现持平或有所提升的态势。因此,从实验结果进行分析,所提出的ISMADA约简算法在解决单值中智序决策信息系统中有新对象流入时属性约简结果集的更新问题上,展现出了显著的有效性。数据集Postoperative、Iris和Car下两种算法的属性约简结果完全一致,而在数据集Wine、Dermatology和Avila中,ISMADA算法的属性约简结果中条件属性个数稍少一些,且ISMADA算法分类精度相对更高。该情况的产生主要归因于两种属性约简算法运行机制的差异性。具体而言,SMADA算法与ISMADA算法均属于启发式属性约简方法,其核心差异在于搜索策略的不同,即SMADA算法在面对更新后的数据集时,采用从空集开始逐步添加属性的贪婪搜索策略,每次迭代选择当前最优属性直至满足终止条件。这种策略虽能保证约简的完备性,但由于缺乏对历史信息的利用,可能因初始搜索路径的随机性而陷入局部最优,导致约简结果包含冗余属性。与之不同,ISMADA算法基于增量学习机制,在原有约简结果集的基础上进行动态调整,通过继承历史状态减少对非必要属性的重复评估。从本质上优化了启发式搜索的路径,使得在数据更新过程中ISMADA算法能更有效地识别并保留对分类贡献最大的核心属性,从而生成更精简的约简集。这种差异并非源于属性重要性的动态变化,而是启发式算法在搜索空间和优化目标上的固有特性,即SMADA算法的约简结果更依赖于初始属性排序,而ISMADA算法通过历史信息继承降低了这种依赖性,因此在多数情况下能获得更小的约简集规模。

综上所述,对于单值中智序决策信息系统中对象增加的情形,ISMADA算法在约简的分类精度方面更加占据优势,这主要是由于增量式算法能够有效提取旧单值中智序决策信息系统中保留的原始信息,提升新信息系统的分类性能。

4.3 约简时间比较

为评估ISMADA算法在运行时间上的优势,统计在单值中智序决策信息系统中加入不同比例的对象时ISMADA算法和SMADA算法的运行时间,并将实验结果直观地展示在图1中。从实验结果可以看出,随着数据集中对象数量的持续增加,ISMADA算法和SMADA算法属性约简时的运行时间均有所增加,但相比SMADA算法运行时间的大幅增加,ISMADA算法的运行时间增幅极小,这表明ISMADA算法能极大地降低计算复杂度,减少有新对象加入时属性约简所需时间,进而提高属性约简的效率。

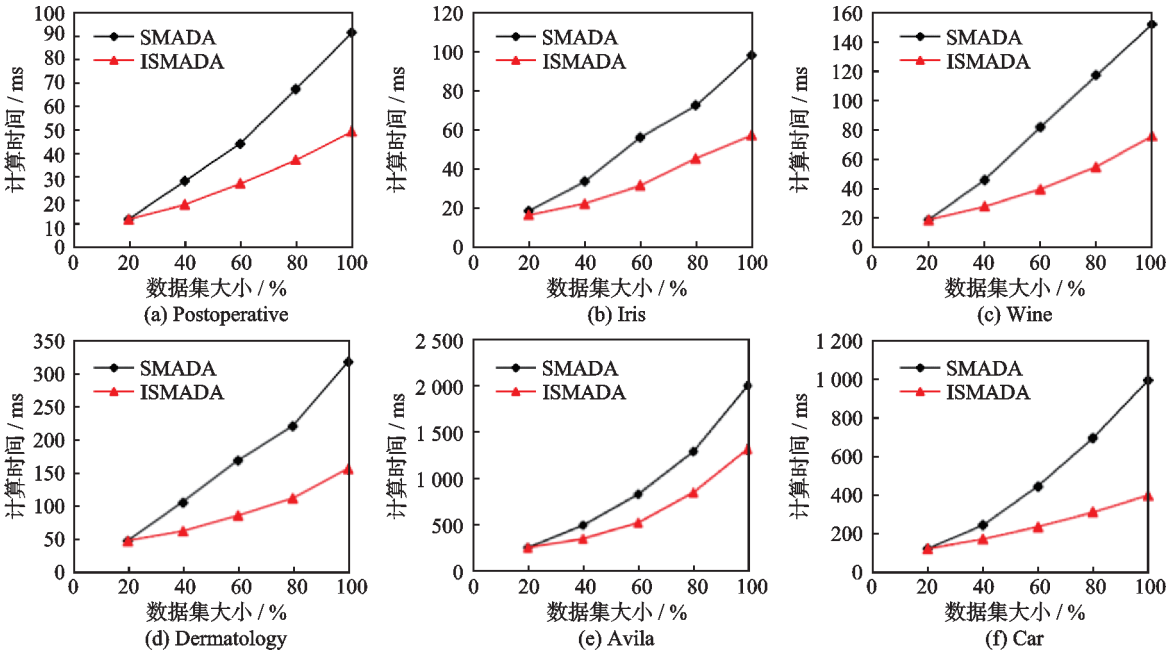

图1 SMADA和ISMADA属性约简的时间消耗对比

Fig.1 Time consumption comparison of attribute reduction for SMADA and ISMADA algorithms

5 结束语

为满足单值中智序决策信息系统中数据动态增加的需求,本文提出一种基于增量单值中智优势条件熵的增量式属性约简算法。首先定义了一种新的单值中智优势条件熵,接着通过分析新增对象与系统原有对象的4种不同优势关系,提出了单值中智条件熵的增量更新机制,并构造增量单值中智优势条件熵以设计增量属性约简算法。在6个数据集上的实验结果表明,所提算法在保持高分类精度的同时,显著降低了算法运行时间。然而,当前算法在处理新增对象时,仍需与已有对象集合进行逐一对比,存在一定的计算冗余。下一步,将聚焦于优化增量处理的存储与查询机制,计划采用哈希表结构存储数据对象的优势类信息,通过预处理优势关系的关键特征构建高效索引,使新增对象可通过哈希映射快速定位匹配的优势类集合,避免对全量数据进行重复对比,从而进一步减少增量过程中的计算冗余。

参考文献:

[1] MA X, WANG J, YU W, et al. Attribute reduction of hybrid decision information systems based on fuzzy conditional information entropy[J]. Computers, Materials and Continua, 2024, 79(2): 2063-2083.

[2] 张维, 苗夺谦, 高灿, 等. 基于粗糙集成学习的半监督属性约简[J]. 小型微型计算机系统, 2016, 37(12): 2727-2732.
ZHANG Wei, MIAO Duoqian, GAO Can, et al. Semi-supervised data attribute reduction based on rough-subspace ensemble learning[J]. Journal of Chinese Computer Systems, 2016, 37(12): 2727-2732.

[3] 庞文莉, 于潇, 郑宇, 等. 基于后悔理论的多粒度直觉模糊三支决策模型[J]. 数据采集与处理, 2025, 40(2): 501-516.
PANG Wenli, YU Xiao, ZHENG Yu, et al. Multi-granularity intuitionistic fuzzy three-way decision model based on regret theory[J]. Journal of Data Acquisition and Processing, 2025, 40(2): 501-516.

[4] LING Z, QI G, MIN H. Intrusion detection using rough-fuzzy set and parallel quantum genetic algorithm[J]. Journal of High Speed Networks, 2024, 30(1): 69-81.

[5] 张万祥, 张贤勇, 杨霁琳, 等. 基于决策代价融合度量的不完备邻域决策粗糙集属性约简[J]. 数据采集与处理, 2025, 40

(3): 807-820.
ZHANG Wanxiang, ZHANG Xianyong, YANG Jilin, et al. Attribute reduction of incomplete neighborhood decision rough sets based on decision-cost fusion measures[J]. Journal of Data Acquisition and Processing, 2025, 40(3): 807-820.

[6] 印振宇, 王平心, 杨习贝, 等. 融合粒度分组与Pareto最优的属性选择[J]. 控制与决策, 2024, 39(9): 2959-2968.
YIN Zhenyu, WANG Pingxin, YANG Xibei, et al. Attribute selection based on granularity grouping and Pareto optimality[J]. Control and Decision, 2024, 39(9): 2959-2968.

[7] GRECO S, MATARAZZO B, SŁOWIŃSKI R. Dominance-based rough set approach as a proper way of handling graduality in rough set theory[M]//Transaction on Rough Sets Ⅶ. Berlin, Heidelberg: Springer-Verlay, 2007: 36-52.

[8] 王雅辉, 钱宇华, 刘郭庆. 基于模糊优势互补互信息的有序决策树算法[J]. 计算机应用, 2021, 41(10): 2785-2792.
WANG Yahui, QIAN Yuhua, LIU Guoqing. Ordinal decision tree algorithm based on fuzzy advantage complementary mutual information[J]. Journal of Computer Applications, 2021, 41(10): 2785-2792.

[9] YANG S, SHI G, ZHANG Y. Acquisition of representative objects and attribute reductions based on generalized decisions of dominance-based rough set approach[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108080.

[10] YANG L, QIN K, SANG B, et al. A novel incremental attribute reduction by using quantitative dominance-based neighborhood self-information[J]. Knowledge-Based Systems, 2023, 261: 110200.

[11] 秦廷桢, 丁卫平, 鞠恒荣, 等. 基于属性树的并行化增量式动态属性约简算法[J]. 模式识别与人工智能, 2022, 35(10): 939-951.
QIN Tingzhen, DING Weiping, JU Hengrong, et al. Parallel incremental dynamic attribute reduction algorithm based on attribute tree[J]. Pattern Recognition and Artificial Intelligence, 2022, 35(10): 939-951.

[12] XU W, YANG Y, DING Y, et al. Incremental feature selection approach to multi-dimensional variation based on matrix dominance conditional entropy for ordered data set[J]. Applied Intelligence, 2024, 54(6): 4890-4910.

[13] 闫振超, 舒文豪, 谢昕. 动态部分标记混合数据的增量式特征选择算法[J]. 计算机科学, 2022, 49(11): 98-108.
YAN Zhenchao, SHU Wenhao, XIE Xin. Incremental feature selection algorithm for dynamic partially labeled hybrid data[J]. Computer Science, 2022, 49(11): 98-108.

[14] 王锋, 姚珍, 梁吉业. 面向动态混合数据的多粒度增量特征选择算法[J]. 软件学报, 2025, 36(3): 1186-1201.
WANG Feng, YAO Zhen, LIANG Jiye. Multi-granulation incremental feature selection algorithm for dynamic hybrid data[J]. Journal of Software, 2025, 36(3): 1186-1201.

[15] DING W, QIN T, SHEN X, et al. Parallel incremental efficient attribute reduction algorithm based on attribute tree[J]. Information Sciences, 2022, 610: 1102-1121.

[16] ZHANG C, DAI J, CHEN J. Knowledge granularity based incremental attribute reduction for incomplete decision systems[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(5): 1141-1157.

[17] NI P, ZHAO S, WANG X, et al. Incremental feature selection based on fuzzy rough sets[J]. Information Sciences, 2020, 536: 185-204.

[18] LUO C, LI T, CHEN H, et al. Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values[J]. Information Sciences, 2015, 299: 221-242.

[19] ZHANG C, LU Z, DAI J. Incremental attribute reduction for dynamic fuzzy decision information systems based on fuzzy knowledge granularity[J]. Information Sciences, 2025, 689: 121467.

[20] SANG B, CHEN H, YANG L, et al. Incremental attribute reduction approaches for ordered data with time-evolving objects [J]. Knowledge-Based Systems, 2021, 212: 106583.

[21] SANG B, CHEN H, YANG L, et al. Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(6): 1683-1697.

[22] WANG W, HUANG B, WANG T. Optimal scale selection based on multi-scale single-valued neutrosophic decision-theoretic rough set with cost-sensitivity[J]. International Journal of Approximate Reasoning, 2023, 155: 132-144.

[23] ZHENG J, WANG Y M, ZHANG K. Dynamic case-based emergency decision-making model under time-varying single-valued neutrosophic set[J]. Expert Systems with Applications, 2024, 250: 123830.

[24] GARG H. A new exponential-logarithm-based single-valued neutrosophic set and their applications[J]. Expert Systems with

Applications, 2024, 238: 121854.

[25] ZHANG X, BO C, SMARANDACHE F, et al. New inclusion relation of neutrosophic sets with applications and related lattice structure[J]. International Journal of Machine Learning and Cybernetics, 2018, 9(10): 1753-1763.

[26] 桑彬彬, 杨留中, 陈红梅, 等. 优势关系粗糙集增量属性约简算法[J]. 计算机科学, 2020, 47(8): 137-143.
SANG Binbin, YANG Liuzhong, CHEN Hongmei, et al. Incremental attribute reduction algorithm in dominance-based rough set[J]. Computer Science, 2020, 47(8): 137-143.

[27] 王文珏, 黄兵. 多尺度单值中智系统中基于优势粗糙集模型的最优尺度选择与约简[J]. 南京大学学报(自然科学), 2022, 58(3): 495-505.
WANG Wenjue, HUANG Bing. Optimal scale selection and reduction based on dominant rough set model in multi-scale single-valued neutrosophic systems[J]. Journal of Nanjing University (Natural Science), 2022, 58(3): 495-505.

作者简介:

骆公志(1972-),**通信作者**,男,教授,研究方向:粗糙集理论及应用,E-mail:lgzyg@163.com。

王聪(1999-),男,硕士研究生,研究方向:粗糙集理论及应用,E-mail:njuptwc@163.com。

(编辑:张黄群)