Advances in Applied Mathematics 应用数学进展, 2024, 13(5), 2172-2179
Published Online May 2024 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2024.135206

Hans汉斯

基于依赖度的时序数据的特征选择方法

甘雨晴

长安大学理学院，陕西 西安

收稿日期：2024年4月28日；录用日期：2024年5月21日；发布日期：2024年5月30日

摘　要

随着大数据时代的不断发展，时序数据广泛存在于生活的各个领域。但现有的信息系统无法存放时序数据或者分类准确率较低。因此，本文构建时序模糊信息表，引入模糊相似关系，提出可以存放时序数据的时序模糊决策粗糙集模型，并研究其性质，给出基于时序依赖度的特征选择方法。

关键词

时序数据，时序模糊决策粗糙集，特征选择

A Feature Selection Method for Time Series Data Based on Dependence

Yuqing Gan

School of Science, Chang'an University, Xi'an Shaanxi

Received: Apr. 28th, 2024; accepted: May 21st, 2024; published: May 30th, 2024

Abstract

With the continuous development of the era of big data, time series data exists in all fields of life. However, the existing information system cannot store time series data or has a low classification accuracy. Therefore, this paper constructs a time series fuzzy information table, introduces fuzzy similarity relations, proposes a rough set model of time series fuzzy decisions that can store time series data, studies its properties, and gives a feature selection method based on time series dependence.

Keywords

Time Series Data, Time Series Fuzzy Decision Rough Set, Feature Selection

文章引用: 甘雨晴. 基于依赖度的时序数据的特征选择方法[J]. 应用数学进展, 2024, 13(5): 2172-2179.
DOI: 10.12677/aam.2024.135206

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

粗糙集理论是 Pawlak 在 1982 年提出的用于处理不精确、不完备数据的方法[1]。粗糙集理论建立在分类的基础上，用严格的等价关系构造上近似和下近似对未知的知识进行划分[2]。为了提高粗糙集处理数据的能力，Dubois 等[3]针对实际样本中的模糊性和不可分辨性可能造成信息丢失的问题，将模糊集和粗糙集相结合，提出模糊粗糙集的概念，至此一系列模糊粗糙集的拓展模型被提出。如，张夏伟等建立了悲观的多覆盖模糊粗糙集模型，分别讨论了悲观多覆盖模糊粗糙集、基于交的覆盖模糊粗糙集、覆盖模糊粗糙集和乐观多覆盖模糊粗糙集之间的关系[4]。李凡等针对模糊粗糙集容易受到噪音数据的影响，提出了变精度模糊粗糙集的概念[5]。邵迎超等将软集理论与模糊粗糙集结合起来，提出了软模糊粗糙集和软模糊粗糙群及它们的同态的概念，讨论了它们相关的性质[6]。李聪等利用模糊粗糙集和多粒度粗糙集各自优点的结合，提出了两类多粒度模糊粗糙集模型[7]。一系列基于模糊粗糙集的特征选择方法也被相继提出：张慧哲等把模糊集合相似度引入模糊粗糙集模型中，提出一种基于变相似度的模糊粗糙集模型，通过定义模糊相似矩阵和不一致程度矩阵，给出属性约简的算法[8]。陈毅宁等引入了基于距离比值尺度的样本集，通过对距离比值尺度的控制，避免了样本分布不确定性对近似集的影响，给出了该模型的基本性质，定义了新的依赖度函数，进而设计了属性约简算法[9]。陆娟等将粗糙集与二型模糊集结合，得到二型模糊粗糙集，并将模糊粗糙集属性约简的模型推广到二型模糊粗糙集框架中，得到了一个二型模糊粗糙属性约简的模型，并举例说明了用此模型进行属性约简的方法[10]。

然而，在人们所保存的数据中还有许多是具有时间特征的数据－时间序列数据。时间序列数据就是按照时间先后顺序记录各个观测样本的数据集[11]。其在现实生活中大量存在，如：金融证券市场中每天的股票价格变化，疫情感染日新增病例的数目，气象中某地区的每天气温与气压的读数以及医学中病人在每个时刻的心跳变化指数等。研究如何从数据量巨大、维度高、变量内部关系复杂的高维时间序列数据中挖掘出与时间、空间有关的隐藏信息，对于揭示对象发展变化的内部规律、不同的对象之间的相互作用关系以及为人们正确认识事物和科学决策提供依据等具有重要的理论价值和实际意义。

目前模糊粗糙集领域关于时间序列数据类型的研究较少，现有的信息系统无法用于存放多元时间序列数据。因此，研究可存放多维时间序列数据的信息系统并对其不确定性知识进行数据挖掘是很有价值的。

2. 预备知识

2.1. 时序数据的距离度量

通常计算两个时间序列之间的距离来度量两个时间序列之间的相似度，距离越小，相似性越高。本节将对常用的时间序列距离进行详细的介绍。

定义 1 设有两条长度相等的时间序列 $P=\{p_1,p_2,\cdots,p_n\}$ 和 $Q=\{q_1,q_2,\cdots,q_n\}$，它们的闵可夫斯基距离(Minkowski Distance)定义为：

$$Dist_m(P,Q)=\left[\sum_{i=1}^{n}(p_i-q_i)^r\right]^{\frac{1}{r}},$$

当 r 取值不同时，距离有不同的含义：

当 $r=1$ 时，为曼哈顿距离；

当 $r=2$ 时，为欧式距离；

当 $r\to\infty$ 时，为切比雪夫距离。

闵可夫斯基距离对于度量的时间序列有严格要求，即序列必须是等长且皆为数值型。因为闵可夫斯基距离所寻找的对应关系是一条时间序列上的点与另一条时间序列上的点一一对应。

如果两条时间序列整体是比较相似的，但时间轴是不对齐匹配的话，那么采用闵可夫斯基距离，将很难有效度量。而动态时间弯曲距离(Dynamic Time Warping, DTW)是目前使用最广泛的时间序列距离度量，与传统的闵可夫斯基距离相比，DTW 寻找的是两个时间序列之间的灵活的对应关系，也就是说一条时间序列上的一个点可以对应到另一个时间序列上的多个点，因此克服了时间不同步的问题。

定义 2 设有两个时间序列 $P=\{p_1,p_2,\cdots,p_n\}$ 和 $Q=\{q_1,q_2,\cdots,q_n\}$，构造一个 $n\times m$ 的距离矩阵 D：

$$D=\begin{bmatrix} d(1,1) & d(1,2) & \cdots & d(1,m) \\ d(2,1) & d(2,2) & \cdots & d(2,m) \\ \vdots & \vdots & \ddots & \vdots \\ d(n,1) & d(n,2) & \cdots & d(n,m) \end{bmatrix},$$

其中，$d(i,j)=\left|p_i-q_j\right|$，

这里采用欧氏距离 $d(i,j)$ 作为向量点 p_i 和 q_j 间的距离函数，$i=1,2,\cdots,n$，$j=1,2,\cdots,m$。为了计算时间序列 P 和 Q 的 DTW 距离 $\mathrm{DTW}(P,Q)$，需找到一条最佳弯曲路径 W，其中弯曲路径 W 中的第 k 个元素可定义为 $W_k(i,j)_k$，由此可得：

$$W=w_1,w_2,\cdots,w_k,\cdots,w_K$$

其中，$w_k=(p_k,q_k)$ 表示两个时间序列 P 和 Q 的匹配关系。弯曲路径长度满足 $\max(n,m)\le K\le n+m-1$。

弯曲路径 W 必须满足以下 3 个约束条件：

1) 边界性：$w_1=(p_1,q_1)$，$w_k=(p_n,q_m)$；

2) 连续性：$W_k(i,j)$ 和 $W_{k-1}(i',j')$ 满足 $i-i'\le 1$，$j-j'\le 1$；

3) 单调性：$i-i'\ge 0$，$j-j'\ge 0$。

可能存在多个 W 满足上述三个条件，DTW 通过动态规划寻找其中累积距离最短的路径：

$$\mathrm{DTW}(P,Q)=\min\sum_{k=1}^{K}D(w_k)=\min\sum_{k=1}^{K}d(p_k,q_k)$$

计算两个时间序列之间的 DTW 距离的一种常用方法是建立累积距离矩阵 φ。为了计算矩阵，我们使用具有以下递推的动态规划：

$$\varphi(i,j)=d(p_i,q_j)+\min\{\varphi(i-1,j-1),\varphi(i-1,j),\varphi(i,j-1)\}$$

$$\mathrm{DTW}(P,Q)=\varphi(n,m)$$

其中，$i=1,2,\cdots,n$，$j=1,2,\cdots,m$，$\varphi(0,0)=0$，$\varphi(i,0)=\varphi(0,j)=\infty$。

2.2. 模糊粗糙集

定义 3 称 $FDIS=(U,C\cup D,V_D,f,g)$ 是模糊决策信息系统，其中 U 为非空有限对象集，称为论域，C 是条件属性集，D 是决策属性，$f:U\times C\to[0,1]$，$f(x,a)$ 表示对象 x 在条件属性 a 下的取值。$g:U\times D\to V_D$，V_D 是决策属性的值域。

定义 4 设 R_B 是模糊决策信息系统 $FDIS=(U,C\cup D,V_D,f,g)$ 上的模糊二元关系，若 $\forall B\subseteq C$ 满足：

1) 自反性：$\forall x\in U$，$R(x,x)=1$；

2) 对称性：$\forall x,y\in U$，$R(x,y)=R(y,x)$；

3) 传递性：$\forall x,y,z\in U$，$R(x,y)\wedge R(y,z)\le R(x,z)$。

称模糊关系 R_B 为 U 上的模糊等价关系。若 R_B 仅满足(1)和(2)，则称模糊关系 R_B 为 U 上的模糊相似关系，也称为模糊相容关系。

若 R_B 是 U 上的模糊相似关系，$\forall x,y\in U$，$B\subseteq C$，令

$$[x]_{R_B}(y)=R_B(x,y),$$

称 $[x]_{R_B}$ 为 x 关于 R_B 的模糊邻域，则 $[x]_{R_B}$ 是 U 上一个模糊集。

定义 5 设 $FDIS=(U,C\cup D,V_D,f,g)$ 是模糊决策信息系统，R_B 是 U 上的模糊相似关系，$\forall B\subseteq C$，$U/D=\{D_1,\cdots,D_r\}$，决策属性 D 关于模糊相似关系 R_B 的下、上近似分别定义为：

$$\underline{R_B}(D_i)(y)=\inf_{y\in U}\max\{1-R_B(x,y),D_i(y)\},$$

$$\overline{R_B}(D_i)(y)=\sup_{y\in U}\min\{R_B(x,y),D_i(y)\},$$

正域定义为：

$$POS_B(D)=\bigcup_{i=1}^{r}\underline{R_B}(D_i),$$

依赖度定义为：

$$\gamma_B(D)=\frac{\sum_{x_i\in U}POS_B^{\delta}(D)(x_i)}{|U|}.$$

3. 时序模糊粗糙集

定义 6 称 $TFDIS=(U,C\cup D,T,V_D,f,g)$ 是时序模糊决策信息系统，其中 $U=\{x_1,x_2,\cdots,x_n\}$ 为非空有限对象构成的论域，$C=\{a_1,a_2,\cdots,a_m\}$ 是条件特征集(属性集)，$D=\{d\}$ 是决策特征，$T=\{t_1,t_2,\cdots,t_k\}$ 是有序时间集，且满足 $0\le t_1<t_2<\cdots<t_k$，$f:U\times C\times T\to[0,1]$，$f(x_i,a_j,t_k)$ 表示 t_k 时刻对象 x_i 关于特征 a_j 的取值，$g:U\times D\to V_D$，V_D 是属性 D 的值域。

表 1 给出了时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，其中 $U=\{x_1,x_2,x_3,x_4,x_5\}$，$C=\{a_1,a_2,a_3,a_4\}$，$T=\{t_1,t_2,t_3\}$，$D=\{d\}$。

Table 1. Time series fuzzy decision information system
表 1. 时序模糊决策信息系统

	a_1	a_2	a_3	d
	t_1,t_2,t_3	t_1,t_2,t_3	t_1,t_2,t_3	
x_1	0,0.1,0.4	0.9,0.1,0.4	0.2,0.9,0,	1
x_2	0.2,0.4,0.1	0.1,0.6,0.3	0.8,0.2,0.1	2
x_3	0.5,0.9,0.1	0.4,0.5,0.1	0.3,0.6,0.7	2
x_4	0.8,0.8,0.8	0.7,0.9,0.3	0.2,0.1,0.8	3
x_5	0.1,0.3,0.6	0.1,0.7,0.8	0.1,0.4,0.2	1

定义 7 [12]给定两个 D 维时序数据

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{D1} & a_{D2} & \cdots & a_{Dm} \end{bmatrix},\quad \mathbf{B}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{D1} & b_{D2} & \cdots & b_{Dn} \end{bmatrix},$$

其中，$i=1,2,\cdots,m$，$j=1,2,\cdots,n$，$d=1,2,\cdots,D$。$\mathbf{A}(:,i)=(a_{1i},a_{2i},\cdots,a_{Di})^T$ 表示 $\mathbf{A}$ 在时刻 i 下各个变量的取值向量；$\mathbf{A}(d,:)=(a_{d1},a_{d2},\cdots,a_{dm})^T$ 表示 $\mathbf{A}$ 在维度 d 下随时间变化的取值向量；$\mathbf{B}(:,j)=(b_{1j},b_{2j},\cdots,b_{dj})^T$ 表示 $\mathbf{B}$ 在时刻 j 下各个变量的取值向量；$\mathbf{B}(d,:)=(b_{d1},b_{d2},\cdots,b_{dn})^T$ 表示 $\mathbf{B}$ 在维度 d 下随时间变化的取值向量。

设多维时间序列 $\mathbf{A}$ 和 $\mathbf{B}$，则基于广义马氏距离的子距离可定义为：

$$d_M\left(A(i),B(j)\right)=\sqrt{\left(A(:,i)-B(:,j)\right)^T\mathbf{M}\left(A(:,i)-B(:,j)\right)},$$

其中，$i=1,2,\cdots,m$；$j=1,2,\cdots,n$，$\mathbf{M}$ 是一个对称的半正定矩阵，称为马氏矩阵。当 $\mathbf{M}=\mathbf{I}$，广义马氏距离就转变为欧氏距离。

多维时间序列 $\mathbf{A}$ 和 $\mathbf{B}$ 的 $\mathrm{DTW}_{\mathrm{M}}(\mathbf{A},\mathbf{B})$ 公式可以表示为：

$$\mathrm{DTW}_{\mathrm{M}}(\mathbf{A},\mathbf{B})=CD(m,n)$$

$$CD(i,j)=d_M\left(\mathbf{A}(i),\mathbf{B}(j)\right)+\min\begin{cases} CD(i-1,j) \\ CD(i,j-1) \\ CD(i-1,j-1) \end{cases}$$

其中，$CD(0,0)=0$，$CD(i,0)=CD(0,j)=\infty$。

定义 8 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，对任意 $x,y\in U$，$\mathrm{x_L}=\left(f(x,a_1,t_l),f(x,a_2,t_l),\cdots,f(x,a_m,t_l)\right)^T$ 是对象 x 在时刻 l 下的属性值向量，$\mathrm{y_S}=\left(f(y,a_1,t_s),f(y,a_2,t_s),\cdots,f(y,a_m,t_s)\right)^T$ 是对象 y 在时刻 s 下的属性值向量。则 $\mathrm{x_L}$ 和 $\mathrm{y_S}$ 的马氏距离定义为：

$$d_M\left(\mathrm{x_L},\mathrm{y_S}\right)=\sqrt{\left(\mathrm{x_L}-\mathrm{y_S}\right)^T\mathbf{M}\left(\mathrm{x_L}-\mathrm{y_S}\right)}$$

其中，$\mathbf{M}$ 是协方差矩阵。

定义 9 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，对任意 $B\subseteq C$，$x_i,x_j\in U$，x_i 和 x_j 在有序时间集 T 上关于特征子集 B 的时序模糊相似关系为：

$$R_B^T\left(x_i,x_j\right)=\exp\left(-\frac{DTW_M\left(x_i,x_j\right)}{2\sigma^2}\right)$$

对任意 $\delta\in(0,1)$，$x\in U$ 在有序时间集 T 上关于特征集 B 的时序模糊邻域粒 $\delta_B^T(x)$ 定义为：$\forall y\in U$，

$$\delta_B^T(x)(y)=\begin{cases} R_B^T(x,y), & R_B^T(x,y)\geq\delta, \\ 0, & R_B^T(x,y)<\delta. \end{cases}$$

决策属性划分为 $U/D=\{D_1,D_2,\cdots,D_r\}$ δ_C^T 是由特征集 C 在 U 上诱导的时序模糊邻域粒，那么 $\forall x\in U$ 的时序模糊决策定义为：

$$\tilde{D}_k^T(x)=\frac{\left|\delta_C^T(x)\cap D_k\right|}{\left|\delta_C^T(x)\right|},\quad k=1,2,\cdots,r$$

$\tilde{D}^T=\left\{\tilde{D}_2^T,\tilde{D}_2^T,\cdots,\tilde{D}_r^T\right\}$ 是一个模糊集，$\tilde{D}_k^T(x)$ 为 x 关于 $\tilde{D}_k^T$ 的模糊隶属度。

定义 10 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，对任意 $B\subseteq C$，决策特征划分为 $U/D=\{D_1,D_2,\cdots,D_r\}$，其对应的模糊决策为 $\tilde{D}^T=\left\{\tilde{D}_1^T,\tilde{D}_2^T,\cdots,\tilde{D}_r^T\right\}$，$R_B^T$ 是由 B 在 U 上诱导的时序模糊相似关系，则模糊决策 $\tilde{D}^T$ 关于特征集 B 的上、下近似定义为：

$$\overline{R}_B^{\delta}\left(\tilde{D}^T\right)=\left\{\overline{R}_B^{\delta}\left(\tilde{D}_1^T\right),\overline{R}_B^{\delta}\left(\tilde{D}_2^T\right),\cdots,\overline{R}_B^{\delta}\left(\tilde{D}_r^T\right)\right\},$$

$$\underline{R}_B^{\delta}\left(\tilde{D}^T\right)=\left\{\underline{R}_B^{\delta}\left(\tilde{D}_1^T\right),\underline{R}_B^{\delta}\left(\tilde{D}_2^T\right),\cdots,\underline{R}_B^{\delta}\left(\tilde{D}_r^T\right)\right\},$$

其中 $\tilde{D}_k^T$ 关于特征集 B 的上、下近似定义为：

$$\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=\sup_{x_j\in U}T\left\{R_B^T\left(x_i,x_j\right),\tilde{D}_k^T\left(x_j\right)\right\},$$

$$\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=\inf_{x_j\in U}S\left\{N\left(R_B^T\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}.$$

定理 1 对任意 $\tilde{D}_k^T,\tilde{D}_g^T\in\tilde{D}^T$，$x_i,x_j\in U$，$B,B_1,B_2\subseteq C$，下列结论成立：

1) $\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\tilde{D}^T\subseteq\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)$；

2) 若 $B_1\subseteq B_2$，有 $\underline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)\subseteq\underline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)$，$\overline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)\subseteq\overline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)$；

3) 若 $\tilde{D}_k^T\subseteq\tilde{D}_g^T$，有 $\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\underline{R}_B^{\delta}\left(\tilde{D}_g^T\right)$，$\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\overline{WR}_B^{\delta}\left(\tilde{D}_g^T\right)$；

4) $N\left(\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\right)=\overline{R}_B^{\delta}\left(N\left(\tilde{D}_k^T\right)\right)$，$N\left(\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\right)=\underline{R}_B^{\delta}\left(N\left(\tilde{D}_k^T\right)\right)$。

证明：1) 对任意 $x_i,x_j\in U$，由定义 9 知 $R_B^T\left(x_i,x_i\right)=1$，从而 $\max\left\{\left\{1-R_B^{\delta}\left(x_i,x_i\right)\right\},\tilde{D}_k^T\left(x_i\right)\right\}=\tilde{D}_k^T\left(x_i\right)$，显然 $\inf_{x_j\in U}S\left\{N\left(R_B^T\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}\le\tilde{D}_k^T\left(x_i\right)$，所以 $\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\tilde{D}^T$。同理可得 $\tilde{D}^T\subseteq\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)$，即 $\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\tilde{D}^T\subseteq\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)$。

2) 由定义 9 和定义 10 可知 $\forall x_i,x_j\in U$，$B_1\subseteq B_2\subseteq C$，$\underline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)(x_i)=\inf_{x_j\in U}S\left\{N\left(R_{B_1}^T\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}$

$=\inf_{x_j\in U}S\left\{\left\{1-R_{B_1}^T\left(x_i,x_j\right)\right\},\tilde{D}_k^T\left(x_j\right)\right\}\le\inf_{x_j\in U}S\left\{\left\{1-R_{B_2}^T\left(x_i,x_j\right)\right\},\tilde{D}_k^T\left(x_j\right)\right\}=\inf_{x_j\in U}S\left\{N\left(R_{B_2}^T\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}$

$=\underline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)(x_i)$。

由 x_i 的任意性可得 $\underline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)\subseteq\underline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)$。

同理可类似求得：$\overline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)(x_i)\le\overline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)(x_i)$。由 x_i 的任意性可得 $\overline{R}_{B_2}^{\delta}\left(\tilde{D}_k^T\right)\subseteq\overline{R}_{B_1}^{\delta}\left(\tilde{D}_k^T\right)$。

3) 由定义 9 和定义 10 可知 $\forall x_i,x_j\in U$，$\tilde{D}_k^T\subseteq\tilde{D}_g^T$，$\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=\inf_{x_j\in U}S\left\{N\left(R_B^T\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}$

$\le\inf_{x_j\in U}S\left\{N\left(R_B^T\left(x_i,x_j\right)\right),\tilde{D}_g^T\left(x_j\right)\right\}=\underline{R}_B^{\delta}\left(\tilde{D}_g^T\right)(x_i)$。由 x_i 的任意性可得 $\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\underline{R}_B^{\delta}\left(\tilde{D}_g^T\right)$。

同理类似求得：$\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=\sup_{x_j\in U}T\left\{R_B^T\left(x_i,x_j\right),\tilde{D}_k^T\left(x_j\right)\right\}\le\sup_{x_j\in U}T\left\{R_B^T\left(x_i,x_j\right),\tilde{D}_g^T\left(x_j\right)\right\}=\overline{R}_B^{\delta}\left(\tilde{D}_g^T\right)(x_i)$。

由 x_i 的任意性可得 $\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\subseteq\overline{R}_B^{\delta}\left(\tilde{D}_g^T\right)$。

4) 由定义 10 可知 $\forall x_i,x_j\in U$，$\tilde{D}_k^T\subseteq\tilde{D}^T$，$N\left(\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)\right)=1-\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=1-\inf_{x_j\in U}S\left\{N\left(R_B^{\delta}\left(x_i,x_j\right)\right),\tilde{D}_k^T\left(x_j\right)\right\}=\sup_{x_j\in U}T\left\{R_B^{\delta}\left(x_i,x_j\right),1-\tilde{D}_k^T\left(x_j\right)\right\}=\overline{R}_B^{\delta}\left(N\left(\tilde{D}_k^T\right)(x_i)\right)$。由 x_i 的任意性可得 $N\left(\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\right)=\overline{R}_B^{\delta}\left(N\left(\tilde{D}_k^T\right)\right)$。同理类似可求 $N\left(\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)\right)=1-\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)(x_i)=1-\sup_{x_j\in U}T\left\{R_B^T\left(x_i,x_j\right),\tilde{D}_k^T\left(x_j\right)\right\}=\inf_{x_j\in U}S\left\{\left(1-R_B^T\left(x_i,x_j\right)\right),1-\tilde{D}_k^T\left(x_j\right)\right\}=\underline{R}_B^{T,\delta}\left(N\left(\tilde{D}_k^T\right)(x_i)\right)$。由 x_i 的任意性可得 $N\left(\overline{R}_B^{\delta}\left(\tilde{D}_k^T\right)\right)=\underline{R}_B^{\delta}\left(N\left(\tilde{D}_k^T\right)\right)$。

4. 基于时序依赖度的时序数据的特征选择

本节将定义基于时序模糊邻域粗糙集模型的特征依赖度，并基于此提出一种时序数据的特征选择算法，见表 2。

定义 11 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，对任意 $B\subseteq C$，决策特征划分为 $U/D=\{D_1,D_2,\cdots,D_r\}$，其对应的模糊决策为 $\tilde{D}^T=\{\tilde{D}_1^T,\tilde{D}_2^T,\cdots,\tilde{D}_r^T\}$，$R_B^T$ 是由 B 在 U 上诱导的时序模糊相似关系，$\tilde{D}^T$ 关于特征子集 B 的正域定义为：

$$POS_B^{\delta}\left(\tilde{D}^T\right)=\bigcup_{k=1}^{r}\underline{R}_B^{\delta}\left(\tilde{D}_k^T\right),$$

正域的大小反应了特征子集 B 的分类能力。根据正域的定义，$\tilde{D}^T$ 关于特征子集 B 的时序依赖度为：

$$\gamma_B^{\delta}\left(\tilde{D}^T\right)=\frac{\sum_{x\in U}POS_B^{\delta}\left(\tilde{D}^T\right)(x)}{|U|}.$$

性质 1 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，对任意 $B_1\subseteq B_2\subseteq C$，有以下性质成立：

1) $POS_{B_1}^{\delta}\left(\tilde{D}^T\right)\subseteq POS_{B_2}^{\delta}\left(\tilde{D}^T\right)$；
2) $\gamma_{B_1}^{\delta}\left(\tilde{D}^T\right)\subseteq\gamma_{B_2}^{\delta}\left(\tilde{D}^T\right)$。

定义 12 给定一个时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，决策特征划分为 $U/D=\{D_1,D_2,\cdots,D_r\}$，其对应的模糊决策为 $\tilde{D}^T=\{\tilde{D}_1^T,\tilde{D}_2^T,\cdots,\tilde{D}_r^T\}$，对任意 $a\in B$，$B\subseteq C$，$\delta\in(0,1)$。若满足 $\gamma_{B-\{a\}}^{\delta}\left(\tilde{D}^T\right)=\gamma_B^{\delta}\left(\tilde{D}^T\right)$，称 a 在 B 中是不必要的(或冗余的)；否则，称 a 在 B 中是必要的。若满足下列条件，则称特征子集 B 是 C 的一个特征约简集：

1) $\gamma_B^{\delta}\left(\tilde{D}^T\right)=\gamma_C^{\delta}\left(\tilde{D}^T\right)$；
2) $\gamma_{B-\{a\}}^{\delta}\left(\tilde{D}^T\right)<\gamma_B^{\delta}\left(\tilde{D}^T\right)$。

Table 2. Feature selection algorithm for time series data based on time series dependence
表 2. 基于时序依赖度的时序数据的特征选择算法

算法 1 基于时序依赖度的时序数据的特征选择算法
输入：时序模糊决策信息系统 $TFDIS=(U,C\cup D,T,V_D,f,g)$，邻域半径 δ。
输出：特征约简 red。
Step1：初始化：$red=\varnothing$；
Step2：$\forall a\in C$，计算特征 a 的时序模糊相似关系 R_a^T 和 $C-\{a\}$ 的时序模糊相似关系 $R_{C-\{a\}}^T$，时序模糊邻域粒 $\delta_C^T(x)$ 和 $\delta_{C-\{a\}}^T(x)$；
Step3：$\forall a\in C$，根据公式(4.6)与公式(4.8)计算计算时序模依赖度 $\gamma_{B\cup\{a\}}^{\delta}\left(\tilde{D}^T\right)$ 和时序外重要度 $sig_{out}^T(a,B,D)=\gamma_{B\cup\{a\}}^{\delta}\left(\tilde{D}^T\right)-\gamma_B^{\delta}\left(\tilde{D}^T\right)$；
Step4：选取满足下面条件的属性 a：$\max sig_{out}^T(a,B,D)=\gamma_{B\cup\{a\}}^{\delta}\left(\tilde{D}^T\right)-\gamma_B^{\delta}\left(\tilde{D}^T\right)$；
Step5：若 $sig_{out}^T(a,B,D)=0$，则停止循环；否则，$red=red\cup\{a\}$；返回 Step3；
Step6：对任意 $a\in red$，计算时序内重要度 $sig_{in}^T\left(a,B,\tilde{D}^T\right)=\gamma_B^{\delta}\left(\tilde{D}^T\right)-\gamma_{B-\{a\}}^{\delta}\left(\tilde{D}^T\right)$；
Step7：若 $sig_{int}^T\left(a,B,\tilde{D}^T\right)=0$，则 $red=red-\{a\}$；否则停止循环；
Stpe8：返回特征约简 red。

定义 13 设 $TFDIS=(U,C\cup D,T,V_D,f,g)$ 为序模糊决策信息系统。对任意 $a\in B$，$B\subseteq C$，$\delta\in(0,1)$，特征 a 关于 B 的时序内重要度定义：

$$sig_{in}^{T}\left(a,B,\tilde{D}^{T}\right)=\gamma_{B}^{\delta}\left(\tilde{D}^{T}\right)-\gamma_{B-\{a\}}^{\delta}\left(\tilde{D}^{T}\right).$$

对任意 $a\in C-B$，特征 a 关于 B 的时序外重要度定义为：

$$sig_{out}^{T}\left(a,B,D\right)=\gamma_{B\cup\{a\}}^{\delta}\left(\tilde{D}^{T}\right)-\gamma_{B}^{\delta}\left(\tilde{D}^{T}\right).$$

由定义 13,可得：$0\le sig_{int}^{T}\left(a,B,\tilde{D}^{T}\right)\le 1$。若 $sig_{int}^{T}\left(a,B,\tilde{D}^{T}\right)>0$，则 a 为核心属性；若 $sig_{int}^{T}\left(a,B,\tilde{D}^{T}\right)=0$，则 a 为不必要属性，即 a 可以从属性集 B 中去除。同理，有 $0\le sig_{out}^{T}\left(a,B,\tilde{D}^{T}\right)\le 1$。若 $sig_{out}^{T}\left(a,B,\tilde{D}^{T}\right)=0$，则 a 为不必要属性；若 $sig_{out}^{T}\left(a,B,\tilde{D}^{T}\right)>0$，则 a 是相对必要属性，可通过筛选 $sig_{out}^{T}\left(a,B,\tilde{D}^{T}\right)$ 的最大值作为候选特征约简集。

5. 结论

本章主要研究时序模糊决策信息系统上的时序数据的特征选择方法。首先定义了时序模糊决策信息系统，引入时序马氏距离，提出时序模糊相似关系，进而提出了时序模糊决策粗糙集，并讨论了其的性质。定义了在时序模糊决策粗糙集上的下近似、上近似、正域和依赖度。通过定义特征内外重要度，提出了基于时序依赖度的特征选择方法。

参考文献

[1] Pawlak, Z. (1982) Rough Sets. *International Journal of Computer and Information Sciences*, **11**, 341-356. https://doi.org/10.1007/BF01001956

[2] 张文修, 梁怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003.

[3] Dubois, D. and Prade, H. (1990) Rough Fuzzy Sets and Fuzzy Rough Sets. *International Journal of General Systems*, **17**, 191-209. https://doi.org/10.1080/03081079008935107

[4] Zhang, X.W. (2016) Pessimistic Multi-Covering Fuzzy Rough Sets. *Journal of Xiamen University* (*Natural Science*), **55**, 918-921.

[5] Li, F., Liu, Q.H. and Yang, G.W. (2008) Definition of Variable Precision Fuzzy Rough Sets. *Control and Decision*, **23**, 1206-1210.

[6] Shao, Y.C., Du, Z.B., Yang, J.L., *et al*. (2013) Soft Fuzzy Rough Sets. *Computer Engineering and Applications*, **49**, 125-128.

[7] Li, C. (2016) The Study on Multi-Granulatlon Fuzzy Rough Set. *Journal of Mathematics*, **36**, 124-134.

[8] Zhang, H.Z., Wang, J. and Hong, B. (2009) Attribute Reduction of Fuzzy Rough Sets Based on Variable Similar Degree. *Pattern Recognition and Artificial Intelligence*, **22**, 393-399.

[9] Chen, Y.N. and Chen, H.M. (2020) Attribute Reduction of Fuzzy Rough Set Based on Distance Ratio Scale. *Computer Science*, **47**, 67-72.

[10] Lu, J. and Li, D.Y. (2017) Model for Type-2 Fuzzy Rough Attribute Reduction. *Computer Science*, **44**, 28-33.

[11] James, D. (2020) Time Series Analysis. Princeton University Press, Princeton.

[12] Yang, X., Ma, J.M. and Zhao, M.J. (2023) Feature Selection of High-Dimensional Time-Series Data Based on Neighborhood Mutual Information. *Computer Engineering*, **49**, 135-142, 149.