ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 34,No. 3,May 2019,pp. 538－547
DOI:10. 16337/j. 1004-9037. 2019. 03. 017

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

一种基于模糊粗糙集的快速特征选择算法

张　晓[1]　杨燕燕[2]

(1. 西安理工大学应用数学系,西安,710048;2. 清华大学自动化系,北京,100084)

摘　要: 模糊粗糙集由于能够处理实数值数据,甚至是混合值数据中的不确定性受到人们的广泛关注,其最重要的应用之一是特征选择,相关的特征选择方法已有不少研究,但其快速的特征选择算法研究很少。实际中的数据一般含有噪声点或信息含量低的样例,如果对数据集先筛选出代表样例,再对筛选的样例集进行数据挖掘便会降低挖掘计算量。本文基于模糊粗糙集,先根据样例的模糊下近似值对样例进行筛选,然后利用筛选样例的模糊粗糙信息熵构造特征选择的评估度量,并给出相应的特征选择算法,从而降低了算法的计算复杂度。数值试验表明该快速算法具有有效性,并且对控制筛选样例个数的参数给出了建议。

关键词: 模糊粗糙集;样例选择;特征选择;信息熵

中图分类号: TP18　　文献标志码:A

Fast Feature Selection Algorithm Based on Fuzzy Rough Sets

Zhang Xiao[1], Yang Yanyan[2]

(1. Department of Applied Mathematics, Xi′an University of Technology, Xi′an, 710048, China; 2. Department of Automation, Tsinghua University, Beijing, 100084, China)

Abstract: Fuzzy rough set theory has been paid much attention since it can be used to deal with the uncertainty in the real-valued data or even the mixed data. One of the most important applications of fuzzy rough sets is feature selection, and there have existed many related feature selection methods. However, little attention has been paid on fast feature selection algorithms. Data collected in practice generally include noises or possess some instances with less information. Considering to previously select representative instances from the original data set and perform data mining algorithms on the selected instances set, one may reduce the computation of the algorithms. In view of the advantage of instance selection, the instances are firstly selected based on fuzzy rough sets according to the values of the fuzzy lower approximation of instances in this paper. Then, the evaluation measure of feature selection is constructed by using fuzzy rough set-based information entropy of the selected instances, and the corresponding feature selection algorithm is provided to alleviate the computational complexity. Some numerical experiments are conducted to show the efficiency of the proposed fast algorithm, and the reasonable suggestion of the critical parameter is given to determine the number of the selected instances.

Key words: fuzzy rough sets; instance selection; feature selection; information entropy

---

基金项目: 国家自然科学基金(61602372,61806108)资助项目;西安理工大学博士研究启动基金(109-256081504)资助项目;中国博士后基金(2018M631475)资助项目

收稿日期: 2018-04-23;**修订日期:** 2019-02-19

引　言

经典的粗糙集理论[1]是由波兰数学家Pawlak在1982年提出的,它是一种处理数据中的不确定性的有效工具,然而经典粗糙集只能处理符号值(名义值)的数据。模糊粗糙集[2]作为经典粗糙集的最重要的推广之一,可以用来处理实数值甚至是混合值的数据。目前,模糊粗糙集已经成功应用于机器学习和数据挖掘领域[3],其最受人们关注的应用之一就是特征选择(属性约简)。关于模糊粗糙集特征选择的研究工作已存在不少[4-10],但其快速的特征选择算法的研究还很少,据作者所知,仅文献[11]在特征选择算法迭代步骤提供了加速策略,从而减少了算法的计算时间。

实际中的数据一般包含信息量较低的样例或噪声点,如果对样例进行筛选,利用筛选得到的样例进行挖掘知识将会减少计算的复杂度。文献[12]提供了3种样例选择的启发式算法,其中之一的算法思想即选择隶属模糊正域的值不小于给定阈值的那些样例。文献[13]使用一种模糊粗糙度量来刻画样例的质量并给出了wrapper式的样例选择方法。文献[14]针对k-最近邻分类规则提出了一种加权抽样技术来筛选代表样例。事实上,特征选择和样例选择是相对独立的工作,也有一些文献基于模糊粗糙集研究特征和样例同时选择的方法[15-18]。例如,文献[18]给出了一个基于频率的启发式算法来交替选取特征和样例,以达到特征和样例同时被选取的目的。

由于现有的基于模糊粗糙集的特征选择算法的复杂度一般是$O(n^2m^2)$,其中n为数据集中的样例个数,m为特征个数。当数据集中有较多的样例时,现有的特征选择算法会消耗大量的计算时间和存储空间。注意到特征选择和样例选择还有另外一个结合点,即是先对数据进行样例选择,然后利用筛选的代表样例进行特征选择,从而减少特征选择算法的计算时间。因此,本文是对数据集筛选代表样例进行特征选择达到加速计算的目的,不同于文献[11]在特征选择算法迭代步骤进行加速的策略,这也为特征选择的快速算法提供了一种新的思路。

基于文献[12]中样例选择的思想,本文先对样例进行筛选,即筛选那些模糊下近似值不低于给定阈值的那些样例,然后在文献[7]的基于模糊粗糙集构造信息熵进行特征选择的工作基础上,只利用筛选样例的信息熵进行特征选择以降低算法的复杂度,从而提供了一种快速的特征选择算法。数值试验表明该算法具有有效性,且对筛选样例多少的关键参数给出了合理的建议。

1　预备知识

1.1　模糊粗糙集

设U是一个论域,$F(U\times U)$为$U\times U$上的模糊幂集。如果$R\in F(U\times U)$,R称为一个在$U\times U$上的模糊关系。如果对任意的$x\in U$有$R(x,x)=1$,R称为是自反的;如果对任意的$x,y\in U$有$R(x,y)=R(y,x)$,R称为是对称的;如果对任意的$x,y,z\in U$有$R(x,y)\geqslant T(R(x,z),R(z,y))$,则$R$称为是$T$-传递的,其中$T$为三角范数。另外,如果$R$是自反、对称和$T$-传递的,则称$R$是$U$上的一个模糊$T$-相似关系。

文献[2]在模糊T-相似关系R上给出了模糊集$X\in F(U)$的一对下、上近似算子:对任意的$x\in U$

$$\underline{R}\,X(x)=\inf_{y\in U}\max\{1-R(x,y),X(y)\} \tag{1}$$

$$\overline{R}\,X(x)=\sup_{y\in U}\min\{R(x,y),X(y)\} \tag{2}$$

式中:$\underline{R}\,X(x)$是对象x隶属于模糊集X的确定程度,而$\overline{R}\,X(x)$是对象x隶属于模糊集X的可能程度,$(\underline{R}\,X,\overline{R}\,X)$称为是$X$的模糊粗糙集。

式(1)和式(2)是模糊粗糙集最初的一对下、上近似算子,后来也有不少文献对其进行了推广,而式(1)

和式(2)是应用最为广泛的一对近似算子,故本文的研究工作也是在其基础之上展开的。

1.2 模糊信息系统和模糊决策系统

一个模糊信息系统是一个二元组(U,A),其中$U=\{x_1,x_2,\cdots,x_n\}$为论域,x_i为对象(样例);$A=\{a_1,a_2,\cdots,a_m\}$是一个有限非空的属性(特征)集;对于每个$a\in A$,有一个映射$a:U\to V_a$,V_a称为属性a的值域,且每个属性a都可定义一个模糊关系$R_{\{a\}}$。由任意的属性子集$B\subseteq A$可定义一个模糊关系$R_B=\bigcap_{a\in B}R_{\{a\}}$。

一个模糊决策系统是一个二元组$(U,A\cup D)$,$A\cap D=\varnothing$,其中(U,A)是一个模糊信息系统,A称为条件属性集,$D=\{d\}$称为决策属性集,d是符号值的属性,成立一个映射$d:U\to V_d$,且$V_d=\{d(x):x\in U\}$称为决策属性d的值域。

在决策属性集D上定义一个等价关系,即

$$R_D=\{(x,y)\in U\times U:d(x)=d(y)\}$$

则R_D产生U的一族划分

$$U/R_D=\{[x_i]_D:x_i\in U\}$$

式中$[x_i]_D=\{x_j\in U:(x_i,x_j)\in R_D\}$称作是对象$x_i$所属的决策类。需要指出的是,分明决策类$[x_i]_D$的特征函数为

$$[x_i]_D(x_j)=\begin{cases}1 & x_j\in[x_i]_D\\ 0 & x_j\notin[x_i]_D\end{cases}$$

定义 1[5] 设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$F(U)$是U上的模糊幂集。模糊集$X\in F(U)$的基数定义为

$$|X|=\sum_{i=1}^{n}X(x_i)\tag{3}$$

1.3 λ-信息熵

定义 2[7] 设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$B\subseteq A$。决策属性集D相对于条件属性集B的λ-条件熵定义为

$$H_\lambda(D|B)=-\frac{1}{n}\sum_{i=1}^{n}\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|\log\frac{\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|}{\left|[x_i]_B^{\lambda_i}\right|}\tag{4}$$

式中

$$[x_i]_B^{\lambda_i}(x_j)=\begin{cases}0 & 1-R_B(x_i,x_j)\geqslant\lambda_i\\ \lambda_i & 1-R_B(x_i,x_j)<\lambda_i\end{cases}\tag{5}$$

是对象x_i关于B的模糊粒,$\lambda_i=\underline{R_A}\,[x_i]_D(x_i)$。

注释 1[7] 如果$\left|[x_i]_B^{\lambda_i}\right|=0$且$\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|=0$,在这种情况下,定义

$$\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|\log\frac{\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|}{\left|[x_i]_B^{\lambda_i}\right|}=0\tag{6}$$

定理 1[7] 设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$C\subseteq B\subseteq A$,则

$$H_\lambda(D|C)\geqslant H_\lambda(D|B)\tag{7}$$

由定理1知,λ-条件熵关于属性子集是单调的。

2　基于模糊决策系统筛选样例的特征选择算法

本节利用文献[12]的样例选择思想,先对模糊决策系统的样例进行筛选,然后利用筛选样例构造新的λ-信息熵,并给出相应的特征选择算法。

设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,属性子集$B\subseteq A$。对U中每一个对象(样例)$x_i(i=1,2,\cdots,n)$,计算对象x_i隶属其所在决策类$[x_i]_D$的下近似值$\lambda_i=\underline{R_A}\,[x_i]_D(x_i)$。已知$\underline{R_A}\,[x_i]_D(x_i)$度量了对象$x_i$隶属$[x_i]_D$的确定程度,因此$\underline{R_A}\,[x_i]_D(x_i)$的值越小,对象$x_i$隶属$[x_i]_D$的确定程度越低,这也说明了对象$x_i$包含的不确定信息越多。在实际中,边界点和噪声点包含更多的不确定信息,故隶属其所在的决策类的下近似值可能会很小。如果在数据挖掘过程中忽略掉这些有较小下近似值的样例,那么会减少计算时间。给定阈值$\alpha\in[0,1]$,记

$$U_\alpha=\{x_i:\underline{R_A}\,[x_i]_D(x_i)\geqslant\alpha,x_i\in U,i=1,2,\cdots,n\}\tag{8}$$

为由参数α确定的$(U,A\cup D)$的筛选样例集。

利用筛选样例集来完成数据挖掘的任务要使最终的结果不能和完整的数据集所得的结果相差太大,因此阈值α要合理地选取。α取值大小直接决定了筛选样例的多少。如果筛选样例过多,从而不能有效地减少计算时间;而筛选样例过少又会损失较多的信息,具体的α取值建议将在数值试验部分给出。

定义3　设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$B\subseteq A$,U_α是筛选样例集。决策属性集D相对于条件属性集B的U_α-λ-条件熵定义为

$$H_\lambda^{U_\alpha}(D|B)=-\frac{1}{|U_\alpha|}\sum_{x_i\in U_\alpha}\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|\log\frac{\left|[x_i]_B^{\lambda_i}\cap[x_i]_D\right|}{\left|[x_i]_B^{\lambda_i}\right|}\tag{9}$$

式中$|U_\alpha|$为U_α的基数。

由定义3易知U_α-λ-条件熵可以看作是λ-条件熵的一种推广,不同之处在于U_α-λ-条件熵只考虑筛选样例的信息熵,而λ-条件熵考虑所有样例的信息熵。因而由定理1知U_α-λ-条件熵也是单调的,即对$C\subseteq B\subseteq A$有$H_\lambda^{U_\alpha}(D|C)\geqslant H_\lambda^{U_\alpha}(D|B)$。应该指出的是注释1对定义3同样成立,且由文献[7]定理3易知$H_\lambda^{U_\alpha}(D|A)=0$恒成立。

定理2　设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$B\subseteq A$,U_α是筛选样例集,则$H_\lambda^{U_\alpha}(D|B)$的最大值为$|U_\alpha|/e$。

证明:由文献[7]定理5易证。

定理3　设$(U,A\cup D)$是一个模糊决策系统,$U=\{x_1,x_2,\cdots,x_n\}$,$B\subseteq A$,U_α是筛选样例集,则$H_\lambda^{U_\alpha}(D|B)=0$当且仅当$\underline{R_B}\,[x_i]_D(x_i)=\underline{R_A}\,[x_i]_D(x_i)=\lambda_i$对任意的$x_i\in U_\alpha$成立。

证明:由文献[7]定理6易证。

如果一个新的条件属性添加到条件属性子集,则U_α-λ-条件熵就会单调地减少,从而U_α-λ-条件熵减少的值就反映了添加的属性相对条件属性子集的重要程度。

定义4　设$(U,A\cup D)$是一个模糊决策系统,$B\subseteq A$,U_α是筛选样例集。对任意的条件属性$a\in A\backslash B$,a相对于D对B的U_α-重要性定义为

$$\mathrm{SIG}_\lambda^{U_\alpha}(a,B,D)=H_\lambda^{U_\alpha}(D|B)-H_\lambda^{U_\alpha}(D|B\cup\{a\})\tag{10}$$

利用U_α-重要性度量,给出相应的特征选择算法。

算法1 基于模糊决策系统筛选样例的特征选择算法

输入:模糊决策系统$(U,A\cup D)$,$U=\{x_1,x_2,\cdots,x_n\}$,阈值α;

输出:属性子集B

① 初始化$U_\alpha=\varnothing$。对每一个对象$x_i\in U$,根据式(1)计算$\lambda_i=\underline{R_A}[x_i]_D(x_i)$。如果$\lambda_i\geqslant\alpha$,添加$x_i$到$U_\alpha$;

② 初始化$B=\varnothing$,$H_\lambda^{U_\alpha}(D|B)=|U_\alpha|/\mathrm{e}$;

③ 对每个条件属性$a_j\in A\backslash B$,计算$\mathrm{SIG}_\lambda^{U_\alpha}(a_j,B,D)$;

④ 对满足$\mathrm{SIG}_\lambda^{U_\alpha}(a_{j_0},B,D)=\max\limits_j\mathrm{SIG}_\lambda^{U_\alpha}(a_j,B,D)$的属性$a_{j_0}$,如果$H_\lambda^{U_\alpha}(D|B\cup\{a_{j_0}\})\geqslant 0$且$H_\lambda^{U_\alpha}(D|B)>0$,则添加$a_{j_0}$到$B$中,并返回③;否则转⑤。

⑤ 输出B并终止算法。

该算法的时间复杂度是多项式级的。实际上,该算法第1步的时间复杂度为$O(|U|^2|A|)$,第3步的时间复杂度至多为$O(|U_\alpha||U||A|)$。另外,第3步至多迭代$|A|$次,第4步的时间复杂度为$O(|A|)$。综上,算法1的时间复杂度为$O(|U||A|(|U|+|U_\alpha||A|))$。

3 数值试验

本节通过一些数值试验对算法1的有效性进行评估。试验主要使用算法1搜索1个特征子集,评估参数α对特征选择在特征个数、计算时间及获取精度等方面的影响。为了达到目的,从UCI数据库下载了8个数据集,关于数据集的描述如表1所示。

表1 数据集描述

Tab. 1 Description of data sets

Data set	Abbreviation of data set	Number of instances	Number of conditional attributes			Number of classes
			All	Real-valued	Nominal	
Wine	Wine	178	13	13	0	3
Wisconsin diagnostic breast cancer	WDBC	569	30	30	0	2
Libras	Libras	360	90	90	0	15
Steel plates faults	Steel	1 941	27	27	0	7
Cardiotocography	CTG	2 126	20	20	0	3
Statlog(Heart)	Heart	270	13	6	7	2
Horse colic	Horse	368	22	7	15	2
Credit approval	Credit	690	15	6	9	2

3.1 数据预处理和试验设计

对每个数据集,分别用U,A和D标记论域、条件属性集和决策属性集。如果其中存在一些实数值的条件属性,则对这些属性的属性值先进行标准化,即对实数值的属性$a\in A$有

$$\tilde{a}(x_i)=\frac{a(x_i)-\min\limits_j a(x_j)}{\max\limits_j a(x_j)-\min\limits_j a(x_j)}\quad x_i\in U \tag{11}$$

于是对任意的$x_i \in U$,有$\tilde{a}(x_i) \in [0,1]$。这里为了符号的简单,仍然用a标记标准化的条件属性。

试验设计如下:给定一个预处理过的数据集,用十折交叉验证方法得到试验结果。具体地,所有样例被平均等分为10份,每一份轮流作为测试集,剩下的9份作为训练集。对任意一个训练集中的标准化条件属性a,定义一个模糊关系

$$R_{\{a\}}(x_i,x_j)=1-|a(x_i)-a(x_j)| \tag{12}$$

式中x_i和x_j为该训练集中的对象(样例);若该训练集中有符号值的条件属性a,则定义

$$R_{\{a\}}(x_i,x_j)=\begin{cases}1 & a(x_i)=a(x_j)\\ 0 & a(x_i)\neq a(x_j)\end{cases} \tag{13}$$

由此每一个训练集都转化为一个模糊决策系统。利用算法1在该训练集上对给定的阈值α选取样例进而选择特征子集。选择的样例和特征子集用来构造k-最近邻分类器(其中k=1,即1NN)和线性支撑向量机(LSVM),其中分类器的参数均为默认设置。构造好的分类器用来对测试集获取分类精度(测试精度)以检验算法1的有效性。这个过程对每一对训练集和测试集都执行一次,因而最终报告的试验结果是10次试验结果的平均值。

这里要指出的是,需对每一个对象得到的模糊下近似值$\lambda_i=\underline{R_A}[x_i]_D(x_i)$进行标准化,即

$$\tilde{\lambda}_i=\frac{\lambda_i-\min\limits_j \lambda_j}{\max\limits_j \lambda_j-\min\limits_j \lambda_j} \tag{14}$$

再令筛选样例的阈值α取值范围设置为0到1,步长为0.05。对模糊下近似值进行标准化的原因是需要对所有数据集的α取值统一标准。

3.2　试验结果

图1—4分别描述了各个数据集在不同的阈值α下选择的样例的平均个数、选择的特征的平均个数、特征选择的平均时间和获取的平均分类精度。由图1和图2很容易看到随着α值的增加,算法1选择的样例和特征的平均个数都单调地减少。此外,实数值的数据集Wine,WDBC,Libras,Steel和CTG随着α趋于1,选取的样例或特征的平均个数也趋于0;而对另外3个混合数据集Heart,Horse和Credit来说,当α趋于1时选取的样例或特征依然比较多,这也说明了实数值样例的标准化模糊下近似值大多小

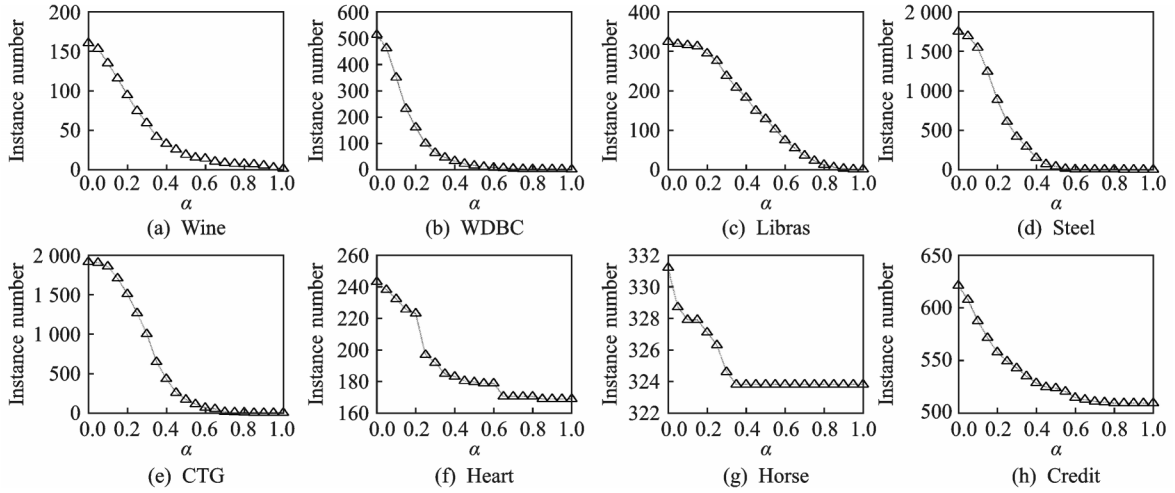

图1　不同阈值α下筛选样例的平均个数

Fig.1　Average number of selected instances with different threshold values α

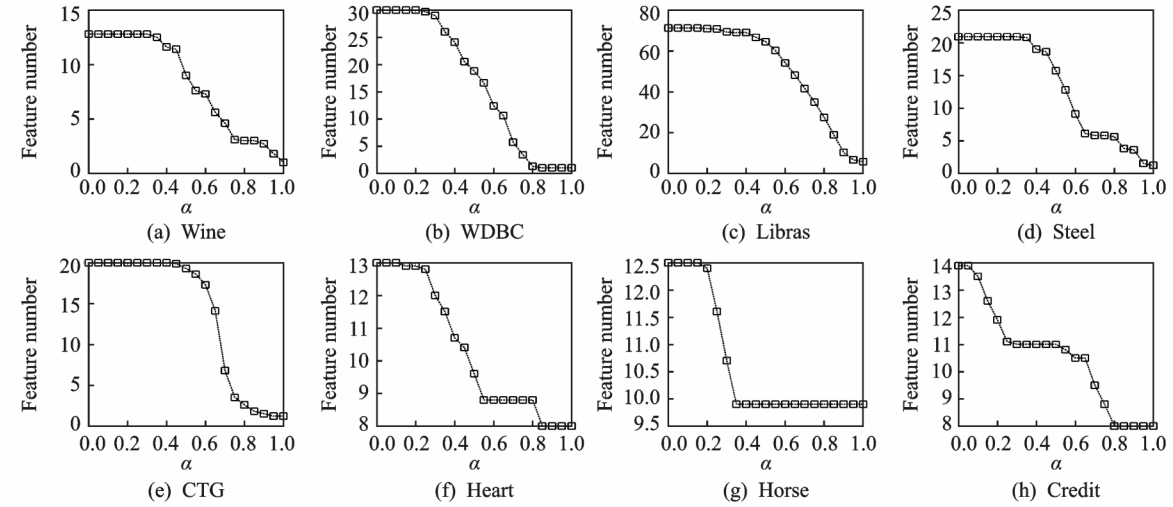

图2 不同阈值α下选择特征的平均个数

Fig.2 Average number of selected features with different threshold values α

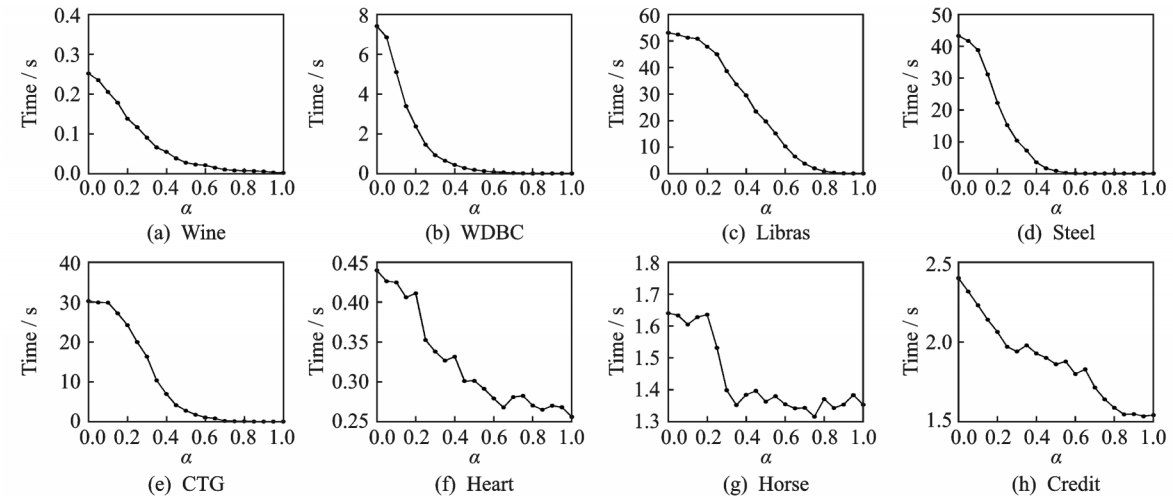

图3 不同阈值α下特征选择过程的平均运行时间

Fig.3 Average running time of feature selection process with different threshold values α

于1而混合值样例的标准化模糊下近似值大多等于1,这主要由本文针对实数值和符号值的条件属性所定义的模糊关系决定。由图3也容易看到随着α值的增加,特征选择过程的平均运行时间也大致地单调减少,尤其对实数值的数据集Wine,WDBC,Libras,Steel和CTG而言,运行时间在α大致对应的区间(0,0.5)上减少得最快。由图4可以看到,当α趋于1时,实数值的数据集Wine,WDBC,Libras,Steel和CTG获取的分类精度急剧地减少,这是由于α趋于1造成选取的样例和特征过少而导致分类器拟合不足;对于混合数据集Heart,Horse和Credit,阈值α的变化对获取的精度并没有太大影响,这是因为变化的α使得选取的样例和特征仍然比较多,从而仍能较好地训练分类器。

从试验结果来看,阈值α的选取至关重要。表2和表3中分别列出在1NN和LSVM下每个数据集所对应的最佳阈值及其相应的试验结果。需要指出的是,这里的最佳阈值是获取的分类精度不会显著

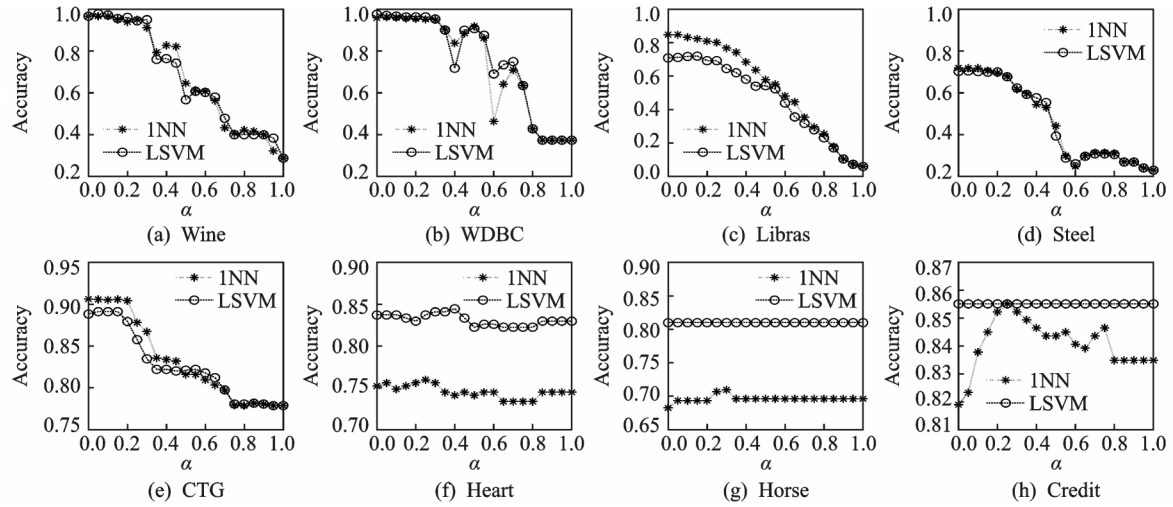

图4　不同阈值α下获取的平均分类精度

Fig.4　Average classification accuracy obtained by different threshold values α

表2　1NN所对应的最佳阈值α及相应的试验结果

Tab. 2　Best threshold and the related experimental results obtained by 1NN

Data set	Best α	Average number of selected instances	Average number of selected features	Average running time	
				Time/s	Ratio/%
Wine	0.15	115.4	12.8	0.18	70.7
WDBC	0.30	63.8	29.0	0.92	12.4
Libras	0.05	318.6	71.3	52.3	98.7
Steel	0.10	1 542.9	20.9	38.75	89.7
CTG	0.20	1 507.3	20.0	24.21	80.0
Heart	1.00	168.8	8.0	0.26	58.2
Horse	1.00	323.8	9.9	1.35	82.5
Credit	1.00	509.3	8.0	1.54	64.3

表3　LSVM所对应的最佳阈值α及相应的试验结果

Tab. 3　Best threshold and the related experimental results obtained by LSVM

Data set	Best α	Average number of selected instances	Average number of selected features	Average running time	
				Time/s	Ratio/%
Wine	0.30	59.0	12.8	0.09	35.8
WDBC	0.35	45.9	26.0	0.64	8.7
Libras	0.25	275.3	70.8	44.89	84.7
Steel	0.20	884.0	20.9	22.18	51.3
CTG	0.15	1 706.5	20.0	27.17	89.8
Heart	1.00	168.8	8.0	0.26	58.2
Horse	1.00	323.8	9.9	1.35	82.5
Credit	1.00	509.3	8.0	1.54	64.3

低于$\alpha=0$时所获取的分类精度的最大阈值。这里,采用Paired-t检验来验证分类精度的显著不同,其中显著性水平设为0.05。另外,表2和表3的最后一列指的是最佳阈值下的特征选择时间占$\alpha=0$时特征选择时间的比例,其值越小则意味着节约的计算时间越多。

综合表2和表3可以看出,对几乎所有数据集而言,最佳阈值α能有效地减少特征选择的计算时间而且对最终获取的分类精度没有显著影响。进一步地,对于实数值的数据集,合理的阈值α大概在区间[0.1,0.3]附近选取;对于混合数据集,阈值α可取为1。

4 结束语

本文提出了一种基于模糊粗糙集的快速特征选择算法,其思想是对样例先进行筛选,然后在筛选样例上进行特征选择。具体地,基于文献[12]的样例选择的思想,本文对模糊决策系统先进行样例筛选,即选择模糊下近似值不低于给定阈值α的那些样例,然后定义了筛选样例的单调信息熵用来作为特征选择的评估度量,并给出了相应的特征选择算法。试验结果表明本文提出的特征选择算法能有效减少计算时间且不会明显降低特征子集所得的精度,另外也分别针对实数值的数据集和混合数据集给出了控制筛选样例个数的阈值α的建议。

参 考 文 献:

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[2] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General System, 1990, 17(2/3): 191-209.

[3] Vluymans S, D′eer L, Saeys Y, et al. Applications of fuzzy rough set theory in machine learning: A survey[J]. Fundamenta Informaticae, 2015, 142(1-4): 53-86.

[4] Tsang E C C, Chen Degang, Yeung D, et al. Attributes reduction using fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2008, 16(5): 1130-1141.

[5] Hu Qinghua, Yu Daren, Xie Zongxia. Information-preserving hybrid data reduction based on fuzzy-rough techniques[J]. Pattern Recognition Letters, 2006, 27(5): 414-423.

[6] Jensen R, Shen Q. Fuzzy-rough sets assisted attribute selection[J]. IEEE Transactions on Fuzzy Systems, 2007, 15(1): 73-89.

[7] Zhang Xiao, Mei Changlin, Chen Degang, et al. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy[J]. Pattern Recognition, 2016, 56: 1-15.

[8] Wang Changzhong, Qi Yali, Shao Mingwen, et al. A fitting model for feature selection with fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2017, 25(4): 741-753.

[9] 徐菲菲, 苗夺谦, 魏莱, 等. 基于互信息的模糊粗糙集属性约简[J]. 电子与信息学报, 2008, 30(6): 1372-1375.
Xu Feifei, Miao Duoqian, Wei Lai, et al. Mutual information-based algorithm for fuzzy-rough attribute reduction[J]. Journal of Electronics & Information Technology, 2008, 30(6): 1372-1375.

[10] 赵军阳, 张志利. 基于模糊粗糙集信息熵的蚁群特征选择方法[J]. 计算机应用, 2009, 29(1): 109-111.
Zhao Junyang, Zhang Zhili. Ant colony feature selection based on fuzzy rough set information entropy[J]. Journal of Computer Applications, 2009, 29(1): 109-111.

[11] Qian Yuhua, Wang Qi, Cheng Honghong, et al. Fuzzy-rough feature selection accelerator[J]. Fuzzy Sets and Systems, 2015, 258: 61-78.

[12] Jensen R, Cornelis C. Fuzzy-rough instance selection[C]//2010 IEEE International Conference on Fuzzy Systems. Piscataway, NJ: IEEE, 2010: 1-7.

[13] Verbiest N, Cornelis C, Herrera F. FRPS: A fuzzy rough prototype selection method[J]. Pattern Recognition, 2013, 46(10): 2770-2782.

[14] Tsang E C C, Hu Qinghua, Chen Degang. Feature and instance reduction for PNN classifiers based on fuzzy rough sets[J]. International Journal of Machine Learning and Cybernetics, 2016, 7(1): 1-11.

[15] Anaraki J R, Samet S, Lee J H, et al. SUFFUSE: Simultaneous fuzzy-rough feature-sample selection[J]. Journal of Advances in Information Technology, 2015, 6(3): 103-110.

[16] Derrac J, Cornelis C, García S, et al. Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection[J]. Information Sciences, 2012, 186(1): 73-92.

[17] He Qiang, Xie Zongxia, Hu Qinghua, et al. Neighborhood based sample and feature selection for SVM classification learning [J]. Neurocomputing, 2011, 74(10): 1585-1594.

[18] Mac P N, Jensen R. Simultaneous feature and instance selection using fuzzy-rough bireducts[C]//2013 IEEE International Conference on Fuzzy Systems. Piscataway, NJ: IEEE, 2013: 1-8.

作者简介:

张晓(1986-),女,讲师,博士,研究方向:粒计算、粗糙集和统计计算,E-mail: zhangxiao@xaut.edu.cn。

杨燕燕(1986-),女,博士,研究方向:粗糙集、模糊集和机器学习,E-mail: yangyanyan@mail.tsinghua.edu.cn。

(编辑:刘彦东)