基于近似 Markov Blanket 和动态互信息的特征选择算法

姚 旭 王晓丹 张玉玺 权 文
(空军工程大学导弹学院计算机工程系 三原 713800)

摘 要 针对大量无关和冗余特征的存在可能降低分类器性能的问题,提出了一种基于近似 Markov Blanket 和动态互信息的特征选择算法。该算法利用互信息作为特征相关性的度量准则,并在未识别的样本上对互信息进行动态估值,利用近似 Markov Blanket 原理准确地去除冗余特征,从而获得远小于原始特征规模的特征子集。通过仿真试验证明了该算法的有效性。以支持向量机为分类器,在公共数据集 UCI 上进行了试验,并与 DMIFS 和 ReliefF 算法进行了对比。试验结果证明,该算法选取的特征子集与原始特征子集相比,以远小于原始特征规模的特征子集获得了高于或接近于原始特征集合的分类结果。

关键词 特征选择,相关性,Markov Blanket,互信息
中图法分类号 TP391 **文献标识码** A

Feature Selection Algorithm-based Approximate Markov Blanket and Dynamic Mutual Information

YAO Xu WANG Xiao-dan ZHANG Yu-xi QUAN Wen
(Department of Computer Engineering, Missile College, Air Force Engineering University, Sanyuan 713800, China)

Abstract To resolve the poor performance of classification owing to the irrelevant and redundancy features, feature selection algorithm based on approximate Markov Blanket and dynamic mutual information was proposed. The algorithm uses mutual information as the evaluation criteria of feature relevance, which is dynamically estimated on the unrecognized samples. Redundancy features were removed exactly by approximate Markov Blanket. So a small size feature subset can be attained with the proposed algorithm. To attest the validity, we made experiments on UCI data sets with support vector machine as the classifier, compared with DMIFS and ReliefF algorithms. Experiments result suggest that, compared with original feature set, the feature subset size obtained by the proposed algorithm is much less than original feature set and performance on actual classification is better than or as good as that by original feature set.

Keywords Feature selection, Relevance, Markov Blanket, Mutual information

1 引言

随着计算机和数据库技术的迅速发展,人们获得的信息量越来越大。人类知识积聚的同时,"维数灾难"的问题也随之而来;特征维数过高和训练样本不足也常常导致"过拟合"现象的发生;高维特征空间中存在着无关特征和冗余特征更导致了问题的复杂化。为了解决这些问题,特征选择应运而生。

特征选择是统计学、机器学习和数据挖掘等领域中的经典研究问题。通过特征选择,去除无关和冗余特征,最终提高分类系统的性能,降低计算代价。搜索策略和评价准则是特征选择过程中的两个关键步骤。好的搜索策略可以加快选择速度,好的评价准则可以保证所选择的子集包含丰富的信息。信息熵和互信息等信息测度是目前普遍采用的评价准则,因为它能够以量化的形式度量特征间的不确定程度,并且能有效地度量特征间的非线性关系。信息测度已被多次引入到特征选择的过程中。如 Sylvain 等[1]提出了基于互信息的特征选择算法并将它应用于故障诊断和识别中,取得了良好的效果;Guo 等[2]用互信息来度量两个变量之间的统计相关性,用来进行步态识别;Estévez 等[3]将正规化互信息和遗传算法相结合,提出了一种混合式的特征选择算法;赵军阳等[4]将互信息和模糊粗糙集结合,提出最大互信息最大相关熵标准,并基于这一标准,设计了一种新的特征选择算法等。

特征选择本质上就是选取一个与类别的相关性大、彼此之间的相关性小的特征子集,即最大相关最小冗余。本文利用互信息和条件互信息作为度量相关和冗余的标准。我们知道,随着特征的选取,数据集中未被识别的样本数逐渐减少,如果不断删除已识别的样本,使互信息和条件互信息在未识别的样本上动态估值,不仅能保证信息度量的准确性,而且能够提高系统的运行效率。本文在分析 Markov Blanket 原理的基础上,以动态互信息为度量标准,提出了一种基于近似 MarkovBlanket 和动态互信息的特征选择算法。为了验证算法的有效性,以支持向量机(Support Vector Machine, SVM)作为分类器,在 UCI 数据集上进行试验,并与其它两种特征

到稿日期:2011-09-23 返修日期:2011-12-22 本文受国家自然科学基金项目(60975026)资助。
姚 旭(1982—),女,博士生,主要研究方向为智能信息处理和机器学习,E-mail:ffxy132@163.com;王晓丹(1966—),女,教授,博士生导师,主要研究方向为智能信息处理和机器学习;张玉玺 男,博士生;权 文 女,博士生。

选择方法进行对比，最后给出试验结果和分析；本文第 2 节介绍了 Markov Blanket 和互信息相关知识；第 3 节在分析特征相关性度量的基础上，提出了一种基于近似 Markov Blanket 和动态互信息的特征选择算法；第 4 节给出试验结果和分析；最后进行总结。

2 相关知识

2.1 Markov Blanket

1996 年，Koller 和 Sahami[5] 首次将 Markov Blanket 引入到特征选择中，在删除特征的过程中以特征中是否存在 Markov Blanket 为标准。下面给出 Markov Blanket 的一些基本概念，关于 Markov Blanket 的更多知识可参见文献[6]。

定义 1(Markov Blanket) 给定一个特征 f_i，设特征子集 $M_i \subset F(f_i \notin M_i)$，称 M_i 是 f_i 的 Markov Blanket 当且仅当在给定 M_i 的条件下，f_i 和 $F-M_i-\{f_i\}$ 是独立的，即 $P(F-M_i-\{f_i\}|f_i,M_i)=P(F-M_i-\{f_i\}|M_i)$。

推论 1 如果特征子集 M_i 是 f_i 的 Markov Blanket，那么在给定 M_i 的条件下，f_i 与类 C 也是独立的，即 $P(C|f_i,M_i)=P(C|M_i)$。

定义 2(相对冗余)[7] 设 M 是一个特征集合，如果特征 f_i 在 $M-\{f_i\}$ 中存在 Markov Blanket，那么 f_i 相对于 M 来说是冗余的。

定义 3(近似 Markov Blanket) 设有特征 f_i 和 f_j，C 为类别，R 为一种度量准则，$R(f_i;C)>R(f_j;C)$，f_i 是 f_j 的一个近似 Markov Blanket，当且仅当 $R(f_j;C|f_i)>R(f_j;C)$。

在 Markov Blanket 理论中，当去除某些冗余特征后在开始阶段被剔除的特征仍然是冗余的[5]，但对于近似 Markov Blanket 则不一定成立。由于强相关特征不存在近似 Markov Blanket，因此在任何阶段强相关特征都不会被删除。所以，利用近似 Markov Blanket 同样可以得到一个最优子集。

2.2 互信息和条件互信息

互信息(Mutual Information，MI)是为了衡量两个变量间相互依赖强弱程度而引入的，它表示两个变量间共同拥有信息的含量，也可以理解为在已知一个变量的情况下，另外一个变量不确定性的减少程度。给定两个随机变量 X 和 Y，若它们的边缘概率分布分别为 $p(x)$ 和 $p(y)$，联合概率分布为 $p(x,y)$，则它们之间的互信息 $I(X;Y)$ 定义为

$$\begin{cases} I(X;Y)=\int_y\int_x p(x,y)\log\dfrac{p(x,y)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y & \text{(连续变量)} \\ I(X;Y)=\sum_y\sum_x p(x,y)\log\dfrac{p(x,y)}{p(x)p(y)} & \text{(离散变量)} \end{cases}$$

条件互信息(Condition Mutual Information，CMI)是指在给定某个随机变量的条件下，其它两个变量之间的相互关联程度。假如随机变量 Z 是已知的，那么变量 X 和 Y 关于 Z 的条件互信息为

$$\begin{cases} I(X;Y|Z)=\int_z\int_y\int_x p(x,y,z)\log\dfrac{p(x,y|z)}{p(x|z)p(y|z)}\mathrm{d}x\mathrm{d}y\mathrm{d}z \\ I(X;Y|Z)=\sum_z\sum_y\sum_x p(x,y,z)\log\dfrac{p(x,y|z)}{p(x|z)p(y|z)} \end{cases}$$

式中，$p(x,y,z)$ 为联合概率分布，$p(x|z)$，$p(y|z)$，$p(x,y|z)$ 均为条件概率分布。

3 基于近似 Markov Blanket 和动态互信息的特征选择算法

3.1 特征相关性度量

特征按其与类别的关系可以分为强相关特征、弱相关特征和无关特征[8]。强相关特征影响着类别的分布，由于缺少必然改变类别的分布情况，因此是最优子集的一部分；弱相关特征在一定条件下影响类分布，但不一定是必需的；无关特征对类别分布没有影响，应首先删除。因此一个最优特征子集应该包括所有强相关特征和部分弱相关特征，不包含无关特征。那么在进行特征选择之前，就要定义相关特征和冗余特征的度量标准。

信息度量标准是特征选择方法中常用的一种评价准则。信息度量主要利用信息熵等量化特征相对于分类类别的不确定性程度，以判定其包含的分类信息含量。信息度量是一种无参的、非线性的标准，且它不需要预先知道样本数据的分布。由于信息熵能很好地量化特征相对于类别的不确定性程度，因此它在特征选择算法中得到了广泛关注。除信息熵外，信息增益、互信息和最小描述长度等也都经常出现在特征选择算法中，其中信息增益用来说明类别的先验熵与其它特征组成的后验熵的差，而互信息则表示在给定类别的情况下，特征不确定性减少的程度。事实上，由贝叶斯公式可知，它们两者是等价的。在特征的相关性和冗余性分析中，采用互信息作为特征间的度量标准。当两变量完全无关或互相独立时，它们的互信息为 0，意味着两者之间不存在相同的信息；而互信息值越大，意味着所包含相同的信息也越多。因此，$I(f_i;C)$ 越大，表示类别 C 对特征 f_i 的依赖性越大。可见，用互信息可以很好地度量特征和类别的相关性。在算法中，首先利用特征与类别的互信息 $I(f_i;C)$ 选出一组与类别有相关性的特征集合，再进行冗余分析。下面分析特征之间的冗余度量准则。

同样地，可以根据互信息来度量两个特征之间的相互关联程度。但是，以此来确定特征是否冗余也存在一定的困难。例如当两个特征彼此不完全相关时，很难判断哪一个特征是冗余的，而且很可能要对全部特征计算 $\frac{d(d-1)}{2}$（其中 d 为特征的维数）个特征间的相互关联度，这在高维数据集中效率是极低的。因此直接计算特征间的互信息来判断冗余特征是不可行的。由 Markov Blanket 的定义可知，Markov Blanket 方法可以有效地去除冗余特征。很多文献也将 Markov Blanket 应用于特征选择中，如文献[9-12]。由于 Markov Blanket 是通过特征子集来计算的，其计算量比较大，因此我们利用近似 Markov Blanket 来近似地确定冗余特征。用互信息和条件互信息作为度量准则，用 $I(f_i;C)$ 表示特征 f_i 与类别 C 的相关性，$I(f_i;f_j)$ 表示特征 f_i 与 f_j 之间的相关性。$I(f_j;C|f_i)$ 表示给定特征 f_i 的条件下，特征 f_j 与类别 C 的相关性。利用近似 Markov Blanket 确定冗余特征的判断准则如下：

如果 $I(f_i;C)>I(f_j;C)$，$I(f_j;C|f_i)>I(f_j;C)$，那么 f_i 是 f_j 的一个近似 Markov Blanket。即特征 f_j 是冗余的，应该删除。

3.2 基于近似 Markov Blanket 和动态互信息的特征选择算法

在度量特征的相关性程度时，数据样本集给定以后，特征在这个样本集的概率分布也就随之确定下来。这种确定性使得度量标准不能准确反映信息或不确定性的动态变化情况，因为特征选择是一个动态的过程，即随着已选特征增多，类别 C 的不确定性逐渐降低，同时样本空间中不可识别的样本数量也呈减少的趋势，这在某种程度上说明不发生变化的相关性度量标准包含了部分"假"信息[13]。因此，如果在特征选择过程中，不断地删除已被识别的样本，使得评价标准在未识别样本上动态估值，这不仅可以更准确地反映相关程度，而且提高了计算效率。刘华文[13]提出了一种基于动态互信息的特征选择算法 DMIFS，算法通过每次选择一个与类别相关性最大的特征加入到已选特征集合中，再不断删除该特征能识别的样本，使得互信息在不断更新的样本集上动态估值，从而保证了相关性度量的准确性。但是，算法中也存在一个问题，即它只考虑了特征与类别的相关性，没有考虑特征之间的相关性。因此所选择的特征中可能存在冗余。从 3.1 节的分析中已经知道，近似 Markov Blanket 是一种去除冗余和无关特征的有效方法。本文以动态互信息为评价标准，采用近似 Markov Blanket 去除冗余和无关特征，提出一种基于近似 Markov Blanket 和动态互信息的特征选择算法，此算法简称为 AMBDMI。算法分为两个阶段，第一阶段先选择与类别有相关性的特征集合，第二阶段用近似 Markov Blanket 去除冗余特征。设样本数据集为 D，类别为 $C=(c_1,c_2,\cdots,c_m)$，特征集合 $F=(f_1,f_2,\cdots,f_d)$，已选特征集合记为 S，算法具体步骤见表 1。

表 1 基于近似 Markov Blanket 和动态互信息的特征选择算法(AMBDMI)

输入：训练数据集 $D=(x_1,x_2,\cdots,x_N)$，其中 $x_i=(x_{i1},x_{i2},\cdots,x_{id})$；类别 $C=\{c_1,c_2,\cdots,c_m\}$；$F=(f_i\mid i\in 1,2,\cdots,d)$ 是原始特征集合；阈值 p。
Step1 参数初始化，已选特征集合 $S=\phi$，$D_l=\phi$。
Step2 对于 F 中所有特征值 f_i 计算 $I(f_i;C)$。如果 $I(f_i;C)=0$，则从 F 中去除 f_i。
Step3 对 F 中剩余的特征按照 $I(f_i;C)$ 的值降序排列，所构成的特征集合记为 F'。
Step4 取 F' 中的第一个特征，即 $f=\arg\max I(f_i;C)$，将 f 加入到已选特征集合 S 中，即 $S=S+\{f\}$，$F'=F'-\{f\}$。
Step5 由特征 f 得到 f 所能识别的样本集合 D_l，更新数据集 $D=D-D_l$。如果 D 为空或者在总样本中所占比例小于阈值 p，算法停止。
Step6 取 F' 中的下一个特征 f，如果 f=NULL，算法停止；否则执行 Step7。
Step7 对任意的 $f_i\in S$，如果 $I(f;C\mid f_i)>I(f;C)$，则从 F' 中将 f 删除，即 $F'=F'-\{f\}$，返回 Step6。否则，将 f 加入到 S 中，并从 F' 中删除 f，即 $S=S+\{f\}$，$F'=F'-\{f\}$，返回到 Step5。
输出：特征子集 S

4 试验结果及分析

通过试验，我们将验证所提的基于近似 Markov Blanket 和动态互信息的特征选择算法 AMBDMI 的可行性和有效性。

4.1 试验数据

试验中的数据均来自 UCI[14] 数据库中的数据集，选择了其中 8 组数据(特征维数范围为 4～60，样本范围为 150～699)，关于试验数据的详细描述如表 2 所列。试验前，对数据集中的数据进行归一化处理。

表 2 UCI 数据集各数据描述

Problem	# Train	# Attributes	# Classes
Breast-cancer-wisconsin	699	10	2
Hepatitis	155	19	2
Ionosphere	351	34	2
Sonar	208	60	2
Iris	150	4	3
Glass	214	10	7
Ecoli	336	8	8
Soybean	307	35	19

4.2 试验设计

为了验证所提算法 AMBDMI 的性能，在公共数据集 UCI 上进行试验，并与文献[13]提出的算法 DMIFS 和经典算法 ReliefF 进行对比。首先计算 8 个数据集上利用所有特征进行分类时的分类正确率；然后比较不同的特征选择算法在各个数据集上的分类正确率，并给出最好情况下不同特征选择算法所得到的子集规模；最后比较在相同子集规模的条件下，算法 AMBDMI 与 ReliefF 的分类正确率。

试验中以 SVM 为分类器，它来自 PRTool(http://www.prtools.org)工具箱，试验机器配置为 2G 内存，2.80G CPU，算法基于 Matlab7.10(R2010a)实现。

4.3 试验结果和分析

试验从分类正确率的角度对本文提出的算法 AMBDMI 与算法 DMIFS 和 ReliefF 进行比较。为了使试验结果更加可靠，训练过程采用 10 重交叉验证，进行 10 次试验，所有试验结果取 10 次试验的平均值。其中分类正确率后半部分数据处于置信水平为 95%的置信区间。表 3 给出了 8 个数据集上未进行特征选择时的 SVM 分类正确率。

表 3 各数据集上未进行特征选择的分类正确率及置信水平为 0.95 的置信区间(%)

数据集	分类正确率	数据集	分类正确率
Breast-cancer-wisconsin	96.68±1.56	Iris	95.80±3.56
Hepatitis	62.15±8.38	Glass	79.42±5.56
Ionosphere	91.25±3.22	Ecoli	86.02±4.64
Sonar	88.47±5.02	Soybean	90.96±3.19

由表 3 可以看出，除数据集 Hepatitis、Glass 以外，其余数据集中都只含有少量的噪声数据；数据集 Hepatitis 在应用所有特征进行分类的条件下，分类正确率仅为 62.15%，且置信区间较大；数据集 Glass 在应用所有特征进行分类的条件下，分类正确率为 79.42%，说明这两个数据集本身含有大量的噪声数据。表 4 给出了 3 种特征选择算法在 8 个数据集上的分类正确率比较。

表 4 不同特征选择算法得到的分类正确率及置信水平为 0.95 的置信区间(%)

数据集	算法		
	DMIFS	ReliefF	AMBDMI
Breast-cancer-wisconsin	95.94±.69	96.69±1.57	96.11±1.76
Hepatitis	65.77±6.72	66.37±7.4	66.18±7.69
ionosphere	88.29±2.86	91.25±2.16	90.32±2.40
Sonar	74.43±7.29	86.34±5.74	85.40±5.85
Iris	95.39±3.52	95.80±3.74	95.40±3.52
Glass	80.24±7.09	80.04±6.33	80.28±7.08
Ecoli	80.62±4.64	80.61±6.13	80.66±5.34
Soybean	89.99±4.51	91.67±4.54	91.86±3.52

从表 4 可以看出，在 Breast-cancer-wisconsin、Hepatitis、ionosphere、Soybean 4 个数据集上，AMBDMI 算法的分类正

确率较 DMIFS 算法分别提高了 0.17%、1.01%、2.03%、1.87%。在 Iris、Glass、Ecoli 3 个数据集上，算法 AMBDMI 与算法 DMIFS 的分类正确率基本持平。但在数据集 Sonar 上，与 DMIFS 相比，AMBDMI 算法的分类正确率提高了 10.97%。AMBDMI 算法与 ReliefF 算法的分类正确率基本持平。在数据集 Hepatitis、Glass、Soybean 上，相对于在未经过特征选择的数据集上的分类情况，算法 AMBDMI 的分类正确率分别提高了 4.03%、0.86%、0.90%。分类正确率不仅没有降低，反而提高了，这说明算法不仅降低了特征的维数，而且在一定程度上减少了个别特征噪声的影响，提高了 SVM 的分类正确率。分析 8 个数据集上每种特征选择算法的置信区间可以看出，在数据集 Breast-cancer-wisconsin、ionosphere、Iris、Ecoli、Soybean 上，3 种算法都比较稳定，但在数据集 Hepatitis、Sonar 和 Glass 上，不太稳定。在 Sonar 数据集上，算法 AMBDMI 相对于 DMIFS 较为稳定。从表 3 可以看出，在数据集 Hepatitis、Sonar 和 Glass 上的置信区间相对于其他 5 个数据集来说都较高。因此算法的不稳定性是与数据集本身有关的。

表 5 给出了最好情况下不同特征选择算法的最小特征子集规模比较。从表 5 可以看出，相对于 ReliefF 算法，DMIFS 和 AMBDMI 算法所得到的特征子集规模要小很多。这是因为 ReliefF 算法只能去除无关特征，不能去除冗余特征。因此用所选特征子集训练分类器，虽然正确率较高，但时间复杂度也较高。在数据集 Breast-cancer-wisconsin、Hepatitis 、Ecoli 和 Soybean 上，相对于算法 DMIFS，由算法 AMBDMI 得到的特征子集规模较小，且分类正确率相当。在数据集 Sonar 上，虽然由 AMBDMI 得到的子集规模大于由 DMIFS 得到的子集规模，但是 AMBDMI 的分类正确率明显高于 DMIFS。

表 5　最好情况下不同特征选择算法的最小特征子集规模比较

数据集	算法		
	DMIFS	ReliefF	AMBDMI
Breast-cancer-wisconsin	5	9	4
Hepatitis	6	16	2
ionosphere	2	15	2
Sonar	2	22	4
Iris	1	3	1
Glass	3	8	3
Ecoli	7	6	4
Soybean	14	23	13

为了更好地说明算法 AMBDMI 在较小的特征子集规模的条件下能够得到较高的分类正确率，将相同子集规模下算法 AMBDMI 与 ReliefF 的分类正确率进行比较，子集规模取由算法 AMBDMI 所得到的子集规模，如表 6 所列。

表 6　相同子集规模下算法 AMBDMI 与 ReliefF 分类正确率比较

数据集	算法		子集规模
	ReliefF	AMBDMI	
Breast-cancer-wisconsin	95.79±1.58	96.11±1.76	4
Hepatitis	60.35±5.49	66.18±7.69	2
ionosphere	78.64±4.35	90.32±2.40	2
Sonar	55.21±7.00	85.40±5.85	4
Iris	95.40±3.52	95.40±3.52	1
Glass	44.10±7.42	80.28±7.08	3
Ecoli	58.17±6.13	80.66±5.34	4
Soybean	73.62±5.28	91.86±3.52	13

从表 6 可以看出，在相同特征子集规模下，除了 Iris 数据集，在数据集 Breast-cancer-wisconsin、Hepatitis、ionosphere、Sonar、Glass、Ecoli 和 Soybean 上，算法 AMBDMI 得到的分类正确率相对于 ReliefF 算法分别提高了 0.32%，5.83%，11.68%，30.19%，36.18%，22.49%，18.24%。因此，算法 AMBDMI 能够有效地去除冗余特征，得到最小的特征子集，且并不以牺牲正确率为代价。

结束语　本文提出了一种基于近似 Markov Blanket 和动态互信息的特征选择算法。首先利用互信息度量特征与类别间的相关性，去除与类别无关的特征，然后利用近似 Markov Blanket 原理去除冗余特征。在特征选择的过程中，不断地用已选特征删除能够识别的样本，从而使互信息和条件互信息在未识别的样本上动态估值，保证了相关性度量的准确性。试验结果表明，该算法能够在保证 SVM 分类精度的情况下，有效地减少特征维数，加速了 SVM 训练，具有较好的性能。

参 考 文 献

[1] Verron S, et al. Fault detection and identification with a new feature selection based on mutual information [J]. Journal of Process Control, 2008, 18(5): 479-490

[2] Guo Bao-feng, Nixon M S. Gait Feature Subset Selection by Mutual Information [J]. IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans, 2009, 39(1): 36-46

[3] Estévez P A, et al. Normalized Mutual Information Feature Selection [J]. IEEE Transactions on Neural Networks, 2009, 20(2): 189-201

[4] 赵军阳，张志利. 基于最大互信息最大相关熵的特征选择方法[J]. 计算机应用研究，2009，26(1)：233-235

[5] Koller D, Sahami M. Toward Optimal Feature Selection [C]// Proceedings of International Conference on Machine Learning. ICML 1996. 1996: 284-292

[6] Pearl J. Probabilistic Reasoning in Intelligent Systems [M]. Morgan Kaufmann, San Mateo, CA, 1988

[7] 崔自峰，等. 一种近似 Markov Blanket 最优特征选择算法[J]. 计算机学报，2007，30(12)：2074-2081

[8] John G H, Kohavi R, Pfleger K. Irrelevant feature and the subset selection problem [C]// Proceedings of the 11th International Conference on Machine Learning. New Jersey, 1994: 121-129

[9] Yaramakala S, et al. Speculative Markov Blanket Discovery for Optimal Feature Selection [C]// Proceedings of the Fifth IEEE International Conference on Data Mining. 2005, 5: 1550-4786

[10] Knijnenburg T A, et al. Artifacts of Markov blanket filtering based on discretized features in small sample size applications [J]. Pattern Recognition Letters, 2006, 27: 709-714

[11] Zhao Hui, et al. Optimal feature selection based on Bayesian networks [C]// Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition. Beijing, China, Nov. 2007: 597-601

[12] Castro P A D, et al. Learning Bayesian Networks to Perform Feature Selection [C]// Proceedings of International Joint Conference on Neural Networks. Atlanta, Georgia, USA, June 2009: 467-473

[13] 刘华文. 基于信息熵的特征选择算法研究[D]. 吉林：吉林大学，2010

[14] Hettich S, Bay S D. The UCI KDD Archive [DB/OL]. http://kdd.ics.uci.edu/, 1999