

基于正则化互信息和差异度的集成特征选择

姚旭 王晓丹 张玉玺 薛爱军

(空军工程大学防空反导学院 西安 710051)

摘要 如何构造差异性大的基分类器是集成学习研究的重点,为此提出迭代循环选择法;以最大化正则互信息为准则提取最优特征子集,进而基于此训练得到基分类器;同时以错分样本个数作为差异性度量准则来评价所得基分类器的性能,若满足条件则停止,反之则循环迭代直至结束。最后用加权投票法融合所选基分类器的识别结果。通过仿真实验验证算法的有效性,以支持向量机为分类器,在公共数据集 UCI 上进行实验,并与单 SVM 及经典的 Bagging 集成算法和特征 Bagging 集成算法进行对比。实验结果显示,该方法可获得较高的分类精度。

关键词 集成学习,集成特征选择,互信息,差异性

中图分类号 TP391 文献标识码 A

Ensemble Feature Selection Based on Normalized Mutual Information and Diversity

YAO Xu WANG Xiao-dan ZHANG Yu-xi XUE Ai-jun

(School of Air and Missile Defense, Air Force Engineering University, Xi'an 710051, China)

Abstract How to generate classifiers with higher diversity is an important problem in ensemble learning, consequently, an iterative algorithm was proposed as follows: base classifier is trained using optimal feature subset which is selected by maximum normalized mutual information, simultaneously, the attained base classifier is measured by the diversity based on the number of miss classified samples. The algorithm stops if satisfy, otherwise iterates until end. Finally, weighted voting method is utilized to fusion the base classifiers' recognition results. To attest the validity, we made experiments on UCI data sets with support vector machine as the classifier, and compared it with Single-SVM, Bagging-SVM and AB-SVM. Experimental results suggest that our algorithm can get higher classification accuracy.

Keywords Ensemble learning, Ensemble feature selection, Mutual Information, Diversity

1 引言

集成学习是近年来机器学习领域研究的经典问题。其中个体分类器的生成方法是集成学习研究的一个重点。围绕如何产生差异性大的个体分类器,研究者们提出了多种方法,如扰动训练样本集的方法、扰动特征空间的方法、扰动分类器的方法等。文献[1]证明基于特征划分的集成方法比基于数据划分的集成方法具有更好的性能。如何有效地生成特征子集是该类方法需要解决的核心问题。Ho 的随机子空间方法^[2]和 Robert Brylla 提出的 Attribute-Bagging^[3]方法就是这类方法的典型代表。然而这类方法只是随机地选择特征,并没有对特征的相关冗余进行分析。目前最常采用的是通过特征选择方法来获得特征子集,Opitz D 将其称为集成特征选择^[1]。与传统的特征选择不同,集成特征选择不再是选择与原问题具有同等或更优预测能力的最佳特征子集,而是选择多个特征子集,用于生成多个个体分类器以达到提高集成中个体差异度的目的。如 Oliveira 等^[4]提出了基于多目标优化的 GA 的特征选择方法;李霞等^[5]针对不平衡数据集提出了一种新的集成特征选择方法;孙亮等^[6]提出了一种基于广义粗集的

集成特征选择方法等。

张宏达等^[7]从差异度量方法、差异度与分类器集成性能的关系以及差异度在集成优化中的应用 3 个方面对当前分类器集成研究进展进行了分析,指出对差异性的研究在集成学习领域中占有基础性地位。本文重点研究了如何构造差异性大的基分类器,在分析特征相关冗余度和分类器差异性的基础上,提出了一种基于正则化互信息和差异度的集成特征选择方法。算法中利用特征的相关冗余准则选择特征子集,使得特征与类别的相关性最大,彼此之间的冗余度最小,从而保证了较高的分类正确率。利用基分类器间的差异度来度量特征子集之间的相关性,生成使得基分类器间的差异性大的不同特征子集,从而提高了集成的性能。为了验证算法的有效性,以 SVM 作为分类器,在 UCI 数据集上进行了实验,最后给出实验结果和分析。

2 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

到稿日期:2012-09-11 返修日期:2012-12-24 本文受国家自然科学基金项目(60975026,61273275)资助。

姚旭(1982-),女,博士生,主要研究方向为智能信息处理、机器学习等;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、机器学习等。

个变量不确定性的减少程度。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;互信息值越大,意味着所包含的相同信息也越多。因此,如果 $I(f_i; C)$ 越大,表示类别 C 对特征 f_i 的依赖性越大。 $I(f_i; f_j)$ 越大,表示特征 f_i 和特征 f_j 之间的相关性越大。因此用互信息可以很好地度量特征和类别的相关性以及特征之间的相关性。

正则化互信息(Normalized Mutual Information, NMI)就是对互信息的归一化处理,将其值限定在 $[0, 1]$ 范围内。对于两个随机变量 X 和 Y ,如果 $H(X), H(Y)$ 表示变量 X 和 Y 的熵, $I(X; Y)$ 表示互信息,那么它们的正则化互信息定义为:

$$NI(X, Y) = \frac{I(X; Y)}{\min\{H(X), H(Y)\}} \quad (1)$$

3 基于正则化互信息和差异度的集成特征选择方法

3.1 差异度

集成的差异性确保集成将弱学习器提升为强学习器的必要因素,是集成中评价成员分类器优劣的一个重要标准。集成差异性的定义目前仍未统一,人们最普遍接受的是 Dietterich^[8] 的集成差异性定义:各个分类器对待测样本做出不同错误的趋势。关于分类器差异性度量,研究者们提出了诸多方法,如 Q 统计^[9]、不一致性度量^[9]、Kappa 一致性^[10] 等,这些方法的共同点是都考虑了分类器同时正确分类的样本。作者认为,撇开基分类器的分类精度,好的基分类器应该是错误不相关的,这样才有利于集成性能的提高,因此只需要考虑每个基分类器错分的那些样本,而分类器都正确分类的样本对集成性能的提高贡献不大,还可能使得对差异度的刻画变得模糊。基于以上分析,本文提出了一种以分类器错误不相关度为指标的差异性度量标准——互补性差异度:

$$\eta_{ij} = \frac{n_i + n_j - 2n_{ij}}{n_i + n_j} \quad (i, j = 1, 2, \dots, L) \quad (2)$$

那么分类器间的差异度即为各基分类器两两之间差异度的平均值,即

$$\eta = \frac{2}{L(L-1)} \sum_{j=2}^L \sum_{i=1}^{j-1} \eta_{ij} \quad (3)$$

式中, n_i 为第 i 个分类器的错分样本数, n_j 为第 j 个分类器的错分样本数, n_{ij} 为第 i 和 j 个分类器都错分的样本数, L 为基分类器的个数。

显然,两个分类器错误不相关度越高,即第 i 个分类器错分的样本被第 j 个分类器正确分类的比率越大,同时,第 j 个分类器错分的样本被第 i 个分类器正确分类的比率越大,两个分类器的互补性差异度越大,其组合效果越好。从式(2)可以看出, $\eta_{ij} \in [0, 1]$, $\eta_{ii} = 0$, 当 $n_{ij} = 0$ 时 $\eta_{ij} = 1$ 。

3.2 特征的相关冗余度量标准

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。强相关特征影响着类别的分布,如果缺少了则必然会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。因此特征选择的任务就是去除无关特征以及相对于其它特征冗余的弱相关特征。信息测度是目前普遍采用的特征相关性评价准则,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。Bonlander^[11] 首次引入了互信

息的概念来度量特征空间中的相关性。比起其它的度量标准,互信息有两个显著的优点,一是可以度量任意关系的两个变量,二是对特征空间的变换具有不变性,即它的值不依赖于坐标轴的选择。

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;互信息值越大,意味着所包含的相同的信息也越多。因此互信息可以很好地度量特征和类别的相关性以及特征之间的相关性。很多文献都利用互信息来度量特征的相关和冗余,比较典型的有 Kwak^[12] 等提出的 MIFS-U 方法、Peng 等^[13] 提出的 mRMR 方法、Pablo 等^[14] 提出的 NMIFS 方法等。L. T. Vinh^[15] 分析了文献[12-14]中方法的缺陷,提出了一种新的相关冗余分析标准:

$$J(f_j) = NI(C; f_j) - \frac{1}{|S_{i-1}|} \sum_{f_i \in S_{i-1}} NI(f_i; f_j) \quad (4)$$

式中, NI 表示正则化互信息,是为了保持相关性和冗余性两部分的平衡引入的^[14]。在式(4)中 $NI(C; f_j) = \frac{I(C; f_j)}{\log_2(|\Omega_C|)}$,

$NI(f_i; f_j) = \frac{I(f_i; f_j)}{\log_2(N)}$ 。其中 $|\Omega_C|$ 为类别空间的维数, N 为特征的量化水平,详细说明参见文献[15]。

3.3 算法框架

Hansen 和 Salamon^[16] 从理论上证明了集成方法能否奏效的关键在于,个体分类器的错误率低于 0.5 且错误互不相关。因此,要想提高集成的性能,就要尽量构造分类正确率高的基分类器,并且保证基分类器间的错误相关性尽可能小,即使基分类器间的差异性尽可能大。由特征的相关冗余度量标准可知,一个好的特征子集应该包含与类别的相关性尽可能大而彼此之间的相关性尽可能小的特征。基于正则化互信息的相关冗余度量标准能够有效地去除无关和冗余特征,因此基于此准则选择的最优特征子集训练的基分类器具有较高的分类正确率和较低的时间复杂度,从而满足了集成中基分类器的一个条件——分类正确率高。从提高分类器集成性能的另一个因素考虑,如果利用不同的特征子集训练的基分类器彼此之间的错误相关性小,那么集成的效果就好。基于特征选择的集成方法就是通过生成相关性小的不同特征子集来构造差异性大的基分类器集合,最终提高集成的性能。评价基分类器间关系的最直接的方法就是计算分类器间的差异度,选择差异度大的分类器集合用于集成学习,从而提高分类器集成的性能。

Bagging 是一种基于样本扰动的集成方法。但由于 SVM 是一种“稳定”的分类器,对样本扰动不敏感,导致单纯的基于 Bagging 集成的方法在以 SVM 为分类器时的效果并不理想,但 Bagging 集成在神经网络、决策树等分类器上的成功应用证明: Bagging 取样技术是一种很好的样本扰动机制。本文将样本扰动和特征扰动相结合,通过分析基分类器间差异度和特征的相关冗余标准,提出一种基于正则化互信息和差异度的集成特征选择方法,简称 NMI-D。算法借鉴特征选择算法中顺序前进法的基本思想,首先应用 Bagging 方法中的 bootstrap 抽样方法生成两个训练子集,然后利用 3.2 节介绍的特征相关冗余分析标准在每个训练子集上选择各自对应的特征子集,并将训练子集在相应的特征子集上进行投影生成新的训练子集,在新的训练子集上训练各基分类器。再利用 3.1 节介绍的差异度量方法计算生成的两个基分类器间的

差异度。如果该差异度大于给定的阈值 σ , 那么将这两个基分类器加入已选基分类器集合。然后按照同样的方法继续训练基分类器, 并计算它与已选基分类器集合中分类器间的差异度, 根据该差异度与阈值 σ 的关系对新生成的基分类器进行取舍。如此循环迭代, 直到基分类器个数达到事先设定的阈值。最后利用加权投票法融合各基分类器的输出结果。算法的具体步骤如下:

输入: 训练集 N , 特征集为 $F = \{f_i | i=1, \dots, D\}$, 测试样本 x , 训练样本子集规模 N_{sub} , 集成规模 L , 阈值 σ

输出: 测试样本的判决类标签

Step1 运用 Bagging 方法中的 bootstrap 抽样方法产生两个规模为 N_{sub} 的训练样本子集 T_1, T_2 ;

Step2 在训练样本子集 T_1, T_2 上进行如下操作: 利用互信息求出所有特征与类别的相关性 $I(f_i, C)$, 选出使互信息最大的特征 f 将其加入到 $S_i (i=1, 2)$ 中, 即 $f = \arg \max I(f_i, C), S_i = S_i + \{f\}$;

Step3 当 $|S_i| < |F|$ 时按以下准则进行递归计算:

$$f = \arg \max_{f_i \in F - S_{i-1}} NI(C; f_i) - \frac{1}{|S_{i-1}|} \sum_{f_j \in S_{i-1}} NI(f_i; f_j)$$

得到两个最优特征子集 S_1, S_2 ;

Step4 将训练子集 T_1, T_2 投影到相应的特征子空间上, 用投影后的训练子集训练 SVM 作为集成的基分类器, 得到 SVM_1, SVM_2 ;

Step5 将测试样本投影到特征子空间上, 并用相应的基分类器分类, 利用式(2)计算分类器间的差异度 η ;

Step6 如果 $\eta > \sigma$, 则将 S_1, S_2 放入已选特征子集集合 S 中, 将 SVM_1, SVM_2 放入基分类器集合 CS 中, 即 $S = \{S_1, S_2\}, CS = \{SVM_1, SVM_2\}$ 。否则, 转向 Step1;

Step7 当 $|S| \geq L$ 时, 转向 Step9; 否则, 继续运用 bootstrap 抽样方法产生规模为 N_{sub} 的训练样本子集 T_i , 利用 Step2-Step3 中特征子集生成方法生成特征子集 S_i , 训练基分类器 SVM_i , 并利用式(3)计算 SVM_i 与 CS 中基分类器成员的差异度 η_i 。

Step8 如果 $\eta_i > \sigma$, 则将 S_i 加入到集合 S 中, SVM_i 加入到集合 CS 中, 转向 Step7。否则, 直接转向 Step7;

Step9 通过以上步骤, 可以得到 L 个特征子集 $S = \{S_i | i=1, 2, \dots, L\}$ 及其对应的 L 个基分类器 $CS = \{SVM_i | i=1, 2, \dots, L\}$ 。把测试样本映射到每个特征子空间 S_i 上, 用 CS 中对应的基 SVM 对测试集进行分类, 采用加权投票法融合 CS 中各基分类器的分类结果。

4 实验结果及分析

4.1 实验数据

实验数据均来自 UCI 数据库, 选择了其中 8 组数据(特征维数范围为 8~60, 样本范围为 208~6435), 详细描述如表 1 所列。实验前, 对数据进行了归一化处理。

表 1 UCI 数据集各数据描述

数据集	训练集	测试集	维数	类别
Sonar	208	—	60	2
Glass	214	—	10	7
Soybean	307	—	35	19
Ecoli	336	—	8	8
Ionosphere	351	—	34	2
Segment	2310	—	19	7
Waveform	3000	2000	21	3
Satimage	4435	2000	36	6

4.2 实验结果和分析

为了验证算法 NMI-D 的性能, 在公共数据集 UCI 上进

行实验, 并与单分类器(Single)及著名的 Bagging 集成算法和特征 Bagging(AB)集成算法进行对比实验。在估计分类正确率时采用五重交叉验证, 并利用双边估计 t 检验法来计算置信水平为 0.95 的分类正确率置信区间, 作为最终结果。计算公式如下:

$$\frac{|\bar{x} - u|}{\sigma/\sqrt{n}} \geq t_{0.025}(n-1) \quad (5)$$

式中, u, σ 分别表示五重交叉验证的均值和标准差, $t_{0.025}(4) = 2.7764$ 。

实验分两部分: 第一部分比较固定基分类器数情况下 4 种算法的分类精度; 第二部分研究集成规模与集成精度之间的关系。实验中以 SVM 为分类器, 它来自 PRTool(<http://www.prtools.org>) 工具箱, 采用径向基核函数(Radial Basis Function, RBF)的 SVM。RBFSVM 有高斯宽度 σ 和规则化参数 C 两个参数, 任何一个的改变都会导致分类器性能的改变。通过选择合适的 C 和 σ 可以有效避免过拟合。对 RBFSVM 的性能分析发现^[17], C 值过小, 分类器学习能力就不好, 但当 C 在一个合适的范围内取值时, RBFSVM 的性能可以简单地通过调整 σ 值改变, 且 σ 对分类器的影响更大。文献[18]分析了在 RBFSVM 中如果 σ 取相同值而带来的一些问题, 提出了通过将训练每个基分类器的样本集的标准差作为该基分类器的 σ 值来控制基分类器的分类精度, 以避免参数 σ 在所有基分类器中的取值相同带来的问题。因此在本文中采用文献[18]的做法, 把训练每个基分类器的样本集的标准差作为该基分类器的 σ 值。惩罚因子 $C=1000$ 。实验机器配置为 2G 内存, 2.80G CPU, 算法基于 Matlab7.10(R2010a)实现。

1. 固定基分类器下的分类精度比较

实验中采用五重交叉验证, 进行 10 次实验, 所有结果取 10 次实验的平均值。文献[19]的实验指出, 对于多数模式识别问题, 基分类器数目达到 25 个左右时集成分类器的泛化性能已经趋于稳定, 因此在这部分实验中基分类器数设置为 25。经过多次实验, 差异度的阈值 σ 设为 0.68。实验结果如表 2 所列。

表 2 4 种算法的分类精度比较(基分类器数为 25)

Dataset	Single	Bagging	AB-SVM	NMI-D
Sonar	85.42±3.64	85.48±3.56	86.49±3.26	87.47±3.02
Glass	80.26±3.48	80.19±2.87	81.44±3.15	82.55±2.69
Soybean	91.79±3.52	92.54±3.51	93.16±3.34	94.31±2.89
Ecoli	80.53±2.24	81.23±2.49	82.02±2.37	82.65±2.84
Ionosphere	92.34±2.44	92.19±2.31	92.87±1.95	93.68±2.21
Segment	93.79±0.22	94.86±0.37	94.98±1.02	95.33±1.15
Waveform	84.36±1.82	89.35±1.77	90.37±1.95	91.31±1.54
Satimage	84.56±2.32	84.83±1.78	85.34±2.50	86.16±1.87
Average	86.61±2.46	87.66±2.33	88.33±2.44	89.18±2.28

表 2 中黑体表示 Single-SVM 分类精度好于 Bagging-SVM 分类精度的数据集。从表 2 可以看出, 相比于其它 3 种算法, NMI-D 算法在 8 个数据集上的分类精度均有所提高, 其中, 较 Single-SVM 提高了约 2.57%, 较 Bagging-SVM 提高了约 1.52%, 较 AB-SVM 提高了约 0.85%; AB-SVM 算法介于 NMI-D 算法和 Bagging-SVM 算法之间, 而 Bagging-SVM 算法在 Ionosphere 和 Glass 数据集上较 Single-SVM 的分类精度分别下降了 0.07% 和 0.15%, 在其它 6 个数据集上, 也较 Single-SVM 算法提升的幅度不大(在 Sonar 数据集上仅提

高了约 0.06%)。从以上的实验结果可以看出, Bagging 集成方法应用于 SVM 分类器时效果不理想。单纯的基于样本扰动的集成效果不如基于样本扰动和特征扰动的集成效果。而基于样本扰动和按一定准则选取特征子集的特征扰动要优于基于样本扰动和随机选取特征子集的特征扰动。实验结果证明了本文所提算法的有效性。

2. 基分类器数与分类精度之间的关系

第一部分只比较了固定基分类器数为 25 的情况下 3 种集成算法的分类精度。为了验证基分类器数对集成精度是否影响, 对 3 种集成算法在不同基分类器数下的分类精度进行比较, 在数据集 Ecoli, Sonar, Segment 和 Satimage 上进行了实验。实验结果如图 1 所示。

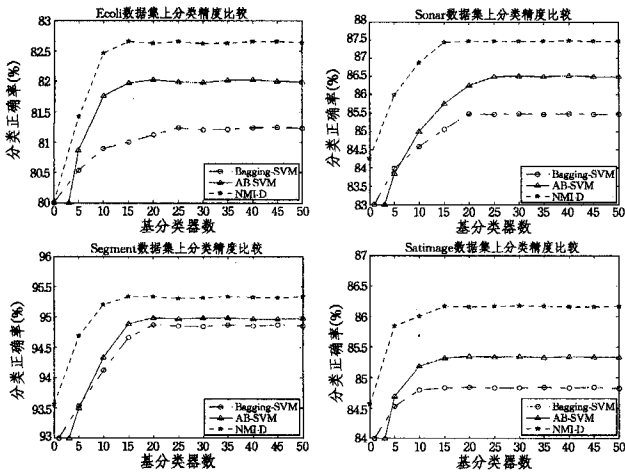


图 1 3 种算法在不同基分类器数下的分类精度比较

从图 1 可以看出, 在 4 个数据集上, NMI-D 算法在同等基分类器数下均取得了较 Bagging-SVM 算法和 AB-SVM 算法高的分类精度, Bagging-SVM 算法在基分类器数较少 (<5) 时分类精度较 AB-SVM 算法好, 而当基分类器数达到一定数目以后, AB-SVM 算法的分类精度明显较 Bagging-SVM 算法好。这是因为, AB-SVM 算法对特征空间的随机抽样可能导致重要特征的缺失, 因此其基分类器精度不高, 而随着基分类器的增加, 基分类器之间的差异性使得它们的错误形成互补, 从而提高了分类精度。而 Bagging-SVM 算法是基于样本随机抽样的, 特别是随着样本集的增大, 随机抽样对分类器的影响不大, 因此它的基分类器的差异性小, 使得集成效果不如 AB-SVM。再来分析 NMI-D 算法, 由于它是在 Bagging 方法形成的训练子集上再进行特征选择而得到最终的训练子集, 而且特征子集的选择基于分类器间的差异性, 因此, 它能得到比 Bagging-SVM 算法和 AB-SVM 算法差异性都高的基分类器, 而且由于 NMI-D 的特征子空间是通过特征选择算法得到的, 因此特征子集具有较好的分类能力, 保证了基分类器的分类精度, 所以可以获得更好的集成性能。

结束语 本文在分析特征的相关冗余度量标准和分类器间差异性的基础上, 提出了一种基于正则化互信息和差异度的集成特征选择方法。该算法利用正则化互信息作为特征相关性的度量准则, 运用 Bagging 方法中的 bootstrap 抽样技术结合基于正则化互信息的特征度量准则生成特征子集, 并引入基分类器差异度作为评价特征子集之间相关性大小的准则, 对所选的特征子集进行取舍从而得到差异性大的特征子集, 在其对应的特征空间上训练基分类器, 最后用加权投票法

融合各个基分类器的识别结果。仿真实验结果显示, 该方法可获得较高的分类精度。

参考文献

- [1] Opitz D. Feature selection for Ensembles [C]// Proceedings of American Association for Artificial Intelligence, 1999, 379-384
- [2] Ho T K. The random subspace method for constructing decision forests[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832-844
- [3] Brylla R, Osunab R G, Queka F. Attribute Bagging: Improving accuracy of classifier ensembles by using random feature subsets [J]. Pattern Recognition, 2003, 36(6): 1291-1302
- [4] Oliveira L S, Morita M, Sabourin R. Multi-Objective Genetic Algorithm Create Ensemble of Classifiers [C] // Proc. OFEMO 2005. Guanajuato, Mexico, 2005: 592-606
- [5] 李霞, 王连喜, 蒋盛益. 面向不平衡问题的集成特征选择[J]. 山东大学学报: 工学版, 2011, 41(3): 7-11
- [6] 孙亮, 韩崇昭, 沈建京, 等. 集成特征选择的广义粗集方法与多分类器融合[J]. 自动化学报, 2008, 34(3): 298-304
- [7] 张宏达, 王晓丹, 韩钧, 等. 分类器集成差异性研究[J]. 系统工程与电子技术, 2009, 31(12): 3007-3012
- [8] Dietterich T G. Ensemble methods in machine learning [C] // Proc. The 1st Int'l Workshop on Multiple Classifier Systems (MCS 2000). Italy, LNCS, Springer, 2000: 1-15
- [9] Kuncheva L I, Skurichina M, Duin R P W. An experimental study on diversity for bagging and boosting with linear classifiers [J]. Information Fusion, 2002, 3: 245-258
- [10] Dietterich T G. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization [J]. Machine Learning, 2000, 40: 139-158
- [11] Bonnländer B V, Weigend A S. Selecting input variables using mutual information and nonparametric density evaluation [C] // ISSN 94. 1994: 42-50
- [12] Kwak N, Choi C-H. Input feature selection by mutual information based on Parzen window [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(12): 1667-1671
- [13] Peng Han-chuan, Long Fu-hui, Ding C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226-1238
- [14] Estévez P A, Tesmer M, Perez C A, et al. Normalized Mutual Information Feature Selection [J]. IEEE Transactions on Neural Networks, 2009, 20(2): 189-201
- [15] Vinh L T, Lee S, Park Y-T, et al. A novel feature selection method based on normalized mutual information [J]. Springer Science Business Media, 2012, 37(1): 100-120
- [16] Hansen L K, Salamon P. Neural network ensembles [J]. IEEE Transactions on Pattern analysis and Machine Intelligence, 1990, 12(10): 993-1001
- [17] Valentini G, Dietterich T G. Bias-variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods [J]. Journal of Machine Learning Research, 2004, 5: 725-775
- [18] 王晓丹, 孙东延, 郑春颖. 一种基于 AdaBoost 的 SVM 分类器 [J]. 空军工程大学学报: 自然科学版, 2006, 7(6): 54-57
- [19] Breiman L. Bagging Predictors [J]. Machine Learning, 1996, 24(2): 123-140