

基于随机子空间的 最小最大模块化支持向量机^{*}

余 艺 吴家皋 李 云

(南京邮电大学 计算机学院 南京 210023)

摘 要 最小最大模块化支持向量机(M3-SVM)是一种对大规模数据进行模式分类的有效方法. 为进一步提高 M3-SVM 对高维大规模不平衡数据的分类性能,文中分析多种随机子空间策略,并将其与 M3-SVM 相结合,以实现降维和增加特征层面上的集成机制,从而得到一类基于随机子空间的最小最大模块化支持向量机(M3-SVM-RS). 在现实数据集上验证随机子空间策略的有效性,同时通过实验分析 M3-SVM-RS 中各个子模块(基分类器)之间的差异性.

关键词 随机子空间(RS),最小最大模块化支持向量机(M3-SVM),多样性
中图法分类号 TP 39

Integration of Random Subspace and Min-Max Modular SVM

YU Yi, WU Jiao-Gao, LI Yun

(School of Computer Science and Technology, Nanjing University of Posts and Telecommunications,
Nanjing 210023)

ABSTRACT

The min-max modular support vector machine (M3-SVM) is a powerful tool for dealing with large-scale data. To improve the classification performance of M3-SVM for unbalanced data with high dimension, several random subspace strategies are analyzed and combined with M3-SVM to reduce the dimensionality and add the ensemble mechanism on feature level. Thus, the min-max modular support vector machine with random subspace is proposed. The experimental results on real-world datasets including unbalanced data indicate that the proposed random subspace strategy enhances the classification of M3-SVM. Moreover, the diversity between sub-modules (base learner) in M3-SVM is discussed.

Key Words Random Subspace (RS), Min-Max Modular Support Vector Machine (M3-SVM), Diversity

^{*} 国家自然科学基金(No. 61073114)、江苏省自然科学基金(No. BK20131378)资助项目

收稿日期: 2013-05-13

作者简介 余艺,女,1989年生,硕士研究生,主要研究方向为智能信息处理. E-mail: yy2007915@126.com. 吴家皋,男,1969年生,博士,副教授,主要研究方向为大规模数据管理. 李云(通讯作者),男,1974年生,博士,教授,主要研究方向为机器学习、模式识别. E-mail: liyun@njupt.edu.cn.

1 引言

如何从大规模数据中挖掘出有用知识是目前的研究热点. 但对于大规模数据, 一些传统的简单模式分类技术, 如 KNN、多层感知机 (Multi-layer Perceptron, MLP) 及 SVM 等^[1], 已无法直接应用. 事实表明集成学习是处理大规模数据的一种有效手段^[2].

基于涌现理论和“部分对部分”的样本分解策略的最小最大模块化支持向量机 (Min-Max Modular SVM, M3-SVM)^[3-5] 是处理大规模数据时表现较优的集成学习方法, 目前已应用到专利分类、人脸识别、脑电信号处理、语音识别和蛋白质亚细胞定位等领域^[6-11].

对于大规模数据, M3-SVM 通常通过将原始问题划分成许多规模较小的子问题, 训练相应的子模块 (基分类器) 来解决这一系列子问题. 再通过一定的 MIN-MAX 集成策略对所有子模块的预测结果进行整合, 最终得到大规模问题的解. 然而, 若大规模数据同时具有较高的特征维数和大量的冗余特征、不相关特征, 将极大影响 M3-SVM 中子模块的性能, 从而降低集成学习的性能. 而且 M3-SVM 只是从样本分解的角度来集成学习, 没有考虑特征层面的集成模式. 因此本文将基于特征排序和随机子空间方法来有效处理“维数灾难”问题及通过增大特征层面的差异来提高 M3-SVM 的性能. 特征排序是特征选择的模式之一^[12-17], 通过设计评价准则, 获取所有特征的权重, 并对其进行排序. 而随机子空间 (Random Subspace, RS) 的主要思想是从原始特征空间中随机采样得到一系列特征子空间, 并基于所挑选的随机子空间构建基学习器, 通过随机森林实现集成学习^[18].

本文提出的基于随机子空间的 M3-SVM (Min-Max Modular SVM with RS, M3-SVM-RS) 是在对特征进行排序后, 采用不同的随机子空间策略来获取 M3-SVM 中不同子模块上的特征子空间, 再训练子模块上的基分类器, 最终利用 MIN-MAX 规则集成得到学习结果.

2 最小最大模块化支持向量机

令 $S = \{X_l, Y_l\}_{l=1}^L$ 表示一个 K ($K > 2$) 类分类问题的训练样本集合, $X_l \in \mathbf{R}^D$ 表示 D 维空间中第 l 个样本的特征向量, L 表示样本总数, 期望输出

$Y_l \in \{C_1, C_2, \dots, C_K\}$ 表示每个样本对应的类别标记. M3-SVM 的整个学习过程包括任务分解和 MIN-MAX 规则集成.

2.1 任务分解

文献 [3] 中 Lü 已对 K 类问题的任务分解进行分析, 可通过“一对一”的分解策略将多类问题划分为 $K(K-1)$ 个二类问题. 假设 K 类中每个类的训练样本为

$$X_i = \{X_l^{(i)}\}_{l=1}^{L_i}, i = 1, 2, \dots, K, \sum_{i=1}^K L_i = L,$$

其中 $X_l^{(i)}$ 是属于 C_i 类的第 l 个训练样本, L_i 指 C_i 类含有的训练样本总数. 每个二类问题都是独立的, 假设一个为正类 (标记为 1), 一个为负类 (标记为 -1), 并定义为

$$S_{ij} = \{X_l^{(i)}, 1\}_{l=1}^{L_i} \cup \{X_l^{(j)}, -1\}_{l=1}^{L_j}, \\ i = 1, 2, \dots, K; j = 1, 2, \dots, K; i \neq j.$$

如果二类问题仍是大规模的或不平衡的, 可继续采用“部分对部分”^[4] 的分解策略对其进行划分得到更小更均衡的二类子问题. 假设属于类别 C_i 的样本集被分解成 M_i 个子集, 表示为

$$X_{iu} = \{X_l^{(iu)}\}_{l=1}^{L_{iu}}, \\ i = 1, 2, \dots, K; u = 1, 2, \dots, M_i,$$

其中 L_{iu} 是数据子集 X_{iu} 所含样本的数目, 且有等式

$$X_i = X_{i1} \cup X_{i2} \cup \dots \cup X_{iM_i}$$

和

$$\sum_{u=1}^{M_i} L_{iu} = L_i$$

成立. 当然其他类别的训练样本也可采用同样方法进行划分, 如 X_j 可划分为 M_j 个子集. 那么原始的两类问题 (X_i 和 X_j) 可分解得到 $M_i \times M_j$ 个更小更平衡的二类子问题, 其训练样本集合为

$$S_{ij}^{uv} = \{ (X_l^{(iu)}, 1) \}_{l=1}^{L_{iu}} \cup \{ (X_l^{(jv)}, -1) \}_{l=1}^{L_{jv}}, \\ u = 1, 2, \dots, M_i; v = 1, 2, \dots, M_j; \\ i = 1, 2, \dots, K; j = 1, 2, \dots, K; i \neq j,$$

L_{jv} 是更小样本子集合 X_{jv} 所含负类训练样本的个数, 且

$$\sum_{v=1}^{M_j} L_{jv} = L_j.$$

于是, 一个 K 类问题就可划分为

$$\sum_{i=1}^K \sum_{j=1, j \neq i}^K M_i \times M_j$$

个二类子问题. 还可采用不同的划分方法, 如随机划分、超平面划分方法^[19]等, 将大规模的二类问题划分成一系列规模较小的二类子问题. 超平面划分方

法是根据训练样本到所选择的超平面之间的距离, 将样本集等分割成若干个训练样本子集合, 这是目前较常用的划分方法之一。

2.2 集成模块

任务分解的步骤完成后, 在每个二类样本子集上训练相应的子模块(SVM 基分类器), 然后用 MIN-MAX 规则进行集成。

每个子模块上得到的结果通过 MIN 规则和 MAX 规则^[3] 集成。假设一个二类(正类和负类)问题, 将其划分成一系列二类子问题。MIN 规则对于每个子模块的训练样本集, 如果它拥有相同正类训练样本和不同负类训练样本, 那么它的分类结果取最小值。类似地, MAX 规则对于每个子模块的训练样本集, 如果它拥有相同负类训练样本和不同正类训练样本, 那么它的分类结果取最大值。于是, 上述提到两类问题分解成 $M_i \times M_j$ 个二类平衡子问题结果的集成过程可描述为 M_i 个 MIN 单元和 1 个 MAX 单元的组合, 即

$$O_{ij}^u = \min(O_{ij}^{u,1}, O_{ij}^{u,2}, \dots, O_{ij}^{u,M_j}) = \min_{v=1}^{M_j} \{O_{ij}^{u,v}\},$$

$$O_{ij} = \max_{u=1}^{M_i} \{O_{ij}^u\},$$

其中 O_{ij}^u 是采用 MIN 规则后输出的最小值, $O_{ij}^{u,p}$ 是对二类子问题 $S_{ij}^{u,p}$ 进行学习所得基分类器的输出, O_{ij} 是二类问题 S_{ij} 学习所得集成结果。

2.3 最小最大模块化支持向量机结构

综上所述, 针对大规模两类问题, M3-SVM 首先进行大规模问题的平衡化分解, 得到一系列更小更平衡的子问题, 并训练相应的子模块(基分类器 SVM) 用于解决相应的子问题, 再利用 MIN-MAX 规则集成得到原始大规模问题的解。整个 M3-SVM 结构如图 1 所示。

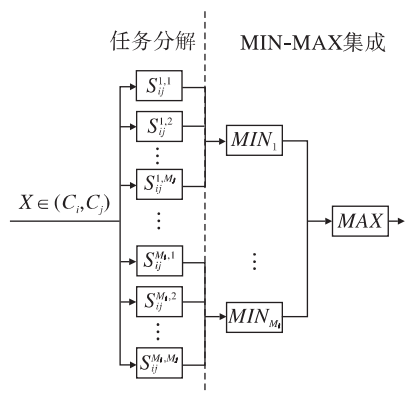


图 1 两类问题的 M3-SVM 结构图

Fig. 1 M3-SVM structure for two-class problem

3 基于随机子空间的最小最大模块化支持向量机

3.1 方法实现策略

为进一步提高 M3-SVM 集成学习的性能, 本文将对每个二类子问题所对应的训练样本进行降维。为保证子模块分类准确率, 且尽可能地增大子模块之间的差异性(多样性), 本文提出的降维方法包括两个阶段: 1) 特征排序, 消除大量不相关特征(排序靠后的特征); 2) 利用随机子空间策略为不同的子模块挑选不同的特征子集。而整个随机子空间策略又包括两部分: 1) 从排序靠前的特征中挑选部分重要特征; 2) 从剩下的特征中挑选部分特征, 两者组合到一起构成子模块上的特征子空间。因而不同子模块上的特征子空间都包括两个部分: 排序靠前的的重要特征和随机挑选的其他特征。这样通过保留一定数量的重要特征, 保证子模块的准确率。而通过随机挑选的特征又不尽相同, 这样就会增大子模块之间的差异性, 从而提高集成学习的性能。

根据特征排序和数据划分之间的顺序, 随机子空间的实现策略主要考虑如下两类。

1) 数据划分前进行特征排序。

(1) RS11: 采用经典的特征排序方法, 基于整个训练样本集对特征进行排序。而对于不平衡数据集, 主要通过对大类样本进行随机采样, 获得平衡的训练样本集再进行特征排序。特征排序好以后, 先消除不相关特征, 再对大规模训练数据进行划分, 划分后通过随机子空间策略获取每个子模块所对应的特征子空间。将特征排序后的前 m_0 个特征直接作为重要特征放入特征子空间中, 然后对不同的子模块, 从剩下特征中再随机抽取 m_1 个特征, 与前面抽取的重要特征共同组成 $m_0 + m_1$ 维的特征子空间。基于选择的特征子空间, 在每个子模块上训练得到相应的基分类器 SVM。

(2) RS12: 特征排序过程与 RS11 一样。对于不同的子模块, 其所对应的特征子空间的构成如下。从排在前半一半的所有特征中随机选择 m_0 个重要特征, 再从剩下的所有特征中随机挑选 m_1 个特征, 共同组成 $m_0 + m_1$ 维的特征子空间^[20]。然后基于选择的特征子空间, 在每个子模块上训练得到相应的基分类器 SVM。

2) 数据划分后进行特征排序。

RS21 和 RS22 随机子空间策略是先通过一定的划分方法对大规模数据进行划分后, 再在每个子模块上利用划分后的训练样本子集对特征进行排序,

然后再选取子模块上的特征子空间. 至于后续的特征子空间的构成方式, RS21 与 RS11 一样, 而 RS22 与 RS12 一样.

3.2 随机子空间的差异性分析

定义 1 差异性 假设有 2 个随机子空间 I_1 和 I_2 , 那么它们之间的差异性可定义为

$$DRS(I_1, I_2) = |I_1 \cup I_2| - |I_1 \cap I_2|,$$

其中, $|\cdot|$ 表示特征子空间的维数, 等式右边第一项表示两个随机子空间进行组合后的特征维数, 第二项为两个随机子空间中相同特征的个数.

如果特征 f 只是出现在其中一个随机子空间, 则 DRS 的值加 1. 这样一对随机子空间在所有特征上的平均差异性为

$$E(DRS) = \sum_{q=1}^Q P(f_q \in I_1) P(f_q \notin I_2) + P(f_q \notin I_1) P(f_q \in I_2),$$

其中, Q 为特征排序后, 通过剔除不相关特征, 剩下的特征集维数; $P(\cdot)$ 表示概率值.

对于随机子空间策略 RS11, 所有随机子空间的前 m_0 个重要特征都相同. 虽然这些特征对子模块的分类性能有用, 但对随机子空间的差异性无用. 而剩下的所有特征(个数为 $Q - m_0$) 中被挑选的概率为 $m_1/(Q - m_0)$, 且随机子空间是独立选取的, 因此平均差异性为

$$\begin{aligned} E(DRS) &= 2(Q - m_0) \frac{m_1}{Q - m_0} \left(1 - \frac{m_1}{Q - m_0}\right) \\ &= 2m_1 \left(\frac{Q - m_0 - m_1}{Q - m_0}\right). \end{aligned}$$

而对于策略 RS12, 前 m_0 个重要特征被挑选的概率为 $2m_0/Q$, 后 m_1 个特征被挑选的概率为 $m_1/(Q - m_0)$, 因此平均差异性为

$$\begin{aligned} E(DRS) &= 2 \left[\left(\frac{1}{2}Q\right) \frac{m_0}{\frac{1}{2}Q} \left(1 - \frac{m_0}{\frac{1}{2}Q}\right) + \right. \\ &\quad \left. (Q - m_0) \frac{m_1}{Q - m_0} \left(1 - \frac{m_1}{Q - m_0}\right) \right] \\ &= 2 \left(\frac{m_0(Q - 2m_0)}{Q} + \frac{m_1(Q - m_1 - m_0)}{Q - m_0} \right). \end{aligned}$$

策略 RS22 和 RS21 是在数据划分后基于每个子模块上的样本子集进行特征排序. 由于每个子模块上的训练样本子集不相同, 因此特征排序结果也会差异较大. 则随机挑选的特征子空间的差异性就难以从概率上进行分析. 但排序结果的差异性会增大

随机子空间的差异性. 因此就差异性而言, 可认为利用 RS22 和 RS21 所挑选的随机子空间之间的差异性大于利用 RS12 和 RS11 所挑选的随机子空间之间的差异性.

4 实验及结果分析

为验证 M3-SVM-RS 的性能, 本文在 3 个现实数据集上^[21]进行实验. 数据集的描述如表 1 所示. 表 1 中 Web Page 数据集和 IJCNN 数据集是不平衡数据集. 实验中直接使用数据集本身所给出的训练样本集和测试样本集.

表 1 实验数据集的基本信息

Table 1 Essential information of experimental datasets

数据集	训练样本	测试样本	特征数	类别
IJCNN	33040	9999	22	2
Adult	32562	16281	123	2
Web Page	49749	14951	300	2

4.1 评价准则

为评价相关方法性能, 主要采用测试准确率. 但对于不平衡数据而言, 仅采用准确率来衡量分类性能是不全面的, 因为数据的不平衡性会导致分类倾向降低少数类的分类性能. 因此, 除常用到的准确率之外, 实验中还采用 F -measure 来衡量 M3-SVM 在不平衡数据集上的分类性能. F -measure 综合考察召回率和准确率, 能反映不平衡数据中少数类的分类性能^[22].

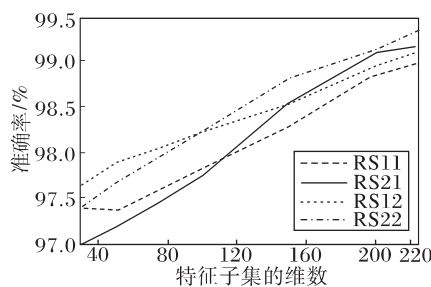
一般情况下, 基学习器的准确率越高, 而基学习器之间的差异性越大, 则集成效果越好. 因此在实验中还评价各种方法中基分类器之间的差异性. 所采用的差异性度量准则为一致性错误差异 (Coincident Failure Diversity, CFD)^[23-24]. CFD 是非对称的差异性度量^[25]方法, 它主要考察所有分类器对同个样例的分类结果之间的差异性. 该度量方法的特点: 对于同个测试样例而言, 当所有的分类器同时分类正确或同时分类错误时, 基分类器之间的差异性最小, 即 $CFD = 0$. 当对于任意一个随机选择的样例, 只有一个基分类器分错时, 基分类器之间的差异性最大, 即 $CFD = 1$. CFD 取值越大, 分类器之间的差异性越大, 反之亦然.

4.2 实验结果分析

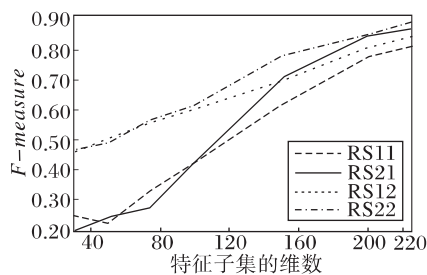
首先, 将 3.1 节所介绍的各种随机子空间策略与 M3-SVM 相结合, 在 Web Page 不平衡数据集上完

成相关实验. 利用超平面划分方法将 Web Page 数据集的多数类划分成 90 个样本子集, 少数类划分成 2 个样本子集, 得到均衡的两类子问题. 特征排序采用的是经典算法支持向量机 - 递归特征消除 (SVM-Recursive Feature Elimination, SVM-RFE) [26], 其用作特征排序时, 不需要算法中的迭代删除过程. 在随机子空间策略中, 第一步挑选的重要特征数 $m_0 = 10$. SVM 中选取 RBF 核函数 ($t = 2$), 设置 $C = 1$.

实验结果如图 2 所示. 从图 2 可看出, 采取随机子空间策略 RS22 与 M3-SVM 相结合, 所得到的无论是准确率还是 F -measure 都较优. 因此后文实验将选用 RS22 作为随机子空间的选取策略.



(a) 准确率
(a) Accuracy rates



(b) F -measure 值
(b) F -measure values

图 2 不同随机子空间策略下的准确率及 F -measure 值

Fig. 2 Accuracy rates and F -measure values with different RS strategies

再采用特征排序方法 SVM-RFE 和所选定的随机子空间策略 RS22 与 M3-SVM 网络相结合, 分别在 Adult 数据集和 IJCNN 数据集上进行一系列实验. 考虑两种数据划分方法: 随机划分 (Random Division, RD) 与超平面划分 (Hyperplane Partition, HP). 将特征选择方法 SVM-RFE 或随机子空间策略 RS22 与不同的数据划分方法进行组合, 可得到不同的集成学习模式:

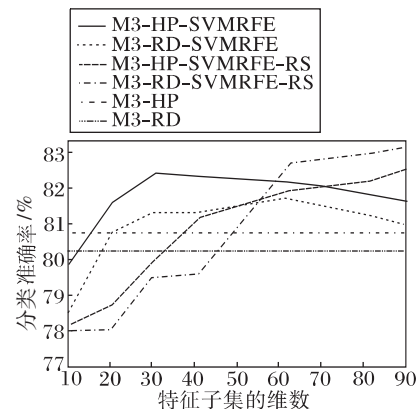
1) 利用两种数据划分方法对数据集进行分解

后, 在每个子模块上利用 SVM-RFE 进行特征选择, 然后直接与 M3-SVM 相结合, 相应的算法名称为 M3-RD-SVMRFE 和 M3-HP-SVMRFE [27];

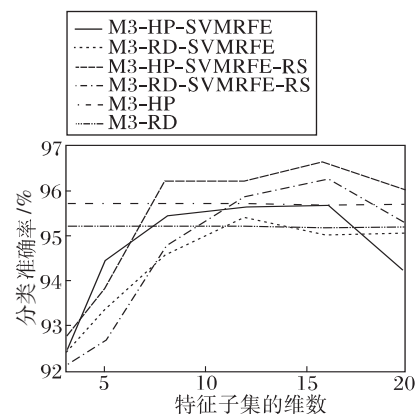
2) 利用两种数据划分方法进行数据分解后, 将本文提出的随机子空间策略 RS22 (采用的特征排序方法为 SVM-RFE) 与 M3-SVM 相结合, 相应的算法名称为 M3-HP-SVMRFE-RS 和 M3-RD-SVMRFE-RS.

此外还考虑原始的基于两种数据划分方法的 M3-SVM, 即 M3-RD 和 M3-HP.

相关实验结果如图 3 和图 4 所示.



(a) Adult



(b) IJCNN

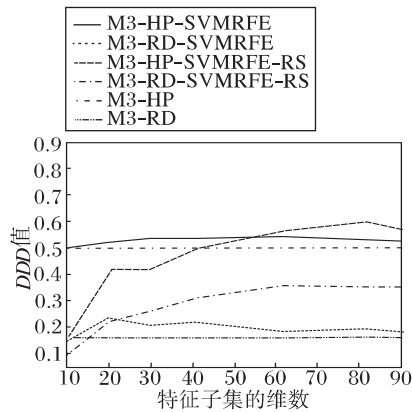
图 3 2 个数据集上不同方法的分类准确率

Fig. 3 Classification accuracy rates of different methods on 2 datasets

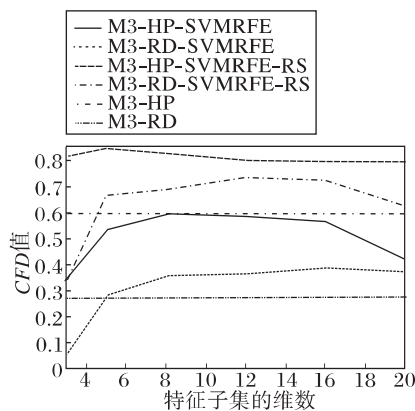
实验中将 Adult 数据集类别为“1”的样本集合划分成 22 个子样本集, 类别为“-1”的样本集合划分成 9 个子样本集; 将 IJCNN 数据集多数类划分成 30 个子样本集, 少数类划分成 5 个子样本集, 以实现两类子问题中训练样本的均衡化.

此外, SVM 的参数设定分别如下: 在 Adult 数据集上采用线性核函数和 IJCNN 数据集上选取 RBF 核函数($t=2$), $C=1$. 对于 Adult 数据集, 在利用随机子空间策略的算法(M3-HP-SVMRFE-RS 和 M3-RD-SVMRFE-RS)中, 第一步所挑选的重要特征数为 $m_0=10$; IJCNN 数据集上, M3-HP-SVMRFE-RS 和 M3-RD-SVMRFE-RS 中第一步所挑选的重要特征数为 $m_0=2$.

上述所有算法在两个数据集上的准确率如图 3 所示. 从图 3 中可看出, 相比不加特征选择的 M3-SVM(M3-RD 和 M3-HP) 和只加入特征选择方法的 M3-SVM(M3-RD-SVMRFE 和 M3-HP-SVMRFE) 而言, 基于随机子空间的 M3-SVM(M3-HP-SVMRFE-RS 和 M3-RD-SVMRFE-RS) 通常随着选择的特征维数增加, 其分类性能也稳步提高, 最终将超越其他方法.



(a) Adult



(b) IJCNN

图 4 2 个数据集上不同方法的一致性错误差异度量

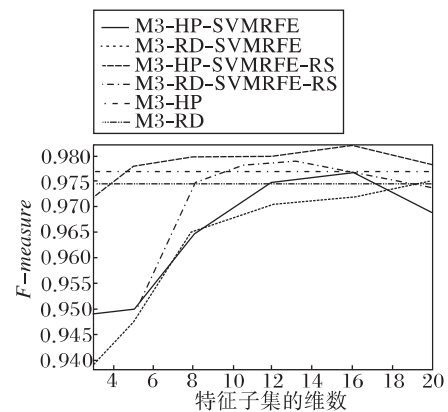
Fig. 4 CFD values of different methods on 2 datasets

图 4 为在 Adult 数据集和 IJCNN 数据集上差异性度量的实验结果. 从图 4 可看出, 基于随机子空间

的 M3-SVM 的 CFD 值通常要高于其他采用相同数据划分方法的 M3-SVM. 这说明基于随机子空间的 M3-SVM 中子模块(基分类器)之间有更大的差异性, 这也从一定程度上说明基于随机子空间的 M3-SVM 可获得更好的分类性能.

结合图 3 和图 4 来看, 在相同数据集上, 所有算法的准确率变化趋势跟差异性的变化趋势是类似的, 即基分类器之间差异性增大, 则集成学习的准确率增大; 而基分类器之间的差异性减小, 则集成的准确率也减少. 这也从一定程度上表明基分类器之间的差异性确实对集成学习的性能有正面影响.

针对不平衡数据, 需要采用 F -measure 来度量分类的综合性能. 在 IJCNN 数据集上实验结果如图 5 所示. 在不平衡数据集上, 本文方法(M3-HP-SVMRFE-RS 和 M3-RD-SVMRFE-RS) 通常能得到更好的 F -measure.

图 5 不平衡数据集 IJCNN 上不同方法的 F -measure 值Fig. 5 F -measure values of different methods on unbalanced dataset IJCNN

根据图 2、图 3(b) 和图 5 可看到, 每个算法的准确率曲线跟 F -measure 曲线的走势较相似, 这说明算法的准确率不是以牺牲少数类为代价来获取的. 而这些算法都是基于 M3-SVM, 因此通过对样本集进行均衡化分解, M3-SVM 可有效解决不平衡问题.

5 结束语

为提高最小最大模块化支持向量机 M3-SVM 对大规模数据的分类性能, 提出一类基于随机子空间的 M3-SVM(M3-SVM-RS). 利用特征排序和随机子空间, 设计多种相关策略, 并简要分析所挑选随机子

空间的差异性. 在现实数据集上的实验表明, 融入随机子空间的 M3-SVM 通常能获得较好的分类性能, 且对于不平衡数据集, 可获得较好的 F -Measure. 实验还表明采用随机子空间可增大 M3-SVM 中子模块 (基分类器) 之间的差异性, 从而一定程度上阐述融入随机子空间的 M3-SVM 性能较好的内在原因. 值得说明的是, 本文方法是基于随机子空间, 所以同样也面临着如何选取合适子空间大小的问题, 这是未来需继续研究和解决的一个重要问题.

参 考 文 献

- [1] Han J, Kamber M. Data Mining: Concepts and Techniques. San Francisco, USA: Morgan Kaufman, 2000
- [2] Dietterich T. Ensemble Methods in Machine Learning // Proc of the 1st International Workshop on Multiple Classifier Systems. Cagliari, Italy, 2000: 1 - 15
- [3] Lü B L, Ito M. Task Decomposition and Module Combination Based on Class Relations: A Modular Neural Network for Pattern Classification. IEEE Trans on Neural Networks, 1999, 10(5): 1244 - 1256
- [4] Lü B L, Wang K A, Utiyama M, et al. A Part-versus-Part Method for Massively Parallel Training of Support Vector Machines // Proc of the International Joint Conference on Neural Networks. Budapest, Hungary, 2004: 735 - 740
- [5] Lü B L, Wang X L. A Parallel and Modular Pattern Classification Framework for Large-Scale Problems // Chen C H, ed. Handbook of Pattern Recognition and Computer Vision. 4th Edition. New Jersey, USA: World Scientific Publishing, 2009: 725 - 746
- [6] Lü B L, Ma Q, Ichikawa M, et al. Efficient Part-of-Speech Tagging with a Min-Max Modular Neural Network. Applied Intelligence, 2003, 19(1/2): 65 - 81
- [7] Lü B L, Shin J, Ichikawa M. Massively Parallel Classification of Single-Trial EEG Signals Using a Min-Max Modular Neural Network. IEEE Trans on Biomedical Engineering, 2004, 51(3): 551 - 558
- [8] Chu X L, Ma C, Li J, et al. Large-Scale Patent Classification with Min-Max Modular Support Vector Machines // Proc of the International Joint Conference on Neural Networks. Hong Kong, China, 2008, 1: 3972 - 3979
- [9] Wu K, Lü B L, Uchiyama M, et al. An Empirical Comparison of Min-Max-Modular k -NN with Different Voting Methods to Large-Scale Text Categorization. Soft Computing, 2008, 12(7): 647 - 655
- [10] Chen Y Y, Lü B L, Zhao H. Parallel Learning of Large-Scale Multi-label Classification Problems with Min-Max Modular LIBLINEAR // Proc of the International Conference on Neural Networks. Brisbane, Australia, 2012, 1: 1 - 7
- [11] Yang Y, Lü B L. Protein Subcellular Multi-localization Prediction Using a Min-Max Modular Support Vector Machine. International Journal of Neural Systems, 2010, 20(1): 13 - 28
- [12] Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 2003, 3(9): 1157 - 1182
- [13] Liu H, Yu L. Toward Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Trans on Knowledge and Data Engineering, 2005, 17(4): 494 - 502
- [14] Sun Z H, Bebis G, Miller R. Object Detection Using Feature Subset Selection. Pattern Recognition, 2004, 37(11): 2165 - 2176
- [15] Jain A K, Duin R P W, Mao J C. Statistical Pattern Recognition: A Review. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4 - 37
- [16] Kudo M, Sklansky J. Comparison of Algorithms That Select Features for Pattern Classifiers. Pattern Recognition, 2000, 33(1): 25 - 41
- [17] Zhao Z. Spectral Feature Selection for Mining Ultrahigh Dimensional Data. Ph. D Dissertation. Arizona, USA: Arizona State University, 2010
- [18] Ho T K. The Random Subspace Method for Constructing Decision Forests. IEEE Trans on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832 - 844
- [19] Wang K L, Zhao H, Lü B L. Task Decomposition Using Geometric Relation for Min-Max Modular SVMs // Proc of the International Symposium on Neural Network. Chongqing, China, 2005: 887 - 892
- [20] Yang M, Bao J, Ji G L. Semi-Random Subspace Sampling for Classification // Proc of the 6th International Conference on Natural Computation. Yantai, China, 2010: 3420 - 3424
- [21] Platt J C. Fast Training of Support Vector Machines Using Sequential Minimal Optimization // Schölkopf B, Burges C J C, Smola A J, eds. Advances in Kernel Methods-Support Vector Learning: Support Vector Learning. Cambridge, USA: MIT Press, 1998: 41 - 65
- [22] Ye Z F, Wen Y M, Lü B L. A Survey of Imbalanced Pattern Classification Problems. CAAI Trans on Intelligent Systems, 2009, 4(2): 148 - 156 (in Chinese)
(叶志飞, 文益民, 吕宝粮. 不平衡分类问题研究综述. 智能系统学报, 2009, 4(2): 148 - 156)
- [23] Kuncheva L I, Whitaker C J. Measure of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. Machine Learning, 2003, 51(2): 181 - 207
- [24] Dietterich T G. Machine Learning Research: Four Current Directions. Artificial Intelligence Magazine, 1997, 18(4): 97 - 136
- [25] Tang E K, Suganthan P N, Yao X. An Analysis of Diversity Measures. Machine Learning, 2006, 65(1): 247 - 271
- [26] Guyon I, Wonten J, Barnhill S. Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning, 2002, 46(1/2/3): 389 - 422
- [27] Li Y, Feng L L. Integrating Feature Selection and Min-Max Modular SVM for Powerful Ensemble [EB/OL]. [2013 - 04 - 10]. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06252481>