

选择性集成学习算法综述

张春霞 张讲社

(西安交通大学理学院 西安 710049)

(西安交通大学机械制造系统工程国家重点实验室 西安 710049)

摘 要 集成学习因其能显著提高一个学习系统的泛化能力而得到了机器学习界的广泛关注,但随着基学习机数目的增多,集成学习机的预测速度明显下降,其所需的存储空间也迅速增加. 选择性集成学习的主要目的是进一步改善集成学习机的预测效果,提高集成学习机的预测速度,并降低其存储需求. 该文对现有的选择性集成学习算法进行了详细综述,按照算法采用的选择策略对其进行了分类,并分析了各种算法的主要特点,最后对选择性集成学习在将来的可能研究方向进行了探讨.

关键词 选择性集成学习;基学习机;集成学习机;多样性;泛化能力

中图法分类号 TP181 DOI号: 10.3724/SP.J.1016.2011.01399

A Survey of Selective Ensemble Learning Algorithms

ZHANG Chun-Xia ZHANG Jiang-She

(Faculty of Science, Xi'an Jiaotong University, Xi'an 710049)

(State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049)

Abstract In recent years, ensemble learning has received lots of attention in machine learning due to its potential to significantly improve the generalization capability of a learning system. With increasing number of ensemble members, however, the prediction speed of an ensemble machine decreases significantly and its storage need increases quickly. The aim of selective ensemble learning is to further improve the prediction accuracy of an ensemble machine, to enhance its prediction speed as well as to decrease its storage need. This paper presents a detailed review of the current selective ensemble learning algorithms and categorizes them into different classes according to their utilized selection strategy. Meanwhile, the main characteristics of each representative algorithm are studied. Finally, the future research directions of selective ensemble learning are discussed.

Keywords selective ensemble learning; base learner; ensemble machine; diversity; generalization capability

1 引 言

模式分类与回归问题是机器学习和模式识别等研究领域中最常遇到的最基本任务之一,该任务的主要目标是利用实际数据构建一个具有较强泛化能力

的预测模型(学习机),使我们可以对新的未知对象给出尽可能精确的估计. 因此,人们关注的一个根本问题是如何提高学习系统的泛化能力. 由于集成学习(ensemble learning)在该方面的显著优势,近年来对集成学习理论和算法的研究一直是机器学习领域的一个热点,国际机器学习界的权威学者

收稿日期:2010-07-27;最终修改稿收到日期:2011-07-16. 本课题得到国家自然科学基金(61075006)、教育部高等学校博士学科点专项科研基金(20100201120048)和中央高校基本科研业务费专项基金资助. 张春霞,女,1980年生,博士,讲师,主要研究方向为集成学习、模式识别、Bootstrap 随机抽样方法等. E-mail: cxzhang@mail. xjtu. edu. cn. 张讲社,男,1962年生,博士,教授,主要研究领域为统计计算、模式识别、全局优化等.

Dietterich^[1]曾在《AI Magazine》杂志上将集成学习列为机器学习领域的四大研究方向之首。

集成学习使用多个学习机来解决同一问题,它通过调用一些简单的分类算法,以获得多个不同的基学习机,然后采用某种方式将这些学习机组合成一个集成学习机。随着集成学习研究队伍的不断壮大,集成学习技术得到了快速发展,各种集成学习算法目前也正在被广泛应用于生物、工程、医学、计算机视觉和图像处理等研究领域^[2-4]。一般地,一个集成学习机的构建分为两步:基学习机的生成和基学习机的合并,现有的许多集成学习算法主要是在这两方面存在差异。在构建集成学习机时,有效地产生泛化能力强、差异大的基学习机是关键,即基学习机的准确性和它们之间的多样性是两个重要因素。

目前,常见的用于生成基学习机的方法可以粗略地分为两大类:一类是将不同类型的学习算法应用于同一数据集上,这种方法得到的基学习机通常被称为是异质类型的(heterogeneous);另一类是将同一学习算法应用于不同的训练集(可基于原有的训练数据集进行随机抽样等方法得到),这种方法得到的基学习机被称为是同质类型的(homogeneous)。对于生成同质类型基学习机的方法,基于它们获取不同训练集所采用的技术,又可以分为对训练集重抽样(如 Bagging^[5]、Boosting^[6])、操纵输入变量(如随机子空间方法^[7]、旋转森林^[8])、操纵输出目标(如误差校正输出编码集成方法(Error correcting output code ensemble)^[9])、注入随机性(如 Random Forest^[10])。

在生成多个基学习机之后,一个很自然的问题是以何种方式将它们进行合并才能得到具有最强泛化能力的集成学习机?对此,研究者们也提出了很多解决办法,徐雷等人^[11]针对模式分类问题,根据基分类器提供的信息水平将现有的合并准则分成了三大类:抽象水平、秩水平和置信值水平。抽象水平是指基分类器的输出是类标签,秩水平假定基分类器的输出是根据分类器的预测效果好坏而对它们赋予的秩序列,而置信值水平则假定每个基分类器的输出是一个概率分布。在置信值水平类的合并准则中,根据是否需要估计额外的参数,它们又可以分为固定的合并准则(fixed rules)和可训练的合并准则(trainable rules)。固定的合并准则是基于基分类器的输出对基分类器直接进行合并,常用的有最大值、最小值、中位数、乘积、均值和多数投票(加权或不加权)准则;而可训练的合并准则是将多个基分类器

的输出作为新的特征再构建一个更高水平的分类器,常见的方法有神经网络、决策树、支持向量机、Bayes 准则、行为知识空间、Dempster-Shafer 理论等。值得注意的是,可训练的合并准则需要额外的数据集估计其中的参数,但在实际问题中,一般只有一个数据集可用,因此该数据集必须被用来同时训练基分类器和合并准则。在这种情况下,有三种方法可以达到上述目的:重复使用策略(reusing strategy)、随机划分策略(validation strategy)和层叠泛化策略(stacked generalization),这三种策略的详细介绍和各自的优缺点可参考文献^[12-13]。

在集成学习的研究初期,大多数方法都是先生成多个基学习机,然后将它们全部用于构建集成学习机。尽管采用这些方法得到的集成学习机的预测效果显著优于单个基学习机,但它们存在一些缺点:与基学习机相比,其预测速度明显下降,且随着基学习机数目的增多,它们所需的存储空间也急剧增多,这对于在线学习更是一个严重问题。因此,人们开始考虑:使用少量的基学习机是否可以达到更好的性能?

2002 年,周志华等人^[14]首先提出了“选择性集成”的概念,肯定地回答了上述问题,并在国内外集成学习界引起了强烈反响。理论分析和试验研究表明,从已有的基学习机中将作用不大和性能不好的基学习机剔除,只挑选一些基学习机用于构建集成则可以得到更好的预测效果。随后,选择性集成学习引起了研究者的关注并提出了一些有效的算法。这里,值得强调的是,选择性集成学习是在假定已生成多个基学习机的基础上,基于某种选择策略只从中选择一部分用于构建最终的集成。换句话说,选择性集成在选择的过程中,不会再生成新的基学习机,这与用训练数据生成一个 Boosting 集成的过程中,直接抛弃精度较差的基学习机是完全不同的。为了给选择性集成学习的初学者提供入门指导,并为新算法的设计提供参考,本文将对现有的选择性集成学习算法进行较为全面的综述,阐述近年来对选择性集成学习进行研究的主要内容和特点,通过分析现有算法的特征,将其分类,并对它们在未来的研究方向进行探讨。据我们所知,尽管文献^[15]曾对选择性集成学习算法进行了简短的综述,但该文未对一些代表性的选择性集成学习方法进行详细说明,且该文献出现的时间较早,而近两年来在集成学习领域中已经又涌现出了很多有效的选择性集成学习算法。因此,本文的研究是文献^[15]工作的进一步完善

和补充。

本文第 2 节对用于解决分类问题的选择性集成学习算法进行详细综述,并分析几类算法的特点;第 3 节讨论选择性集成学习算法在回归预测中的应用;第 4 节是总结与展望,主要探讨选择性集成学习在未来可能的研究方向。

2 选择性集成学习算法

由于分类问题和回归问题具有不同的特点,需要使用不同的方法进行解决,而集成学习最早是在解决分类问题的过程中提出的.因此,我们在本节中先对现有的关于分类问题的选择性集成学习算法进行讨论,下一节中再探讨基于回归问题的相关算法.在此,我们引入一些记号以方便后文的叙述.

假定 $\mathcal{L}_{tr} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_{tr}}$ 为给定的训练集,对每个训练个体 (\mathbf{x}_i, y_i) ,其输入变量 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \in R^p$,输出变量 $y_i \in \Phi = \{\omega_1, \omega_2, \dots, \omega_c\}$, c 为类的个数.同时,令 \mathcal{L}_{val} 、 \mathcal{L}_{ts} 分别表示容量为 N_{val} 和 N_{ts} 的验证集和检验集.在实际应用中,如果只给定了一个数据集,则可以通过随机划分的方法得到相应的训练集、验证集和检验集.此外,我们用 $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$ 表示基于 \mathcal{L}_{tr} 训练的基分类器集合, $\mathcal{S} = \{C_1^*, C_2^*, \dots, C_S^*\}$ 表示基于某种选择性集成学习算法从 \mathcal{T} 中选取出来的基分类器集合.

选择性集成学习的主要思想是基于某种衡量准则,从已有的基分类器中选择一些用于构建集成分类器,以加快分类器的预测速度、降低其存储空间需求并进一步提高分类精度.理论上,最优的基分类器子集 \mathcal{S} 可以通过枚举法(exhaustive enumeration)得到.对 \mathcal{T} 的每个子集 $\tilde{\mathcal{T}}$,衡量基于它构建的集成分类器的预测效果,选择具有最小泛化误差的集成分类器所对应的子集即可,即 $\mathcal{S} = \text{Argmin}_{\tilde{\mathcal{T}} \subset \mathcal{T}} \epsilon(\tilde{\mathcal{T}})$.然而,我们需要对 $\sum_{i=1}^T \binom{T}{i}$ 个子集进行计算,当基分类器的数目 T 较大时,其计算量极大.因而,枚举法在实际中并不可行.

鉴于上述原因,选择性集成学习在近 10 年中得到了广泛的研究,我们在下面的图 1 中给出了选择性集成学习算法的基本框架.

对于现有的选择性集成学习算法,它们主要在评测方法 \mathcal{M} 的选择上存在差异.于是,这些算法大致可以分为以下几类:聚类、排序、选择、优化和其它方法.下面我们对这几种方法的主要思想和特点进行详细分析.

```

输入: 训练集  $\mathcal{L}_{tr}$ , 验证集  $\mathcal{L}_{val}$ , 基分类算法  $\mathcal{C}$ , 基分类器个数  $T$ , 选择的基分类器个数  $S$ , 评测方法  $\mathcal{M}$ 
输出: 选择的基分类器集合  $\mathcal{S} = \{C_1^*, C_2^*, \dots, C_S^*\}$ 
训练过程:
初始化: 令基分类器集合  $\mathcal{T} = \emptyset$ .
For  $t = 1, 2, \dots, T$ 
    基于训练集  $\mathcal{L}_{tr}$ , 采用某种技术(如 Bootstrap 随机抽样方法)获取新的训练集  $\mathcal{L}_{tr}^{(t)}$ .
    应用基分类算法  $\mathcal{C}$  于  $\mathcal{L}_{tr}^{(t)}$  训练得到基分类器  $C_t$ , 将其加入集合  $\mathcal{T}$ .
EndFor (得到初始的基分类器集合  $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$ )
选择过程:
在验证集  $\mathcal{L}_{val}$  上对每个基分类器  $C_t (t = 1, 2, \dots, T)$  进行测试, 输出  $O_t$ .
利用评测方法  $\mathcal{M}$  基于  $O_t (t = 1, 2, \dots, T)$  对  $\mathcal{T}$  中每个元素进行评测, 并从中选择出  $S$  个基分类器  $C_1^*, C_2^*, \dots, C_S^*$ .

```

图 1 选择性集成学习算法的基本框架

2.1 基于聚类的方法

这类方法^[16-21]的主要过程如图 2 所示.

```

输入: 验证集  $\mathcal{L}_{val}$ , 基分类器集合  $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$ , 聚类算法  $\mathcal{A}$ 
输出: 选择的基分类器集合  $\mathcal{S} = \{C_1^*, C_2^*, \dots, C_S^*\}$ 
选择过程:
利用每个基分类器  $C_t (t = 1, 2, \dots, T)$  对验证集  $\mathcal{L}_{val}$  中的个体进行预测, 得到矩阵  $\mathbf{O} = (o_{ij})_{T \times N_{val}}$ , 其元素  $o_{ij}$  对应于第  $i$  个基分类器  $C_i$  对第  $j$  个个体  $\mathbf{x}_j$  的预测结果, 即  $o_{ij} = C_i(\mathbf{x}_j) (i = 1, 2, \dots, T; j = 1, 2, \dots, N_{val})$ .
视矩阵  $\mathbf{O}$  的每行  $\mathbf{o}_i = (o_{i1}, o_{i2}, \dots, o_{iN_{val}})$  为新的特征空间中的一个观测,  $\mathbf{O}$  为包含  $T$  个观测的数据集, 将聚类算法  $\mathcal{A}$  应用于  $\mathbf{O}$ , 找到具有类似预测结果的基分类器子集  $C_1, C_2, \dots, C_S$  (即聚类算法最终形成的类).
对每个基分类器子集进行修剪, 选择出具有代表性的基分类器  $C_1^*, C_2^*, \dots, C_S^*$ .

```

图 2 基于聚类的选择性集成学习算法框架

在上述过程中,有 3 个关键问题需要解决:如何衡量两个基分类器(子集)预测结果(如 \mathbf{o}_i 和 $\mathbf{o}_j, i \neq j$)之间的相似性?聚类算法 \mathcal{A} 选用何种算法?如何确定基分类器子集的个数 S ?

在文献[16]中,Giacinto 和 Roli 定义两个基分类器 C_s 和 C_t 之间的距离为

$$d(C_s, C_t) = 1 - \text{Prob}(C_s \text{ fails}, C_t \text{ fails}), \quad \forall C_s, C_t \in \mathcal{T} \quad (1)$$

其中, $\text{Prob}(C_s \text{ fails}, C_t \text{ fails})$ 为基于验证集 \mathcal{L}_{val} 估计的两个基分类器同时分类错误的概率.同时,他们定义两个基分类器子集 \mathcal{C}_i 和 \mathcal{C}_j 之间的距离为

$$d(\mathcal{C}_i, \mathcal{C}_j) = \max_{C_s \in \mathcal{C}_i, C_t \in \mathcal{C}_j} \{d(C_s, C_t)\}, \quad \forall \mathcal{C}_i \cap \mathcal{C}_j = \emptyset \quad (2)$$

然后,采用层次凝聚的聚类算法(hierarchical agglomerative clustering)来找到具有相似预测结果的基分类器子集.至于 S 的确定,则是在聚类算法的每一步,从每个类中挑选出到其它类的平均距离最大的基分类器,采用简单投票的方式将其组合成集成分类器,并基于验证集衡量其分类精度.最后,

选择具有最高精度的集成分类器对应的基分类器子集,即 $S = \{C_1^*, C_2^*, \dots, C_S^*\}$.

Lazarevic 和 Obradovic^[17] 则基于 O 中观测的欧氏距离,应用 K -均值聚类算法将基分类器进行分组,并利用基分类器子集的准确性和多样性对它们进行修剪.此外,文献[18-21]采用其它一些聚类技术如基于确定性退火(deterministic annealing)的软聚类算法、谱系聚类法、核聚类算法等对多分类器的选择问题进行了研究.

2.2 基于排序的方法

通过对基分类器进行排序来达到修剪集成分类器的目的^[22-29]是比较直观的选择性集成学习方法,它们的步骤大致可以分为两步:基于某种衡量标准(如准确性)对基分类器排序,采用合适的停止准则(如事先指定选取的基分类器个数)选取一定数量的基分类器.在下面的图 3 中,我们列出了这类算法的主要步骤.

```

输入:验证集  $\mathcal{L}_{val}$ , 评测标准  $\mathcal{M}$ (如准确性、预测误差等), 基分类器集合  $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$ 
输出:选择的基分类器集合  $S = \{C_1^*, C_2^*, \dots, C_S^*\}$ 
排序过程:
For  $t = 1, 2, \dots, T$ 
    在验证集  $\mathcal{L}_{val}$  上, 依据评测标准  $\mathcal{M}$  对每个基分类器  $C_i$  ( $i = 1, 2, \dots, T$ ) 进行评价并排序.
    根据事先指定的基分类器个数  $S$  或评测方法自身确定的  $S$ , 从  $\mathcal{T}$  选出性能较好的基分类器  $C_1^*, C_2^*, \dots, C_S^*$ .
EndFor
  
```

图 3 基于排序的选择性集成学习算法框架

在文献[26]中, Martínez-Muñoz 和 Suárez 提出基于 Boosting 主要思想来对 Bagging 集成分类器进行修剪的算法,并给出了两种确定参数 S 的取值方法:(1)直接令 $S \approx 40\% \times T$,因为作者之前进行试验的一些结果^[23]表明该准则效果较好;(2)第一个 Boosting 停止点准则,即在选择到某个基分类器 C_s^* 的加权训练误差 $\epsilon_s > 0.5$ 时,停止选择过程. Martínez-Muñoz 和 Suárez 采用 16 个 UCI 实际数据集和 2 个人工数据集进行试验,结果表明上述两种停止准则性能都较好.值得指出的是,文献[26]中选择基分类器的过程是基于训练集 \mathcal{L}_{tr} (同时也用于基分类器的训练)进行的,在实际应用中,为了避免出现过拟合现象,上述选择过程一般要基于验证集 \mathcal{L}_{val} 进行.

此外, Bryll 等人^[22]先采用绕封(wrapper)的特征选择方法确定特征子集,基于该特征子集训练基分类器,根据其分类准确性对所有的特征子集进行排序,最后选用得到秩数较高的一些子集所训练的基学习机来构建集成分类器. Martínez-Muñoz 和

Suárez^[24]定义了一个参照向量(reference vector)和一个对应于 Bagging 集成中每个基分类器的标号向量(signature vector),利用相应的标号向量偏离参照向量的程度对基分类器进行排序,通过及早停止合并过程而达到修剪 Bagging 集成的目的. Croux 等人^[25]则采用 Out-of-bag 样本^[30]估计基于 Bagging 技术所生成的每个基分类器的泛化误差,并对其排序,通过预先设置的阈值将泛化误差较大的基分类器剔除. Martínez-Muñoz 等人^[27]则对通过排序修剪 Bagging 集成分类器的技术作了详细的分析和研究,并指出可用于对基学习机排序的指标大致包括误差的减小量、 κ 统计量^[31]、互补性、边缘距离等. Rokach^[28]提出了 CAP(Collective-Agreement-based Pruning)选择性集成学习算法,对每个基分类器子集, CAP 基于其中每个成员的预测能力和它们之间的冗余性(用对称的不确定性或 κ 统计量来衡量)对这些子集进行排序,并选择冗余性小且平均预测精度高的基分类器子集. 张春霞和张讲社^[29]也采用 Boosting 的思想提出了对 Double-Bagging 集成分类器^[32]进行修剪的技术,并取得了较好的效果.

2.3 基于选择的方法

依据某种选择标准,只选择部分基分类器来参与集成学习是最直观的选择性集成学习方法,现有的多数相关算法^[31,33-57]都属于此类.前述的基于排序的方法与此类方法密切相关,从广义上讲,基于排序的方法也属于选择类的方法.在选择类方法中,按照它们是否采用统一模型对检验集中的所有个体进行预测,又可以分为静态选择法(static selection method)和动态选择法(dynamic selection method).

2.3.1 静态选择法

该方法^[31,33-46]的主要特点是基于已有的基分类器,从中选择一部分构建集成分类器,并用其对所有的检验个体进行预测.各种方法之间的区别是采用不同的度量标准来选择基分类器.静态选择性集成学习算法的基本框架如图 4 所示.

```

输入:验证集  $\mathcal{L}_{val}$ , 评测标准  $\mathcal{M}$ , 基分类器集合  $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$ , 选择的基分类器个数  $S$ 
输出:选择的基分类器集合  $S = \{C_1^*, C_2^*, \dots, C_S^*\}$ 
选择过程:
初始化:令  $S = \emptyset$ .
For  $s = 1, 2, \dots, S$ 
    在验证集  $\mathcal{L}_{val}$  上, 依据评测标准  $\mathcal{M}$  从  $\mathcal{T}$  中选择效果最好的基分类器  $C_s^*$ .
    将  $C_s^*$  加入集合  $S$  中, 并将其从  $\mathcal{T}$  中删除, 即令  $S = S \cup C_s^*$ ,  $\mathcal{T} = \mathcal{T} \setminus C_s^*$ .
EndFor
  
```

图 4 静态选择性集成学习算法框架

在文献[31]中, Margineantu 和 Dietterich 提出基于 Kappa-Error 图来修剪 AdaBoost 集成分类器, 以保证选择的基分类器的准确性和它们之间的多样性, 进而使它们组成的集成分类器的预测性能与原来的相比没有明显的下降; Tamon 和 Xiang^[33] 进一步改进了该方法, 经理论分析, 他们发现 Boosting 集成的修剪问题甚至用逼近的方法都难以解决, 于是提出了基于边缘 (margin) 来选取最优基分类器子集的启发式方法。

Caruana 等人^[34] 针对多种不同的学习算法如神经网络、支撑向量机 (SVM)、决策树、Bagging 和 Boosting 分类树集成等, 通过将把这些算法采用不同的参数生成了大量基分类器 (2000 个), 并利用前向逐步选择的方法来最大化选择性集成分类器在验证集上的分类效果。Banfield 等人^[35] 基于基分类器的准确性设计了顺序后向选择方法 (sequential backward selection)、AID (Accuracy In Diversity) 修剪算法和协同 (concurrency) 修剪算法。Demir 和 Alpaydin^[36] 基于效用理论 (utility theory), 提出了一种对损失敏感且期望效用最大化的选择性集成学习算法, 试验结果表明: 该方法在实际中通过达到准确性和检验损失之间的折中, 可以成功地选择一些基分类器使得它们组成的集成分类器具有较强的泛化能力。Aksela 和 Laaksonen^[37] 定义了指数误差数 (Exponential Error Count, EEC) 用于度量基分类器误差之间的多样性, 通过最小化 EEC 来选择最优的基分类器子集, 在手写字符数据集上进行试验的结果表明该方法具有较好的效果。Rokach 等人^[38] 针对异质类型基分类器的选择进行了研究, 通过定义 PEM 量 (potential extract measure) 提出了一种选择性投票 (selective voting) 方法。

Shahjahan 和 Murase^[39] 则提出了修剪相互协作神经网络集成的算法 PNNE (Pruning NN ensemble)。对每个 NN, 他们引入了隐节点的一个协作函数来支持衰减过程, 并基于负相关学习 (negative learning) 增加各个 NN 间的多样性, 进而将过拟合的 NN 从中剔除。Hu 等人^[40] 提出了 FS-PP-EROS (Forward Search-Post Pruning-Ensemble multiple ROugh subspaces) 算法, 该算法首先基于粗糙集理论对数据的特征进行约简 (rough set-based attribute reduction), 得到一些特征子集, 然后基于每个特征子集训练一个基分类器, 并采用精度驱动的前向搜索和修剪的策略选择基分类器。唐耀华等人^[43] 利用 $\xi\alpha$ 误差估计法度量个体 SVM 泛化性, 并

基于负相关学习引入差异性, 通过递归删除法选择一组泛化能力优良、差异性大的 SVM 参与集成学习。

针对半朴素贝叶斯分类器 (seminative Bayesian classifier) 族中的超 1-依赖贝叶斯分类器 (Super-Parent-One-Dependence Estimator, SPODE) 在集成学习过程中应该进行选择还是加权合并的问题, Yang 等人^[41] 采用 58 个数据集对 16 种方法进行了大规模的研究, 为该类方法在实际中的应用提供了有效的指导。Folino 等人^[42] 提出基于分格遗传规划 (cellular genetic programming) 的方法来构建 Boosting 集成分类器, 并采用遗传规划中的结构多样性 (structural diversity) 度量 (主要有两个树之间的距离、一个树到空树的距离和 κ 统计量) 来对集成分类器进行修剪。Meynet 和 Thiran^[44] 在信息论框架下建立了基分类器的准确性和多样性之间的联系, 并建议基于信息理论得分 (Information Theoretic Score, ITS) 来修剪集成分类器, 相比于使用多样性来选择最优基分类器子集的方法, ITS 算法更具优势。

Ting 和 Witten^[45] 则提出用交叉确认法选择一个最好性能的基分类器来对检验数据进行预测。对于商业中的破产预测 (bankruptcy prediction) 二分类问题, Hung 和 Chen^[46] 建议基于破产和不破产的期望概率从决策树、后向传播神经网络和 SVM 三个分类器中进行选择, 以汲取它们的优点并克服其缺点。经实际验证, 这种方法比其它的加权或投票方法的预测效果要好。

2.3.2 动态选择法

这类方法^[47-57] 的主要特点是对检验集 \mathcal{L}_{ts} 中的每个个体 \mathbf{x} , 从已有的基分类器 C_1, C_2, \dots, C_T 中动态挑选合适的一部分对其进行预测, 每个个体选用的基分类器子集一般是不同的。图 5 给出了动态选择性集成学习算法的主要步骤。

<p>输入: 验证集 \mathcal{L}_{val}, 检验数据集 $\mathcal{L}_{ts} = \{\mathbf{x}_i\}_{i=1}^{N_{ts}}$, 评测标准 \mathcal{M}, 基分类器集合 $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$, 邻域点个数 k</p> <p>输出: 检验数据的类标签集合 $\Phi = \{y_i\}_{i=1}^{N_{ts}}$</p> <p>选择过程:</p> <p>For $i = 1, 2, \dots, N_{ts}$</p> <p> 针对检验样本点 \mathbf{x}_i, 在验证集 \mathcal{L}_{val} 上寻找它的 k 个最近邻样本点;</p> <p> 根据评测标准 \mathcal{M} 评价每个基分类器 $C_t (t = 1, 2, \dots, T)$ 在 k 个最近邻样本点上的性能;</p> <p> 选取一个或多个较好的基分类器对 \mathbf{x}_i 进行预测, 得到估计的类标签 y_i.</p> <p>EndFor</p>

图 5 动态选择性集成学习算法框架

Woods 等人^[47] 提出了一种基于局部精度的动态选择法 DCS-LA (Dynamic Classifier Selection-

Local Accuracy), 通过在 5 个实际数据集上将 DCS-LA 的分类精度和 ROC (Receiver Operating Characteristic) 曲线下的面积与其它几种合并准则 (行为知识空间 BKS、基于秩的方法和改进的秩方法) 进行比较, 发现 DCS-LA 总能改进具有最高预测精度的单个分类器的性能. Giacinto 和 Roli^[48] 根据多分类器的行为提出了 DCS-MCB 算法 (Dynamic Classifier Selection-Multiple Classifier Behavior), 它与上述的 DCS-LA 算法存在两点差别. 在 DCS-MCB 中, k 是随着检验个体的不同而发生变化的. 另外, 只有当一个基分类器的 LA 明显高于其它分类器时, DCS-MCB 才选择单个分类器对 x 预测; 否则, 它采用简单多数投票的方式构建集成分类器来对 x 分类. 与 DCS-LA 密切相关的还有 Kuncheva^[50] 提出的 DCS-DT 算法 (Dynamic Classifier Selection-Decision Templates), 该方法的思想与 DCS-MCB 类似, 但它采用成对的 t 检验 (paired t -test) 方法来判别一个基分类器是否在 LA 方面具有显著优势, 如果各个基分类器 LA 之间的差别不显著, 则基于决策表矩阵 (decision template matrix) 来构建集成分类器对 x 预测. Didaci 和 Giacinto^[51] 发现 DCS-LA 方法的性能不仅依赖于邻域形状和大小的选择, 同时还与训练个体的局部密度分布 (local density distribution) 有关. 于是, 他们提出了一种自适应选取邻域以更好估计 LA 的方法, 试验结果表明通过适当调整一些额外参数, 可以进一步改进 DCS-LA 的预测效果. Canuto 等人在文献^[52] 中较全面地研究了异质类型基分类器的数目 T 的变化对上述几种 DCS 方法 (DCS-LA, DCS-MCB 和 DCS-DT) 性能的影响, 并得出结论: DCS-DT 在多数情况下效果较好.

此外, Fan 等人^[49] 针对集成学习在对损失敏感 (cost-sensitive) 的大规模数据集上的应用, 提出了动态策略 (dynamic scheduling) 与基于收益的贪婪方法 (benefit-based greedy pruning) 相结合的集成分类器修剪算法, 该方法可以在不损失预测精度的前提下, 剔除约 90% 的基分类器. Ko 等人^[53] 基于 Oracle 的概念提出了 4 种动态选择基分类器子集的方法 (KNORA-ELIMINATE、KNORA-UNION、KNORA-ELIMINATE-W、KNORA-UNION-W). 它们的特点是对检验个体 x , 基于验证集 \mathcal{L}_{val} 确定 x 的 k 个邻域点, 挑选出那些对 k 个邻域点分类正确的基分类器, 并根据简单投票或加权投票的方式合并

被选出的基分类器来预测 x 的类标签. Dos Santos 等人^[54] 提出了优化和动态选择相结合的两阶段选择法, 目的是对检验集 \mathcal{L}_{tr} 中的每个个体都搜索到最优的基分类器子集. 优化阶段主要是为了产生具有高精度的集成分类器总体, 优化目标是集成分类器的分类误差和多样性; 而动态选择阶段则为每个检验个体, 挑选出具有最高置信度的集成分类器, 用于度量置信度的指标可以是模糊性 (ambiguity)、边缘 (margin) 和相对于最近类的力量 (strength relative to the closest class). Cavalin 等人^[56] 则从文献^[54] 提出的方法 DSA (Dos Santos et al's Approach) 出发, 通过利用验证集 \mathcal{L}_{val} 提供的上下文信息 (contextual information) 和基分类器提供的证据 (evidence), 并引入切换策略 (switch mechanism) 以处理类得票相等 (tie-breaking) 和大边缘决策 (large-margin decision) 问题, 提出了一种新的动态选择分类器集成的方法 DSA'. 考虑到实际数据中通常含有噪声, Xiao 等人^[57] 在对噪声具有较强免疫性的启发式数据挖掘方法 GMDH (Group Method of Data Handling) 的基础上, 根据基分类器的准确性和多样性提出了动态选择集成分类器的 GDES-AD 算法 (GMDH-based Dynamic classifier Ensemble Selection according to Accuracy and Diversity). 采用误差的偏差-方差分解对集成分类器的研究表明, GDES-AD 对噪声的强免疫能力主要在于它在降低分类误差的偏差方面更具优势.

此外, Hernández-Lobato 等人^[55] 在贝叶斯理论框架下, 探讨了通过将一种随机学习算法应用于一个给定的训练集上所生成的独立的同质类型基分类器的选择问题, 提出了基于个体的修剪方法 (Instance-Based pruning, IB). Hernández-Lobato 等人经理论分析, 得出结论: 对需要预测的个体 x , 若基于逐一添加基分类器和简单多数投票的合并方式来构建集成分类器, 当对 x 预测的精度达到事先给定的置信水平且对 x 预测的类标签不再发生变化时, 即可停止基分类器的添加. 基于一些分类问题的基准数据集, Hernández-Lobato 等人采用 Bagging 和 Random Forest 技术生成基分类器, 验证了理论分析的正确性和 IB 算法的有效性, 并指出需要选择的基分类器个数 S 是与要预测的个体 x 密切相关的. 当基分类器对 x 的预测结果基本一致时, 只需少数几个来构建集成分类器即可达到较好效果; 对于其它情形, 尤其是 x 接近分类边界时, 则需较多的基分类器.

2.4 基于优化的方法

这类方法^[14,58-69]的主要思想是在基分类器的合并过程中对它们赋予权重,通过稀疏性约束或设置阈值,借助于优化算法来选择最优的基分类器子集.图 6 列出了基于优化的选择性集成学习算法的一般步骤.

<p>输入: 验证集 \mathcal{L}_{val}, 基分类器集合 $\mathcal{T} = \{C_1, C_2, \dots, C_T\}$, 评测标准 \mathcal{M}, 优化算法 A, 阈值 λ</p> <p>输出: 选择的基分类器集合 $S = \{C_1^*, C_2^*, \dots, C_S^*\}$</p> <p>选择过程:</p> <p>对每个基分类器 $C_t (t=1, 2, \dots, T)$ 赋予初始权重 ω_t;</p> <p>根据评测标准 \mathcal{M}, 选择合适的目标函数 f (如遗传算法中的适应值函数等);</p> <p>基于验证集 \mathcal{L}_{val}, 利用算法 A 优化目标函数 f, 得到最优权重向量 $\omega^* = (\omega_1, \omega_2, \dots, \omega_T)$;</p> <p>选择满足条件 $\omega^* > \lambda$ 的元素所对应的基分类器 $C_1^*, C_2^*, \dots, C_S^*$.</p>

图 6 基于优化方法的选择性集成学习算法框架

在基于优化方法的选择性集成学习算法中,文献^[14,58-60]均采用遗传算法(Genetic Algorithm, GA)来优化每个基分类器被赋予的权重,而它们之间的主要区别在于采用了不同的遗传算法编码方法(二进制或实值).在 2002 年,周志华等人^[14]首次从理论上证明了选择性集成学习的有效性,并提出了基于实值编码遗传算法的选择性集成学习方法 GASEN (Genetic Algorithm based Selective ENsemble).在试验过程中,他们将算法中的阈值 λ 设定为 0.05,研究了上述策略修剪 Bagging 和 Boosting 神经网络集成的效果,并得出结论:在所探讨的数据集上, GASEN 方法都可以用较少的神经网络得到更好的泛化效果.同时,他们还指出通过调整遗传算法中的适应值函数、编码方法、遗传算子以及阈值 λ , GASEN 算法的性能都可以得到进一步改进.

王丽丽和苏德富^[61]针对已有的选择性集成学习方法计算复杂性高、效率低的特点,提出将具有快速收敛性质的群体智能算法用于多分类器的选择. Zhang 等人^[62]将多分类器的选择问题看成是二次整数规划问题(quadratic integer programming),并给出了一种半定规划方法 SDP (Semi-Definite Programming),试验结果表明该方法可以比其它启发式方法更好地逼近基分类器的最优子集. Chen 等人^[63]则从概率推理的角度出发提出了一种新的选择性集成学习算法,为了使每个基分类器的权重非负且尽量稀疏(即多数基分类器得到的权重为 0),算法采用左截断的高斯分布(left-truncated Gaussian distribution)对每个基分类器的初始权重赋予先验知识,并基于期望扩散(Expectation Propagation,

EP)算法来估计权重的后验分布.该算法的优点是在 EP 算法的训练过程中,不需要额外的计算即可获得 LOO(Leave-One-Out)误差,结合贝叶斯证据(Bayesian evidence)也可以实现模型选择.

为了避免标准的线性最小二乘回归合并准则(standard linear least squares regression)出现过拟合现象, Reid 和 Grudic^[66]提出将基于层叠泛化策略(stacked generalization)的正则化线性模型用于基分类器的合并,模型中的未知参数可以用岭回归(ridge regression)、Lasso 回归(Lasso regression)和弹性网回归(elastic net regression)等方法估计得到. Li 和 Zhou^[65]在正则化的理论框架下,基于 Hinge 损失函数(对于 $\Phi = \{-1, +1\}$ 的二分类问题,其定义为 $l(y_i, C(\mathbf{x}_i)) = \max(0, 1 - y_i C(\mathbf{x}_i))$)和图的拉普拉斯正则子(graph Laplacian regularizer)将基分类器的选择问题归结为具有稀疏解的二次规划问题,并通过相关的优化算法(文中用的是 MOSEK 软件)进行求解.算法除了能用少数的基学习机达到较强的泛化能力之外,还可以充分利用无标签的数据(unlabeled data)来进一步提高分类系统的性能. Zhang 和 Zhou^[69]也基于正则化理论,采用 Hinge 损失函数和 \mathcal{L}_1 正则项提出了一种基于线性规划的方法,优化的目标是 minimized 集成分类器的训练误差同时控制权向量,以使得最终得到的权向量尽量稀疏.

Dos Santos 等人^[64]针对决策树和 K-NN(K Nearest Neighbors)作为基学习算法,采用 Bagging 和随机子空间技术训练的基分类器,提出以分类误差和多样性为优化目标,并基于单目标和多目标的遗传算法(multi-objective genetic algorithm)进行优化求解.试验结果表明,多目标优化的遗传算法除了能找到近似最优的基分类器子集之外,还可以控制过拟合现象的产生.此外,文献^[64]中以多样性为单目标优化函数的研究还有助于探索基分类器之间的多样性与集成分类器性能之间的关系.

杨晓霜和汪源源^[68]则提出了一种基于 Moore-Penrose 逆矩阵的新型选择性集成学习算法 PISEN (Pseudo-Inverse matrix based Selective ENsemble),它基于矩阵的伪逆理论对每个基学习机的权值进行优化,并通过预先设定的阈值来选择最后的基学习机子集,算法的优点是简单、易于实现且效率较高.基于 8 个 UCI 实际数据集,作者将 PISEN 算法(阈值设为 0.1)与文献^[14]提出的 GASEN 算法(采用该文中的参数设置)进行了比较,发现二者的泛化误

差相当,但前者的计算效率要远远高于后者.此外,王磊^[67]则基于约束投影矩阵法首先训练多个 SVM,然后采用遗传优化和最小化偏离度误差的技术来对 SVM 进行组合.

2.5 其它方法

除了上面提到的几类选择性集成学习算法,还有一些算法不能归入其中,我们在此对其作一简单介绍. Prodromidis 和 Stolfo^[70]基于基分类器的预测结果构建一个决策树,并对树进行修剪,如果某个基分类器的预测结果不在被修剪后的树中,则将它从构建集成分类器的过程中剔除. Tsoumakas 等人^[71]建议采用多重统计检验的方法(如 Turkey 检验、Hsu 检验、Scott 和 Knott 检验等)来选择最优的基分类器集合,使得其中每个基分类器的预测精度基本相当,且集合中至少有一个基分类器的预测性能显著优于未被选择的任何一个基分类器. Partalas 等人^[72]则首次将增强学习(reinforcement learning)的概念引入选择性集成学习中,提出了一种新的有效算法. 盛高斌^[73]针对小数据量聚类的有标记样本问题,提出了一种基于半监督回归的选择性集成学习算法 SSRES (Semi-Supervised Regression Ensemble Selection).

Zhang 和 Chau^[74]提出了基于多子群粒子群优化算法(Multi-Sub-Swarm Particle Swarm Optimization, MSSPSO)的多层次修剪模型,在每一层修剪过程中,模型假定每个基分类器都产生一个明智的输出(oracle output),并把基分类器的选择看成是多模态的优化问题,基于前一层基分类器的输出采用 MSSPSO 算法进行求解,最终选择出包含重要信息的基分类器. Soto 等人^[75]针对二分类问题,将前述的对基分类器进行排序的方法与基于个体(Instance-Based, IB)的动态修剪技术相结合,提出了一种两层的修剪算法. 算法首先采用排序(如基于 Boosting 的主要思想^[26])的方法选出约 20%的基分类器并设置置信值 α ,在对某个个体 x 预测时,通过简单多数投票的方式逐一合并选出的基分类器,直至集成分类器预测的某个类的概率大于 α . 该算法的优点是能明显提高集成分类器的预测精度、降低存储需求(只需存储约 20%的基学习机)、加快预测过程(基学习机的数目减少且 IB 修剪技术加快了对每个个体的预测).

3 基于回归问题的选择性集成学习算法

解决回归问题的多数选择性集成学习算法都是

经过对用于分类问题的相关算法进行推广得到的,如文献[3,14,29,76-81]等.若将这些算法按照其特点进行分类,则也可以类似地分为基于聚类、排序、优化和选择的方法.其中,Partalas 等人^[3]针对选择性集成学习在水质量预测问题中的应用,提出基于学习机的均方误差根(root-mean-squared-error)采用前向选择(forward selection)和后向删除(backward elimination)的策略贪婪地搜索最优的基学习机子集.周志华等人^[14,78]建议采用实值编码的遗传算法和设置阈值的方法来对基学习机进行选择. Hernández-Lobato 等人^[77]基于基学习机之间的互补性对它们进行排序,并采用贪婪算法从中选取约 20%的基学习机.

与分类问题相比,对基于回归问题的选择性集成学习算法进行的研究相对较少,究其原因,大概是由以下因素造成的.选择性集成学习的目的是在不降低甚至进一步提高原集成学习机预测精度的前提下,尽可能减少参与集成学习的基学习机的数目,以大大降低其对存储空间的需求并加快预测速度.然而,在回归问题中,采用某种策略对基学习机进行选择之后,被剔除的基学习机往往较少,而在预测精度的改进方面效果也不太明显.

4 总结与展望

本文对机器学习领域中一些具有代表性的选择性集成学习算法作了一个比较全面的综述,并分析了几类方法的主要特点.选择性集成学习的提出极大地丰富了集成学习的相关理论,并为其它相关领域的研究提供了新技术和新思路,研究前景广阔.近年来虽然仍有很多研究人员致力于该方面的研究,发表了很多研究成果,但目前的方法多是针对某个具体任务而言的,其中尚有一些问题需要在未来研究中得到突破:

(1) 若干关键参数的确定.例如,在基于聚类技术、排序和选择的方法中,如何确定最终保留的基学习机的数目 S ?在基于优化的方法中,如何设置权重的阈值以剔除某些效果较差的基学习机?针对这些参数,目前采用的大多是启发式的方法.如果选取不合适,则会大大影响预测效果.如何根据具体问题,自适应地选取这些关键参数将是一个值得研究的内容.

(2) 选择基学习机的衡量指标的选定.基学习机的准确性和它们之间的多样性在集成学习机的构建过程中起着至关重要的作用,只有当两者达到一

个较好的折中,集成学习机才能具有较强的泛化能力.然而,多样性在实际中难以衡量,且多样性和准确性与集成学习机预测性能之间的有效联系也较难建立.在选择性集成学习算法的设计中,如何选取合适的准则或度量标准将准确性和多样性因素充分考虑在内,也是一个需要解决的关键问题.

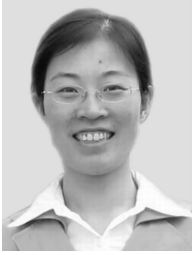
(3) 选择性集成学习算法的应用研究.目前,一些选择性集成学习方法已经应用在疾病诊断、人脸识别、图像挖掘等方面.研究该类方法更广泛的实际应用以及现有的机器学习方法在选择性集成学习中的应用都将有十分重要的意义.

参 考 文 献

- [1] Dietterich T G. Machine learning research: Four current directions. *AI Magazine*, 1997, 18(4): 97-136
- [2] Takemura A, Shimizu A, Hamamoto K. Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the AdaBoost algorithm with feature selection. *IEEE Transactions on Medical Imaging*, 2010, 20(3): 598-609
- [3] Partalas I, Tsoumakas G, Hatzikos E V, Vlahavas I. Greedy regression ensemble selection: theory and an application to water quality prediction. *Information Sciences*, 2008, 178(20): 3867-3879
- [4] Diego I M, Serrano A, Conde C, Cabello E. Face verification with a kernel fusion method. *Pattern Recognition Letters*, 2010, 31(9): 837-844
- [5] Breiman L. Bagging predictors. *Machine Learning*, 1996, 24(2): 123-140
- [6] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55(1): 119-139
- [7] Ho T K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844
- [8] Rodríguez J J, Kuncheva L I, Alonso C J. Rotation Forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(10): 1619-1630
- [9] Kuncheva L I. Using diversity measures for generating error-correcting output codes in classifier ensembles. *Pattern Recognition Letters*, 2005, 26(1): 83-90
- [10] Breiman L. Random forest. *Machine Learning*, 2001, 45(1): 5-32
- [11] Xu L, Krzyzak A, Suen C Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1992, 22(3): 418-435
- [12] Paclik P, Landgrebe TCW, Tax DMJ, Duin RPW. On deriving the second-stage training set for trainable combiners// Roli F, Oza N C, Polikar R. *Multiple Classifier Systems*. Berlin Heidelberg: Springer-Verlag, 2005: 136-146
- [13] Zhang C X. Research of some algorithms in ensemble learning [Ph. D. dissertation]. Xi'an Jiaotong University, Xi'an, 2009(in Chinese)
(张春霞. 集成学习中有关算法的研究[博士学位论文]. 西安交通大学, 西安, 2009)
- [14] Zhou Z H, Wu J X, Tang W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 2002, 137(1-2): 239-263
- [15] Tsoumakas G, Partalas I, Vlahavas I. A taxonomy and short review of ensemble selection//Proceedings of the 18th European Conference on Artificial Intelligence. Patras, Greece, 2008: 1-6
- [16] Giacinto G, Roli F. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 2001, 22(1): 25-33
- [17] Lazarevic A, Obradovic Z. Effective pruning of neural network classifier ensembles//Proceedings of International Joint Conference on Neural Networks. Washington DC, 2001: 796-801
- [18] Bakker B, Heskes T. Clustering ensembles of neural network models. *Neural Networks*, 2003, 16(2): 261-269
- [19] Fu Qiang. Investigations upon the algorithms for selective ensemble of neural networks[Ph. D. dissertation]. Zhejiang University, Hangzhou, 2007(in Chinese)
(傅强. 选择性神经网络集成算法研究[博士学位论文]. 浙江大学, 杭州, 2007)
- [20] Guo Hong-Ling. Research on the kry technologies of selective multiple classifiers ensemble[M. S. dissertation]. Jiangsu University, Zhenjiang, Jiangsu, 2007(in Chinese)
(郭红玲. 多分类器选择关键技术研究. 江苏大学, 镇江, 江苏, 2007)
- [21] Cheng X Y, Guo H L. The technology of selective multiple classifiers ensemble based on kernel clustering//Proceedings of the 2nd Symposium on Intelligent Information Technology Application. Shanghai, China, 2008: 146-150
- [22] Bryll R, Gutierrez-Osuna R, Quek F. Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 2003, 36(6): 1291-1302
- [23] Martínez-Munoz G, Suárez A. Aggregation ordering in bagging//Proceedings of the IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria, 2004: 258-263
- [24] Martínez-Munoz G, Suárez A. Pruning in ordered bagging ensembles//Proceedings of the 23th International Conference on Machine Learning. Pittsburgh, PA, 2006: 609-616
- [25] Croux C, Joossens K, Lemmens A. Trimmed bagging. *Computational statistics & data analysis*, 2007, 52(1): 362-368

- [26] Martínez-Munoz G, Suárez A. Using boosting to prune bagging ensembles. *Pattern Recognition Letters*, 2007, 28(1): 156-165
- [27] Martínez-Munoz G, Hernández-Lobato D, Suárez A. An analysis of ensemble pruning techniques based on ordered aggregation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 245-259
- [28] Rokach L. Collective-agreement-based pruning of ensembles. *Computational Statistics & Data Analysis*, 2009, 53(4): 1015-1026
- [29] Zhang C X, Zhang J S, Zhang G Y. Using boosting to prune double-bagging ensembles. *Computational Statistics & Data Analysis*, 2009, 53(4): 1218-1231
- [30] Breiman L. Out-of-bag estimation. Berkeley: Statistics Department in University of California. Technical Report; 1996
- [31] Margineantu D D, Dietterich T G. Pruning adaptive boosting//Proceedings of the 14th International Conference on Machine Learning. San Francisco, Morgan Kaufmann, 1997: 211-218
- [32] Hothorn T, Lausen B. Double-bagging: combining classifiers by bootstrap aggregation. *Pattern Recognition*, 2003, 36(6): 1303-1309
- [33] Tamon C, Xiang J. On the boosting pruning problem//Proceedings of the 11th European Conference on Machine Learning, Catalonia, Spain, 2000: 404-412
- [34] Caruana R, Niculescu-Mizil A, Crew G, Ksikes A. Ensemble selection from libraries of models//Proceedings of the 21st International Conference on Machine Learning. Banff, Canada, 2004: 137-144
- [35] Banfield R E, Hall L O, Bowyer K W, Kegelmeyer W P. Ensemble diversity measures and their application to thinning. *Information Fusion*, 2005, 6(1): 49-62
- [36] Demir C, Alpaydin E. Cost-conscious classifier ensembles. *Pattern Recognition Letters*, 2005, 26(14): 2206-2214
- [37] Aksela M, Laaksonen J. Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*, 2006, 39(4): 608-623
- [38] Rokach L, Maimon O, Arbel R. Selective voting-getting more or less in sensor fusion. *International Journal of Pattern Recognition and Artificial Intelligence*, 2006, 20(3): 329-350
- [39] Shahjahan M, Murase K. A pruning algorithm for training cooperative neural network ensembles. *IEICE Transactions on Information and Systems*, 2006, E89-D(3): 1257-1269
- [40] Hu Q H, Yu D R, Xie Z X, Li X D. EROS: Ensemble rough subspaces. *Pattern Recognition*, 2007, 40(12): 3728-3739
- [41] Yang Y, Webb G I, Cerquides J, Korb K B, Boughton J, Ting K. To select or to weigh; A comparative study of linear combination schemes for superparent-one-dependence estimators. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(12): 1652-1665
- [42] Folino G, Pizzuti C, Spezzano G. Training distributed GP ensemble with a selective algorithm based on clustering and pruning for pattern classification. *IEEE Transactions on Evolutionary Computation*, 2008, 12(4): 458-468
- [43] Tang Yao-Hua, Gao Jing-Huai, Bao Qian-Zong. Novel selective support vector machine ensemble learning algorithm. *Journal of Xi'an Jiaotong University*, 2008, 42(10): 1221-1225(in Chinese)
(唐耀华, 高静怀, 包乾宗. 一种新的选择性支持向量机集成学习算法. *西安交通大学学报*, 2008, 42(10): 1221-1225)
- [44] Meynet J, Thiran J P. Information theoretic combination of pattern classifiers. *Pattern Recognition Letters*, 2010, 43(10): 3412-3421
- [45] Ting K M, Witten I H. Issues in stacked generalization. *Journal of Artificial Intelligence Research*, 1999, 10: 271-289
- [46] Hung C, Chen J H. A selective ensemble based on expected probabilities for bankruptcy prediction. *Experts Systems with Applications*, 2009, 36(3): 5297-5303
- [47] Woods K, Kegelmeyer Jr W P, Bowyer K. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(4): 405-410
- [48] Giacinto G, Roli F. Dynamic classifier selection based on multiple classifier behavior. *Pattern Recognition*, 2001, 34(9): 1879-1881
- [49] Fan W, Chu F, Wang H X, Yu P S. Pruning and dynamic scheduling of cost-sensitive ensembles//Proceedings of the 18th National Conference on Artificial Intelligence, USA, 2002. USA: American Association for Artificial Intelligence, 2002: 146-151
- [50] Kuncheva L I. Switching between selection and fusion in combining classifiers: An experiment. *IEEE Transactions on System, Man and Cybernetics, Part B*, 2002, 32(2): 146-156
- [51] Didaci L, Giacinto G. Dynamic classifier selection by adaptive k -nearest-neighborhood rule//Roli F, Kittler J, Windatt T. *Multiple Classifier Systems*. Berlin Heidelberg: Springer-Verlag, 2004: 174-183
- [52] Canuto A M P, Abreu M C C, Oliveira L M, Xavier Jr J C, Santos A M. Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles. *Pattern Recognition Letters*, 2007, 28(4): 472-486
- [53] Ko A H R, Sabourin R, Jr A S B. From dynamic classifier to dynamic ensemble selection. *Pattern Recognition*, 2008, 41(5): 1718-1731
- [54] Dos Santos E M, Sabourin R, Maupin P. A dynamic over-produce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition*, 2008, 41(10): 2993-3009
- [55] Hernández-Lobato D, Martínez-Munoz G, Suárez A. Statistical instance-based pruning in ensembles of independent classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 364-369

- [56] Cavalin P R, Sabourin R, Suen C Y. Dynamic selection of ensembles of classifiers using contextual information//Gayer N E, Kittler J, Roli F. Multiple Classifier Systems. Berlin Heidelberg: Springer-Verlag, 2010: 145-154
- [57] Xiao J, He C Z, Jiang X Y, Liu D H. A dynamic classifier ensemble selection approach for noise data. Information Sciences, 2010, 180(18): 3402-3421
- [58] Thompson S. Pruning boosted classifiers with a real valued genetic algorithm. Knowledge-Based Systems, 1999, 12(5-6): 277-284
- [59] Zhou Z H, Tang W. Selective ensemble of decision trees//Proceedings of the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Chongqing, China, 2003: 476-483
- [60] Hernández-Lobato D, Hernández-Lobato J M, Ruiz-Torrubiano R, Valle Á. Pruning adaptive boosting ensembles by means of a genetic algorithm//Corchado et al. International Conference on Intelligent Data Engineering and Automated Learning. Berlin Heidelberg: Springer-Verlag, 2006: 322-329
- [61] Wang Li-Li, Su De-Fu. Swarm intelligence-based selective ensemble with decision tree classifiers. Computer Technology and Development, 2006, 16(12): 55-57(in Chinese)
(王丽丽, 苏德富. 基于群体智能的选择性决策树分类器集成. 计算机技术与发展, 2006, 16(12): 55-57)
- [62] Zhang Y, Burer S, Street W N. Ensemble pruning via semi-definite programming. Journal of Machine Learning Research, 2006, 7: 1315-1338
- [63] Chen H H, Tiño P, Yao X. Predictive ensemble pruning by expectation propagation. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(7): 999-1013
- [64] Dos Santos E M, Sabourin R, Maupin P. Overfitting cautious selection of classifier ensembles with genetic algorithms. Information Fusion, 2009, 10(2): 150-162
- [65] Li N, Zhou Z H. Selective ensemble under regularization framework//Benediksson J A, Kittler J, Roli F. Multiple Classifier Systems. Berlin Heidelberg: Springer-Verlag, 2009: 293-303
- [66] Reid S, Grudic G. Regularized linear models in stacked generalization//Benediksson J A, Kittler J, Roli F. Multiple Classifier Systems. Berlin Heidelberg: Springer-Verlag, 2009: 112-121
- [67] Wang Lei. Constraint projection-based support vector machines selective ensemble algorithms. Computer Science, 2009, 36(10): 234-237(in Chinese)
(王磊. 基于约束投影的支持向量机选择性集成. 计算机科学, 2009, 36(10): 234-237)
- [68] Yang Xiao-Shuang, Wang Yuan-Yuan. Selective ensemble based on Moore-Penrose pseudo-inverse. Opto-Electronic Engineering, 2009, 36(11): 140-144(in Chinese)
(杨晓霜, 汪源源. 基于 Moore-Penrose 逆矩阵的选择性集成. 光电工程, 2009, 36(11): 140-144)
- [69] Zhang L, Zhou W D. Sparse ensembles using weighted combination methods based on linear programming. Pattern Recognition, 2011, 44(1): 97-106
- [70] Prodromidis A L, Stolfo S J. Cost complexity-based pruning of ensemble classifiers. Knowledge and Information Systems, 2001, 3(4): 449-469
- [71] Tsoumakas G, Angelis L, Vlahavas I. Selective fusion of heterogeneous classifiers. Intelligent Data Analysis, 2005, 9(6): 511-525
- [72] Partalas I, Tsoumakas G, Vlahavas I. Pruning an ensemble of classifiers via reinforcement learning. Neurocomputing, 2009, 72(7-9): 1900-1909
- [73] Gao Sheng-Bin. Research of ensemble selective algorithm based on semi-supervised regression and its application [M. S. dissertation]. Zhejiang University of Technology, Hangzhou, 2009(in Chinese)
(盛高斌. 基于半监督回归的选择性集成算法及其应用研究 [硕士学位论文]. 浙江工业大学, 杭州, 2009)
- [74] Zhang J, Chau K W. Multilayer ensemble pruning via novel multi-sub-swarm particle swarm optimization. Journal of Universal Computer Science, 2009, 15(4): 840-858
- [75] Soto V, Martínez-Munoz G, Hernández-Lobato D, Suárez A. A double pruning algorithm for classification ensembles//Gayer N E, Kittler J, Roli F. Multiple Classifier Systems. Berlin Heidelberg: Springer-Verlag, 2010: 104-113
- [76] Rooney N, Patterson D, Nugent C. Reduced ensemble size stacking//Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence. Boca Raton, Florida, 2004: 266-271
- [77] Hernández-Lobato D, Martínez-Munoz G, Suárez A. Pruning in ordered regression bagging ensembles//Proceedings of the 2006 International Joint Conference on Neural Networks. Vancouver, Canada, 2006: 1266-1273
- [78] Wang Jue, Zhou Zhi-Hua, Zhou Ao-Ying. Machine Learning and Its Applications. Beijing: Tsinghua University Press, 2006(in Chinese)
(王珏, 周志华, 周傲英. 机器学习及其应用. 北京: 清华大学出版社, 2006)
- [79] Li Yu, Xu Cheng-Xian. A method for pruning Bagging ensembles and its applications. Systems Engineering-Theory & Practice, 2008, 7: 105-110(in Chinese)
(李毓, 徐成贤. 修剪 Bagging 集成的方法及其应用. 系统工程理论与实践, 2008, 7: 105-110)
- [80] Zhao Sheng-Ying, Gao Guang-Chun. Ant colony optimization-based approach for selective neural network ensemble. Journal of Zhejiang University (Engineering Science), 2009, 43(9): 1568-1573(in Chinese)
(赵胜颖, 高广春. 基于蚁群算法的选择性神经网络集成方法. 浙江大学学报(工学版), 2009, 43(9): 1568-1573)
- [81] Hernández-Lobato D, Martínez-Munoz G, Suárez A. Empirical analysis and evaluation of approximate techniques for pruning regression bagging ensembles. Neurocomputing, 2011, 74(12-13): 2250-2264



ZHANG Chun-Xia, born in 1980, Ph. D, lecturer. Her main research interests include ensemble learning, pattern recognition, Bootstrap random sampling methods and etc.

ZHANG Jiang-She, born in 1962, Ph. D, professor. His research interests mainly contain statistical computing, pattern recognition, global optimization and etc.

Background

In recent years, ensemble learning methods are rapidly growing and attract lots of attention from pattern recognition and machine learning communities due to their potential to greatly improve the prediction accuracy of a learning system. These techniques generally work by means of first generating an ensemble set of base learning machines according to some strategies, and then the outputs from each ensemble member are combined in a suitable way to create the final prediction.

The traditional ensemble learning methods include all the ensemble members to construct the final composite learning machine. However, with the increasing number of ensemble members, the training and prediction speed of an ensemble machine decreases significantly and its storage need increases quickly. In 2002, the professor Zhou Zhihua in Nanjing University first proposed the concept of selective ensemble learning. From theoretical and experimental viewpoint, he proved that only selecting a part of ensemble members to build the ensemble machine will alleviate the above-mentioned problem while not deteriorating the prediction accuracy.

Until now, a large number of this type of techniques have been developed. This paper presents a detailed review of the current selective ensemble learning algorithms and categorizes them into different classes according to their main characteristics. Meanwhile, the advantages and disadvantages of each representative algorithm are studied. Finally, the future research directions of selective ensemble learning are discussed.

This work is supported by the National Natural Science Foundation of China Project titled “Research on Nonnegative Matrix Factorization Methods for K-Nearest Neighbor Classification Rule”, and the Research Fund for the Doctoral Program of Higher Education of China titled “Research on Some Algorithms in Ensemble Learning”.

These two projects attempt to propose some efficient classification methods, especially for small sample size problems. The paper provides a relatively complete survey of selective ensemble learning algorithms and therefore lays a solid foundation for us to develop some related methods.