

集成学习中的多样性度量

孙博, 王建东, 陈海燕, 王寅同

(南京航空航天大学 计算机科学与技术学院, 南京 210016)

摘要: 在集成学习中, 基分类器之间的多样性对于解释多分类器系统的工作机理和构造有效的集成系统具有重要作用, 但至今仍没有统一的度量多样性的方法. 首先总结介绍常用的多样性度量方法, 阐述每种方法评估多样性的角度和方式; 然后从对多样性新的解释和度量、多样性度量在选择性集成中的应用、多样性度量和集成学习精度的关系3个方面探讨多样性度量的研究进展; 最后给出关于多样性度量进一步的研究方向.

关键词: 集成学习; 多样性度量; 精度; 选择性集成; 泛化性能

中图分类号: TP181

文献标志码: A

Diversity measures in ensemble learning

SUN Bo, WANG Jian-dong, CHEN Hai-yan, WANG Yin-tong

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. Correspondent: SUN Bo, E-mail: sunbo87@126.com)

Abstract: Diversity among base classifiers plays an important role in ensemble learning for illustrating the working mechanism of multiple classifier systems and constructing effective ensemble systems. However, at present, there doesn't exist a widely accepted diversity measure. Firstly, some commonly used diversity measures are summarized, and the perspective adopted by each measure is illustrated when evaluating the diversity. Then, the research progresses of diversity measures are investigated in the following three aspects: The recently proposed interpretations and measures for diversity, the application of diversity measures in selective ensemble, and the relationship between diversity measures and ensemble accuracy. Finally, several directions for future research are given.

Key words: ensemble learning; diversity measures; accuracy; selective ensemble; generalization performance

0 引言

集成学习是机器学习中的一个热门的研究方向, Dietterich^[1]将集成学习列为机器学习的4大研究方向之首. 在分类问题中, 通常难以直接获得一个具有较好泛化能力的分类器. 集成学习提供了一种解决方法, 其思想是首先训练得到一些基分类器, 然后以适当的方式将其分类结果进行整合从而得到集成的分类结果. 经典的集成学习方法有 Bagging^[2]、AdaBoost^[3]、Random Subspace^[4]、Rotation Forest^[5]等, 每种方法都在训练集上采用一定的策略生成一些基分类器, 并通过某种组合方式将这些基分类器进行集成.

理论和实验研究表明, 集成学习能够提高基分类器的泛化能力^[6-7]. 由互补且精确的分类器进行集成得到的集成系统优于性能最好的基分类器. 直觉

上, 如果各基分类器的分类结果是相似的, 则将它们集成起来不会提高泛化能力; 如果各基分类器的分类结果是多样化的, 此时一个被某些分类器错误分类的样例可以被另外一些分类器正确分类, 则通过简单多数投票将所有分类器的结果进行组合, 可以得到正确的分类结果, 从而提高基分类器的泛化能力. 因此, 基分类器之间的多样性是影响集成系统泛化性能的重要因素. 如前文提到的 Bagging、AdaBoost 和 Random Subspace 等有效的集成学习方法, 都隐式地利用了分类器间的多样性.

对于多样性而言, 学者们已经从以下几方面进行了研究: 1) 如何显式地定义和度量分类器之间的多样性; 2) 如何使用多样性度量指导构建多分类器系统; 3) 对多样性度量与集成系统精度的关系进行分析, 并对多样性概念本身进行探讨. 本文以这3方面为线索

收稿日期: 2013-09-27; 修回日期: 2013-12-12.

基金项目: 国家自然科学基金重点项目(61139002); 江苏省博士后计划项目(1301013A); 中央高校基本科研业务费专项基金项目(NS2012134, NZ2013306).

作者简介: 孙博(1987-), 男, 博士生, 从事机器学习、数据挖掘的研究; 王建东(1945-), 男, 教授, 博士生导师, 从事人工智能、信息安全等研究.

对多样性度量进行详细的综述.

1 常用的多样性度量

如何定义和度量分类器之间的多样性是一个重要的研究问题,文献中存在多种度量多样性的定义,Kuncheva等^[8]总结了10种多样性度量定义,它们是从不同领域、不同角度提出的,然而,目前还没有关于多样性度量的一个统一的定义.大多数多样性度量都基于基分类器的 oracle 输出(即 0/1 输出),它假定已知每个样例 x 的真实类标签.如果分类器 C 对 x 正确分类,则其 oracle 输出为 1, 否则为 0. 总体上,度量可以分为两大类:成对的 (pairwise) 和非成对的 (non-pairwise) 多样性度量.成对的多样性度量首先计算每一对分类器之间的多样性值,然后用其平均值衡量集成系统的多样性;非成对多样性度量直接计算集成系统的多样性值.

1.1 成对多样性度量

在介绍成对多样性度量前,首先引入如下符号:假设有 L 个基分类器, C_i 和 C_j ($i, j = 1, 2, \dots, L, i \neq j$) 分别为两个不同的分类器; $N^{11}(N^{00})$ 为分类器 C_i 和 C_j 都对其正确(错误)分类的样例数目; $N^{10}(N^{01})$ 为满足以下要求的样例数目:分类器 $C_i(C_j)$ 对其正确分类而分类器 $C_j(C_i)$ 对其错误分类.由此,总的样例数目 N 可以表示为 $N = N^{11} + N^{00} + N^{10} + N^{01}$, 具体如表 1 所示.

表 1 两个分类器的分类结果组合情况

	C_j correct	C_j incorrect
C_i correct	N^{11}	N^{10}
C_i incorrect	N^{01}	N^{00}

4 种成对多样性度量如下:

1) Q 统计.

Q 统计^[9]来自统计学领域,两个分类器 C_i 与 C_j 之间的 Q 统计值为

$$Q_{ij} = \frac{N^{11}N^{00} - N^{10}N^{01}}{N^{11}N^{00} + N^{10}N^{01}}. \quad (1)$$

由式(1)可见,对于每个样例,如果两个分类器总是同时对其正确或错误分类,即其行为是相同的,则有 $N^{10} = N^{01} = 0$, 即 $Q_{ij} = 1$, 此时它们之间的多样性程度最低;反之,如果两个分类器在每个样例上的分类结果都不同,则 $N^{11} = N^{00} = 0$, 即 $Q_{ij} = -1$, 这种情况下多样性程度最高;对于统计独立的两个分类器, Q_{ij} 期望值为 $Q_{ij} = 0$.

对于多分类器系统 C_1, C_2, \dots, C_L , 其 Q 统计多样性值可以通过计算每对分类器之间多样性的平均值得到,即

$$Q = \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^L Q_{ij}. \quad (2)$$

类似地,对于下面 3 种成对多样性度量,也是首先计算每对分类器之间的多样性值,然后将其平均值作为集成系统的多样性值.

2) 相关系数 ρ .

两个分类器 C_i 与 C_j 之间的相关系数^[8]定义为

$$\rho_{ij} = \frac{N^{11}N^{00} - N^{10}N^{01}}{\sqrt{(N^{11} + N^{10})(N^{01} + N^{00})(N^{11} + N^{01})(N^{10} + N^{00})}}. \quad (3)$$

可以看出,对于两个分类器 C_i 和 C_j , ρ_{ij} 与 Q_{ij} 有相同的符号,可以证明它们存在如下关系:

$$|\rho_{ij}| \leq |Q_{ij}|. \quad (4)$$

与相关系数 ρ 相比, Q 统计更简单且更容易计算.

3) 不一致度量 dis.

Skalak^[10]使用不一致度量衡量一个分类器和互补分类器之间的多样性, Ho^[11]利用不一致度量衡量决策森林 (decision forest) 中的多样性.不一致度量关注两个分类器 C_i 与 C_j 分类结果不同的样例,它们之间的不一致度量 dis_{ij} 定义为

$$\text{dis}_{ij} = (N^{10} + N^{01})/N. \quad (5)$$

直觉上,这样的样例越多,两个分类器之间的多样性程度越高;反之,多样性程度越低.容易看出, dis_{ij} 的取值范围为 $[0, 1]$.

4) 双次失败度量 DF.

Giacinto 等^[12]使用双次失败度量为分类器集合生成一个成对多样性矩阵,通过该矩阵选择最不相关的分类器.双次失败度量关注分类器 C_i 和 C_j 均将其错误分类的样例,该度量定义为

$$\text{DF}_{ij} = N^{00}/N. \quad (6)$$

可以看出,这样的样例越多,两个分类器越倾向于在相同的样例上出错.极端地,如果对于每个样例 x , C_i 和 C_j 均将其错误分类,即 $\text{DF}_{ij} = 1$, 则两个分类器的精确性和它们之间的多样性程度达到最低.

1.2 非成对多样性度量

与成对多样性度量不同,非成对多样性度量直接定义在多分类器系统上.首先引入如下符号: z_j ($j = 1, 2, \dots, N$) 为 N 个样例中的第 j 个样例, $l(z_j)$ 为对 z_j 正确分类的分类器的数目, L 为总的分类器数目.以下介绍几种常用的非成对多样性度量:

1) Kohavi-Wolpert 方差 KW.

KW 方差来源于对分类器分类误差的分解.对于一个特定分类器, Kohavi 等^[13]给出了对样例 x 预测类标签 y 的方差表达式. Kuncheva 等^[8]根据这一思想,在只考虑 correct、incorrect 两种分类器输出的形式下,给出了在训练集上使用分类器 C_1, C_2, \dots, C_L 对 x 预

测类标签的方差表达式, 得到如下 KW 方差公式:

$$KW = \frac{1}{NL^2} \sum_{j=1}^N l(z_j)(L - l(z_j)). \quad (7)$$

对于每个样例 z_j , KW 方差度量均要统计出对其正确分类和错误分类的分类器数目. 极端情况下, 对于每个样例 z_j , 所有分类器都给出相同的分类结果, 即 $l(z_j) = L$ 或 $l(z_j) = 0$, 此时 $KW = 0$, 该集成系统的多样性程度最低; 当有一半的分类器对 z_j 正确分类、一半的分类器对其错误分类时, $KW = 1/4$, 集成系统的多样性程度最高.

2) Measurement of interrater agreement κ .

该度量来源于统计学, 用于度量评价者间的可靠性, Fleiss^[14]将其用于度量分类器间的一致程度. κ 度量定义为

$$\kappa = 1 - \frac{\frac{1}{L} \sum_{j=1}^N l(z_j)(L - l(z_j))}{N(L - 1)\bar{p}(1 - \bar{p})}, \quad (8)$$

其中 \bar{p} 为基分类器的平均分类精度, 有

$$\bar{p} = \frac{1}{NL} \sum_{j=1}^N \sum_{i=1}^L y_{j,i}, \quad (9)$$

y_{ji} 为基分类器 C_i 对样例 z_j 的分类结果. 研究表明, κ 与 KW 方差和不一致度量 dis 是相关的, 它们之间的关系^[15]如下所示:

$$\kappa = 1 - \frac{L}{(L - 1)\bar{p}(1 - \bar{p})} KW = 1 - \frac{1}{2\bar{p}(1 - \bar{p})} Dis_{av}, \quad (10)$$

其中 Dis_{av} 为集成系统的不一致度量平均值.

3) 熵度量 E .

本文的熵度量 E 没有使用对数函数, 所以与 Cunningham 等^[16]提出的经典熵度量有所不同, 但其更易于处理且能更快速计算. L 个基分类器集合的熵度量^[15]定义为

$$E = \frac{1}{N} \sum_{j=1}^N \frac{1}{(L - \lceil L/2 \rceil)} \min\{l(z_j), L - l(z_j)\}. \quad (11)$$

对于每一个样例 z_j , 如果所有的分类器给出同样的分类结果, 即它们之间不存在任何多样性, 则此时熵度量的值为 0; 相反, 如果 $\lceil L/2 \rceil$ 个分类器对该样例正确分类, $L - \lceil L/2 \rceil$ 的分类器错误分类, 则熵度量的值为 1, 此时集成系统的多样性程度最高.

4) 难度度量 θ .

难度度量的思想来源于文献 [6], 该文献讨论了样例集中样例的分类“困难”模式. 对于一个从问题分布中随机抽取的样例 x , 用随机变量 X 表示对该样例正确分类的分类器所占的比例, 其取值范围为 $\{0, 1/L, 2/L, \dots, 1\}$. 难度度量^[6]定义为随机变量 X 的方差, 即

$$\theta = \text{Var}(X). \quad (12)$$

该方差的值越小, 随机变量 X 的波动越小, 表明对于每一个样例, 将其正确分类的分类器数目越趋向于 $L/2$, 此时一些分类器难以对样例正确分类, 而另外的分类器容易对它正确分类, 分类器之间的多样性程度越高; 该方差的值越大, 表明一部分样例对于所有分类器而言是难以正确分类的, 而剩余样例对所有分类器而言是容易分类的, 此时各个分类器的行为越相似, 多样性程度越低.

5) 广义多样性 GD.

广义多样性由 Partridge 等^[17]提出, 该度量用随机变量 Y 表示在一个随机抽取的样例 x 上分类错误的分类器比例; 用 p_i 表示 $Y = i/L$ 的概率, $p_i = P\{Y = i/L\}$; 用 $p(i)$ 表示 i 个随机选择的分类器在一个随机抽取的样例上分类错误的概率.

假设随机选择两个分类器, Partridge 等认为, 当一个分类器正确分类、另一个分类器错误分类时, 它们的多样性程度最高, 两个分类器同时错误分类的概率 $p(2) = 0$; 当一个分类器的错误分类总是伴随着另一个分类器的错误分类时, 多样性程度最低, 两个分类器均失败的概率等于一个随机选择的分类器失败的概率. 给出如下公式:

$$p(1) = \sum_{i=1}^L \frac{i}{L} p_i, \quad p(2) = \sum_{i=1}^L \frac{i(i-1)}{L(L-1)} p_i. \quad (13)$$

则广义多样性^[17]定义为

$$GD = 1 - p(2)/p(1). \quad (14)$$

根据以上分析, 若 $p(1) = p(2)$, 则 $GD = 0$, 集成系统多样性程度最低; 若 $p(2) = 0$, 则 $GD = 1$, 集成系统的多样性程度最高.

6) Percentage Correct Diversity Measure(PCDM).

Banfield 等^[18]提出了 PCDM 度量, 对于每个样例, 该度量关注将其正确分类的分类器所占的比例, 计算过程如下所示:

```
Tally = 0
For each example
  If  $T_{low} \leq \text{Classifiers Correct} \leq T_{high}$ 
    Tally = Tally + 1
  End If
End For
PCDM = Tally/N.
```

由以上过程可知, PCDM 度量关注这样的样例: 有 $T_{low} \sim T_{high}$ 个基分类器对其正确分类, 其中 T_{low} 和 T_{high} 是两个阈值. 即对于一个样例 x , 如果在上述范围内的一定比例的分类器对其正确分类, 则至少对该样例而言, 这些分类器被认为是多样化的. 这样, 所有分类器均对其正确分类(正确率高于 T_{high} , 即容易

被正确分类的样例), 或很少分类器对其正确分类(正确率低于 T_{low} , 即难以被正确分类的样例)的样例在该多样性度量下被认为是无用的。

1.3 多样性度量总结

前文介绍的 10 种常用的多样性度量是由不同学者从不同领域、不同角度提出的, 每种方法评估和看待分类器间多样性的方式各不相同, 因此, 难以将它们放在一个框架下进行透彻的比较。目前, 还不存在一种优于任何其他方法的多样性度量^[8]。下面从每种方法所来自的领域或主要思想、理论上与该方法存在联系的其他方法这两个主要方面, 对上述 10 种多样性度量方法进行归纳总结, 得到表 2。表 2 中, 前 4 种成对度量方法均是通过统计满足一定条件的样例来评估分类器间的多样性程度; 后 6 种非成对度量方法均是通过考察满足一定条件的分类器来衡量集成系统的多样性。

表 2 各种多样性度量归纳

名称	来自领域(主要思想)	与其存在联系的方法
Q	统计学	ρ , 式(4)
ρ	统计学	Q , 式(4)
dis	对分类器分类结果不一致的评估	κ , 式(10)
DF	对分类器错误分类结果一致性的评估	-
KW	分类器误差的分解	κ , 式(10)
κ	统计学	KW, dis, 式(10)
E	条件熵	-
θ	对随机样例分类难度的方差评估	-
GD	对两个随机选择分类器的行为比较	-
PCDM	对每个样例分类难易的评估	-

二次失败度量 DF 作为多样性度量方法应用于图像分类问题中^[12]; 4 种成对度量方法均在数字分类问题中得到了应用^[15]; Giacinto 等^[19]使用成对多样性度量(二次失败度量 DF 和 Q 统计)计算每对分类器之间的多样性。但在实际应用中, 如何针对具体问题选择使用合适的多样性度量, 目前还没有相应的理论进行说明。本文认为, 需要根据具体问题中是侧重评估分类器之间的多样性还是集成系统整体的多样性、每种度量方法计算或实现的难易程度、具体问题中所采用的分类器集成方式等方面进行选择。如, 在进行选择性集成时, 有时需要计算每对分类器间的多样性, 此时应从 4 种成对多样性度量中进行选择, 如文献[19]选择使用了 DF 度量和 Q 统计; 有时需要计算不考虑某个分类器 $C_i (1 \leq i \leq L)$ 时剩余分类器集合 $C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_L$ 的多样性, 此时最好使用非成对度量方法来评估系统的整体多样性程度(如文献[20]使用的是 κ 度量)。文献[15]发现, 二次失败度量 DF 和难度度量 θ 与多数投票集成、朴素贝叶斯集成方式的精度有较强的关系, 因此在这两种集成方式下最好选择 DF 或 θ 。

2 多样性度量研究进展

本节对近年来多样性度量的研究进展进行分析探讨。目前, 关于多样性度量的研究, 主要集中在 3 个方面: 1) 从新的角度提出解释和度量多样性的方法; 2) 如何利用多样性度量更好地选择分类器构建集成系统, 以提高集成学习的性能; 3) 探讨多样性度量与集成学习精度的关系。

2.1 新的多样性解释和度量

为了更好地认识多样性, 学者们不断地从新的角度解释多样性, 并提出了一些新的多样性度量方法。如从关系熵、信息理论、多样性正则化、集成误差分解、margin 最大化和异质性等角度来解释多样性, 并提出了针对元组输出形式、利用无标签数据、基于多样性和分类器精度的新的多样性度量方法。

2.1.1 针对元组输出形式的多样性度量

对于分类器的输出形式, 除了前文介绍的 oracle 输出外, 还有以下两种输出形式: 1) c 维元组形式 $D_i(x) = [d_{i,1}(x), d_{i,2}(x), \dots, d_{i,c}(x)]$, 其中 $d_{i,j}(x)$ 为第 i 个分类器 D_i 给出的样例 x 属于第 j 个类的置信度, $j = 1, 2, \dots, c$, c 为不同类标签的数目; 2) 某一明确的类标签 $D_i(x) \in \Omega$, Ω 为类标签集合, 对于每个样例 x , 分类器 D_i 的输出为 Ω 中一个明确的类标签。

文献[21]认为, 针对分类器的 c 维元组输出形式, 在将 c 维元组形式转换为 oracle 形式的过程中会丢失一些分类信息。如, 两个分类器的 4 维元组输出结果 $D_i(x_k) = [0.1, 0.1, 0.8, 0]$ 和 $D_j(x_k) = [0.2, 0.25, 0.3, 0.25]$ 分别为分类器 D_i 和 D_j 对样例 x_k 的分类结果。将它们转换为 oracle 形式时, 根据最大成员规则, 两个分类器对 x_k 的分类结果均为第 3 类 c_3 , 则其 oracle 输出也是相同的。但是两个元组之间较大的差别, 而转换得到的相同的 oracle 输出结果不能精确反映两个分类器提供信息之间的差异。基于这一点, 文献[21]提出了一种新的多样性度量, 以计算输出为 c 维元组形式时分类器之间的多样性。

2.1.2 利用无标签数据度量多样性

大多数常用的多样性度量根据分类器在有标签样例上的分类结果计算多样性, 但有时这种有标签的样例较为缺乏, 相比而言, 无标签样例更丰富且更容易获得。鉴于此, 文献[22]提出了一种利用无标签样例的数据分布信息计算多样性的方法, 并通过实验表明了基于该多样性度量的选择性集成具有良好的性能。然而, 该方法只考虑了连续输出形式下的二分类问题, 下一步还需要解决离散输出形式的情况和多分类问题。

另外, 文献[23]提出了一种利用无标签数据增强集成系统多样性的半监督集成学习方法 UDEED。与

现有的利用无标签数据扩大有标签数据规模的半监督集成方法不同, 该方法最大化基分类器在有标签数据上的精度, 同时最大化分类器之间在无标签数据上的多样性. UDEED 能够通过增强多样性有效地利用无标签数据进行集成学习, 实验数据表明, UDEED 能够达到与其他成功的半监督集成方法相当的性能, 然而对于 UDEED 为什么具有良好的性能, 还需要从理论上进一步分析.

2.1.3 多样性度量与基分类器精度相结合

文献[24]认为, 现有多样性度量应该与某种精确性度量相结合, 以更好地构建集成系统, 并基于此提出了一种同时考虑多样性和精确性的多样性度量. 该度量既以成对的形式计算每对分类器间的多样性, 又计算了所有分类器的多样性. 将其应用到一个基于遗传算法 GA 的集成方法中, 实验结果表明了该多样性度量的有效性.

文献[25]探讨了分类器精度和集成多样性度量(双次失败度量、难度度量)的组合在估计集成系统精度上的效果. 实验表明, 与使用单个度量(如集成多样性度量或分类器精度度量)相比, 该组合度量通常具有更好的性能. 然而, 文献[25]只考虑了两种多样性度量, 对于其他多样性度量组合的情形值得进一步探讨. 另外, 该研究将多样性度量与精度度量线性组合, 考虑给每个度量赋予一个权值, 然后用两个度量的加权和对选择分类器子集进行指导是很好的改进思想. 文献[26]给出了复合多样性函数, 将分类器间的多样性和每个基分类器的分类精度结合起来, 表明了该复合函数与集成精度之间具有较强的相关性, 并通过统计结果表明该复合多样性函数要优于传统的多样性度量, 因此可以用于选择更好的集成系统.

2.1.4 从关系熵和信息理论的角度解释多样性

文献[27]提出了一种基于关系熵的多样性度量, 该方法计算用一个关系矩阵表示的信息熵, 如果关系熵是基于 Q 统计或相关系数 ρ , 则该信息量反映了基分类器间的多样性. 关系熵会随着集成系统多样性程度的提高而增加, 实验表明所提出的关系熵是多分类器系统一个很好的多样性度量. 熵是最好的度量不确定性的概念, 文献[28]给出了一个类似 Q 统计的基于熵的成对多样性度量, 该度量能够充分利用熵在度量分类器不确定性上的优势, 将该度量与 Q 统计和 κ 度量进行了比较, 结果表明所提出的度量与系统精度提高具有更强的相关性.

文献[29]从信息理论的角度解释多样性, 信息理论标准的一个主要优势是它能够捕获到数据的高阶统计信息. 讨论了互信息和分类器误差之间的关系, 并说明集成互信息可自然地扩展得到多样性和精确性两项. 发现集成系统多样性实际上以多阶相关性的

形式存在, 而成对多样性度量只能捕获到低阶成分, 这对集成系统多样性的理解迈进了重要的一步. 文献中给出了用多数投票作为组合方式时的初步结论, 进一步的扩展工作包括评价不同组合方式对集成系统包含的信息进行“解码”的有效性.

受文献[29]的启发, 文献[30]也从信息理论的角度研究集成系统的多样性, 认为文献[29]中的信息理论多样性及其项的表达式非常复杂, 并且没有提出有效的过程来估计现实中的多阶相关性. 鉴于此, 给出了一种新的描述, 使集成多样性及其项的表示更加简单. 在此基础上, 给出了一个近似形式对现实中的集成系统多样性进行估计, 实验表明了所给出的表示形式和近似形式的有效性.

2.1.5 多样性正则化

文献[31]提出了多样性正则机(DRM), 有效地生成了一个由多样化的支持向量机(SVM)构成的集成. 理论分析表明, 在 DRM 中使用的多样性约束可有效降低其假设空间复杂性, 这意味着在集成方法中的多样性控制确实发挥着正则化的作用. 实验表明, DRM 能够极大地提高系统的泛化能力, 并优于一些先进的 SVM 集成方法.

文献[32]在 PAC 学习框架内从理论上分析了多样性对于投票集成泛化性能的影响, 这是在投票集成中对多样性作用进行的第一次 PAC 形式的分析. 该分析表明, 通过增强多样性, 可以降低投票集成的假设空间复杂性, 从而可以期望得到更好的泛化性能. 从统计学习的角度, 这意味着可以将增强多样性看作在集成方法中应用正则化, 这可能在集成学习中引入看待多样性的一个新角度. 受这一结果的启发, 文中提出了一个显式采用多样性正则化的贪婪集成细化方法. 实验结果表明, 该细化方法能够在减小集成规模的情况下显著提高泛化性能. 当前的工作只是将多样性正则化用于贪婪集成细化, 开发显式采用多样性正则化的集成学习算法将是下一步的工作.

2.1.6 从集成误差分解的角度解释多样性

文献[33]从集成分类误差精确分解的角度解释多样性, 对集成分类误差进行了类似回归问题的 bias-variance-covariance 分解和 ambiguity 分解, 并讨论了分解结果是否能为多样性提供一个更明确的定义以及能否更有效地用于构建集成系统.

文献[34]认为, 多样性度量是从误差函数和组合函数这两个直接因素中自然衍生出来的, 从这一角度给出了使用多数投票作为组合函数时对集成分类误差的一个分解. 分解得到 3 项: 个体分类器的平均精度、“好的多样性”和“差的多样性”, 并说明增大“好的多样性”的值会减小多数投票分类误差, 而增

大“差的多样性”的值会提高分类误差. 最后展示了所提出的分解如何更好地了解多数投票分类器集成.

2.1.7 从 margin 的角度评价多样性度量

为了解释 boosting 集成学习算法, Schapire 等^[35]将 margin 的概念引入集成学习领域. 给定 L 个基分类器及其对应的权值 $w = [w_1, w_2, \dots, w_L]^T$, 其中 $w_j \geq 0$ 且 $\sum_j w_j = 1$. 该集成系统在样例 x_i 上的 margin 定义为 $m_i = \sum_{j=1}^L w_j O_{ij}$, 其中 O_{ij} 为第 j 个分类器在样例 x_i 上的输出结果.

margin 概念与集成学习算法的成功具有更明确的关系. 一些研究表明, 一个集成系统的泛化性能与其在训练样例集上 margin 的分布相关, Schapire 等^[35]已证明, 在训练样例集上增大 margin 会改善集成系统的泛化误差边界. 基于此, 文献[36]从 margin 的角度评价多样性度量, 对 6 个现有的多样性度量进行理论分析, 说明了这些度量之间的内在关系, 并将这些度量与 margin 的概念关联起来. 另外, 给出了同时最大化集成系统多样性和集成系统最小 margin 的一致条件, 该分析可以对多样性概念有更深刻的理解.

2.1.8 多样性与异质集成

集成学习大多是关于同质集成的研究, 即利用同一个分类算法生成所有分类器, 只是在生成每个分类器时提供给该分类算法的参数(如训练数据集、属性子集、初始权值等)有所不同. 关于分类器间多样性的研究也主要是在同质集成中进行的, 而在异质集成中对多样性探讨的不多. 异质集成的主要思想是使用不同类型的分类算法生成一些分类器来构成集成系统.

文献[37]对于由训练数据集的不同导致的分类器间的多样性和由分类算法的不同导致的多样性都给出了形式化表示, 并从理论上说明异质性与多样性增益之间存在概率关系. 另外, 文中还给出了一个利用非线性函数来描述多样性与相关性之间的关系, 其中相关性 with 集成系统精度的关系已有研究. 对于在其他机器学习问题(如多类别分类问题)中使用异质分类算法的情况, 还需要进一步研究.

文献[38]从理论和实验上研究了异质性与多样性之间的关系, 为在集成系统中使用异质分类器奠定了基础. 较少文献将同质分类器集成的多样性与异质分类器集成的多样性进行理论和实验比较, 该文献对它们进行了比较并发现异质分类器之间的多样性程度要高于同质分类器间的多样性, 因此使用异质分类器构建集成系统可以提高多样性程度. 下一步需要对异质性与集成系统精度的关系进行研究.

2.2 多样性度量用于选择性集成

如何选择基分类器对集成系统的泛化性能具有

重要的影响. 现实中, 当生成很多基分类器时, 不一定将所有的分类器都融合到集成系统中. 因为有些分类器的精度较差, 或者有些分类器的分类结果相似, 即它们不具有多样性. 这些分类器参与集成将不会提高集成系统的泛化性能. 因此, 需要将低精度或冗余的分类器移除, 这样能够减小系统的规模且不会损失甚至提高集成系统的精度. 这一过程称为集成剪枝、集成细化或选择性集成^[39-41].

选择性集成的目的是在减少集成系统中分类器的同时保持甚至提高系统的预测性能, 从而减少存储和计算开销, 提高预测速度和精度. 多样性度量可用于指导构建有效的分类器集成. 一些学者将多样性度量与其他方法结合起来进行选择集成, 这些方法包括聚类算法、进化算法、排序方法等.

2.2.1 多样性度量与聚类算法结合

多样性度量可以与聚类算法相结合来选择分类器^[12,19,42-43], 这类方法分为两个过程: 1) 簇的生成, 将分类器之间的多样性值看作它们之间的距离, 然后使用某一聚类算法将所有分类器进行划分, 从而得到若干个簇; 2) 簇内选择, 根据某一标准从每个簇中选择预定数目的分类器来构成最终的集成系统.

文献[12]提出了一种基于层次聚类的神经网络选择性集成方法, 基于两个神经网络的复合错误概率计算它们之间的多样性值, 并以此作为它们之间的距离. 两个神经网络作出的一致分类错误越少, 它们之间的距离越大; 反之, 距离越小. 对于聚类算法的每一次迭代, 从每个簇中选择一个离簇中心最近的分类器构成集成系统, 这样每次迭代对应一个集成系统. 最后根据各个集成系统在验证集上的精度, 选择精度最好的作为最终的集成系统.

文献[42-43]均提出了一种基于 k -means 聚类的选择性神经网络集成方法. 首先用所有分类器对验证集进行分类, 从而得到分类结果矩阵, 矩阵中每一列对应一个分类器的分类结果. 然后使用 k -means 聚类算法将矩阵中的所有列向量进行聚类, 从而得到若干个簇. 在聚类过程中, 将两个列向量(代表两个分类器)之间的多样性值看作它们之间的距离, 文献[43]考虑的是回归问题, 使用欧氏距离度量神经网络输出之间的多样性. 最后根据从每个簇中选择一个精度最高的分类器构成最终的集成系统. 由于不同簇中的分类器差别较大, 这样便保证了所选分类器之间的多样性.

文献[19]使用成对多样性度量(双失败度量和 Q 统计)计算每对分类器之间的多样性, 并得到一个成对多样性矩阵, 然后将矩阵中的值看作距离, 使用一个聚类算法对分类器进行聚类. 得到若干个簇后, 从每个簇中随机选一个分类器构造集成系统. 这种方

法也保证了选择的分类器是多样化的。

对于通过聚类算法增强分类器间多样性的方法,如基于 k -means 聚类的选择性集成^[42-43], 还有以下问题需要更好地解决: 1) 对生成的分类器集合进行聚类时, 将它们聚成几个簇主要由人工确定, 因此, 如何更好地利用启发信息确定得到的簇数需要进一步研究; 2) 进行簇内选择时, 文献[43]从每个簇中选择一个精度最高的分类器构成最终的集成, 而文献[12]从每个簇中选择一个离簇中心最近的分类器, 因此, 对不同的簇内选择方法进行理论和实验比较, 以确定更高效的簇内选择方法是值得研究的问题。

2.2.2 多样性度量与进化算法结合

多样性度量也可与进化算法相结合来进行选择性集成^[24,32,44-46]。这类方法将选择性集成问题看作一个优化问题, 试图找到这样一个分类器子集, 该子集最优化某一与集成系统泛化性能相关的目标函数。

在利用遗传算法选择分类器时^[24,44], 从一个初始染色体种群开始, 每个染色体表示一个要进行组合的基分类器集合, 染色体中每一位 (bit) 代表一个分类器。在每次迭代中, 用多样性度量评价染色体的适应度, 然后根据每个染色体的适应度进行轮盘赌选择、单点交叉、二进制突变等遗传操作。当达到某个结束标准时停止迭代过程, 从而得到一个近似最优的分类器集合。适应度函数也可以采用 $Fitness_i = acc_i + \alpha \times div_i$ 的形式, 它考虑了基分类器的多样性和精度的影响。使用遗传算法选择分类器的方法计算代价较大, 需要较高的时间复杂度。

文献[32]研究了多样性对于集成系统泛化性能的影响, 将显式多样性正则化用于贪婪选择性集成。它基于这样的分析: 通过理论分析得出多样性与假设空间复杂度是紧密相关的, 增强多样性可以看作在集成方法中应用正则化, 这提供了一种看待集成系统中多样性的新视角。基于以上分析, 将显式多样性正则化应用于集成细化中, 并提出了一种集成细化方法 DREP。实验结果表明, 由于进行了多样性正则化, DREP 方法能够在减小集成规模的情况下得到更好的泛化性能, 因此, 开发显式的利用多样性正则化的集成学习方法是一个很好的方向。

文献[45]认为, 在选择分类器时, 使用简单贪婪搜索作为优化方法的现有启发式方法缺乏理论或实证保证, 并提出将分类器子集的选择问题转化为一个严格的二次整数规划问题, 使用 SDP (semi-definite programming) 作为解决技术能够得到更好的近似解。计算实验表明, 基于 SDP 的选择算法优于其他两种基于度量的算法。下一步需要探讨是否存在目标函数的一个良好形式, 以便在不执行基数约束的情况下能够找到“真正”的最优子集。

文献[46]给出了基于一个新多样性度量的贪婪集成选择方法 Focused Ensemble Selection (FES)。该方法的思想是将分类器非常容易或困难地进行分类的样例忽略掉, 只关注靠近改变状态的样例。文中提出的考虑当前集成系统分类结果优势的思想对于从事选择性集成的研究者具有重要的价值。

基于进化算法的选择性集成方法中涉及非线性优化过程, 该类方法一般经过很多次迭代后才能得到最终的近似最优解, 得到的集成系统大都具有较好的泛化性能。然而, 这类方法一般时间复杂度较高, 不适合应用于对实时响应要求较高的场合。

2.2.3 多样性度量与排序方法结合

这类方法的主要思想如下: 首先用某种多样性度量计算分类器间的多样性或分类器系统的多样性; 然后根据分类器间的多样性或每个分类器对集成系统多样性的贡献程度对所有分类器进行排序, 根据排序从中选择一部分分类器构成集成系统。

文献[41]给出了第1个选择性集成方法 Kappa pruning。该方法首先在训练集上用 Kappa 统计^[41]度量每对分类器之间的多样性; 然后基于其多样性值按从小到大的顺序对所有分类器对进行排序; 最后根据该次序选择前面的一部分分类器构成集成系统。该方法是基于排序的方法, 它试图选择多样化程度最高的分类器子集。

文献[47]提出了一种基于分类器精度和多样性的选择性集成方法 SAD。首先生成一些基分类器, 根据它们在验证集上的精度选择若干个精度最高的分类器 C_1, C_2, \dots, C_n ; 然后计算精度最高的分类器和没被选择的各个分类器之间的 Q 统计值, 并将它们从小到大排序; 最后从对应的 Q 统计值最小的分类器开始, 选择它到集成系统中, 重复该选择过程直到满足预定的集成规模。由于两个分类器之间的 Q 统计值越小, 它们之间的多样性程度越高, 这样便保证了所选分类器的多样性。

Banfield 等^[20]介绍了4种选择性集成算法, 分别为 accuracy in diversity thinning (AID)、concurrency thinning、kappa thinning 和 sequential backwards selection thinning (SBS), 它们都是基于逐步去除性能较差的分类器的方法。其中在 kappa thinning 方法中, 对每个分类器 $C_i (1 \leq i \leq L)$ 进行如下操作: 计算不考虑该分类器时剩余分类器集合 $C_1, C_2, \dots, C_{i-1}, C_{i+1}, \dots, C_L$ 的 κ 多样性值, 这样每个分类器对应一个分类器集合的多样性值, 找出对应的多样性值最大的分类器并将其去除。直观上, 该分类器的去除对集成系统的多样性贡献最大, 重复执行以上过程直至集成规模达到某个阈值。与 kappa thinning 方法不同, SBS 关注不考虑某个分类器时剩余分类器集合的精

度,该方法将使剩余分类器集合精度最大的分类器去除.以上几个算法可用于组合不同类型的分类器,如决策树、神经网络等.

2.3 多样性度量和集成精度

第 2.2 节探讨了多样性度量在选择性集成中的一些应用,通过显式地使用多样性度量对选择分类器进行指导,从而构造有效的集成系统,以达到减少存储、计算开销,提高预测速度和精度的目的.

1) 多样性在多分类器系统的设计中发挥着重要作用,但在理论上,多样性与集成系统精度的关系仍然不十分明确^[48]. 学者们探讨了多样性度量与基分类器平均精度的权衡是否与集成精度相关联,但没有发现它们之间存在明确的相关性,并指出显式使用多样性度量的有用性仍存在争议.

在文献[8]中, Kuncheva 的研究表明设计多样化分类器的总体动机是正确的,但显式地度量该多样性并有效地使用它来构建更好的集成系统仍然是有待解决的问题. 另外还指出,尽管在某些特殊情况下,多样性与集成精度之间存在明确的关系,但实验结果对多样性度量在构建集成系统中的有用性提出了质疑.

文献[15]研究了各种多样性度量与 10 种分类器集成方式之间的关系. 在标准数据集上进行了一系列实验,发现只有双次失败度量 DF 和难度度量 θ 与多数投票、朴素贝叶斯集成方式的精度有较强的关系,而其他多样性度量方法与考虑的 10 种分类器集成方式之间没有这种关系. 由该结论可知,在多数投票和朴素贝叶斯这两种集成方式下最好选择使用 DF 度量或 θ 度量.

文献[15]认为,如果多样性度量和分类器集成方式之间的相关性不明确,则在设计集成系统时使用多样性度量的有效性是一个尚未解决的问题. 另外,对于某一分类器集成方式而言(如简单多数投票),直接计算集成精度比先计算分类器间的多样性,再试图用该多样性预测集成精度更有意义. 尽管在一些时候更容易计算多样性的值,但理论上多样性与集成精度之间模糊的关系使得优化多样性这一方法并不乐观. 应该找到对多样性概念一个更精确的描述,从而定义更实用的度量并对其进行优化以构造更好的集成系统.

文献[18,20]使用多样性度量进行选择集成,尽管最终获得的集成系统的规模比原始系统要小,但其精度没有得到任何明显的提高. 用于构建集成系统的“过度生产再选择”方法显式地使用多样性度量,它们不能提供证据以表明该方法比直接估计集成精度更有效^[12,18,20,26,31].

尽管现有的多样性度量都反应了多样性的直观概念,但是没有一个度量来自对集成分类误差的精

确分解,没有一个度量展示出与成员分类器平均精度的一个明确权衡以确定集成系统的精度^[8,36]. 文献[33]对集成分类误差进行了类似回归问题的 bias-variance-covariance 分解和 ambiguity 分解,指出现有的常用多样性度量最多是对分解得到的“真实多样性项”的近似,并认为多样性度量的有用性仍然存在争议.

由于 margin 的概念与集成学习算法的成功具有更明确的关系,文献[36]从 margin 的角度探讨了多样性度量与集成精度的关系,展示了 6 种多样性度量与 margin 最大化概念之间的关系,给出了同时最大化集成的多样性和集成最小 margin 的一致条件,并从理论和实验上表明了这 6 种多样性度量不能有效地构建具有良好泛化性能的集成系统. 该文献的分析强调了分类问题中不同理论之间的关系,这有助于设计更好的集成学习算法. 文献[49]探讨了多样性与 margin 理论的关系,发现多样性度量实际上不足以对集成系统总体性能的提高进行评估,且这些度量与集成成员之间的方差具有更紧密的联系. 这一事实可以解释为什么显式地追求多样性不能有效地构建具有最优泛化性能的集成系统.

文献[50]给出了一个基于线性程序的框架来分析分类器集成中的多样性度量. 用该框架分析多样性度量与集成精度之间的关系,分析结果对多样性度量的有用性提出了质疑. 基于这一结果,指出尽管有些多样性度量在它们被提出的领域中可能是有用的,但在分类器集成中将它们作为多样性度量使用时要格外注意.

Kuncheva^[48]认为,由于理论上多样性与集成精度之间不明确的关系,对直接使用多样性构建集成系统的研究开展得不多. 如果多样性与集成精度之间没有明确的关系,则没有必要麻烦地将多样性引入到集成构建中. 然而, Kuncheva 指出,虽然还没有统一的多样性定义,但从直觉上讲,多样性是有用的,研究工作一旦取得突破,便会出现一个理论来解释应该在什么时候、什么场合以及如何才能最好地使用多样性.

2) 针对多样性与集成精度在理论上不明确的关系,一些学者从不同方面对其进行了改善,包括从新的角度研究多样性与集成精度之间的关系、提出各种方法试图找到分类器的多样性与精度之间一个较好的权衡、将现有多样性度量与某种精度度量相结合以构建有效的集成系统等.

文献[51]从相关性 correlation 的角度探讨多样性与集成精度的关系. 由于相关性与集成精度的关系是相对明确的,如, Tumer 等^[52]在简单平均集成中为分类器间的多样性和相关性构建了一个关系,表明简

单平均集成可降低的误差可以表示为相关性的一个函数, 并将训练样例数目与降低相关性的效果关联起来; Breiman^[53]将相关性与随机森林(random forests)的集成精度关联起来, 认为集成泛化误差的边界由分类器集合的平均相关性和基分类器精度决定. 基于此, 文献[51]进行了理论和实验分析, 给出了一个非线性函数将多样性和相关性联系起来, 从而为多样性和集成精度的关系提供了一个代理.

文献[26]提出基分类器的多样性与精度之间存在一个权衡 trade-off, 通过理论分析表明了具有良好泛化性能的分类器集合需要同时具有多样性和精确性, 简要介绍了著名的集成学习方法——负相关集成学习(NCL)^[54]. NCL直接利用模糊分解对成员神经网络的权值进行更新, 通过这种方式试图达到多样性与精度的较好的权衡. 文中对NCL进行了扩展并应用于分类问题中, 同时也介绍了多目标进化集成学习作为达到多样性和精度权衡的一种策略, 其基本思想是将多样性和精度作为两个优化目标, 通过不断向 Pareto 最优移动试图使进化过程搜索到一个较好的权衡.

文献[24]提出了一种同时考虑分类器多样性和精确性的多样性度量, 并将其应用到基于遗传算法 GA 的集成方法中, 实验结果表明了该度量的有效性. 文献[55-56]均提出了一个将分类器间的多样性和分类器的精度组合起来的复合多样性函数, 文献[55]指出用 Random Subspaces 作为集成学习方法时, 提出的复合多样性函数 CDF 与集成精度之间存在较强的相关性. 统计结果表明, 该复合多样性函数要优于传统的多样性度量, 然而该工作只是在 Random Subspaces 集成方法下进行的实验. 文献[56]对其工作进行了扩展, 考虑了除 Random Subspaces 以外的多种集成学习方法, 并得到同样的结论, 即所给出的复合多样性函数要优于传统的多样性度量, 从而可以构建更好的集成系统.

总之, 在分类器集成中增强多样性的总体动机是正确的, 然而在理论上多样性与集成精度的关系还不明确, 显式地度量分类器间的多样性并使用它来构建集成系统的有用性仍存在争议. 因此, 在这两方面还需要进行更多的理论研究.

3 结 论

本文首先介绍了10种常用的多样性度量方法, 对每种方法看待、度量多样性的角度和基本思想进行了详细的分析阐述, 并给出了实际应用中如何选择使用多样性度量的建议. 从3个方面总结探讨了多样性度量的研究进展, 包括从新的角度解释和度量多样性、多样性度量在选择分类器构建集成系统中的应

用、多样性度量与集成学习精度的关系. 进一步的工作包括:

1) 在构建集成系统时, 显式使用现有多多样性度量的有用性仍存在质疑, 需要进行更多的理论和实验研究来进一步探讨多样性与集成精度的关系, 以及显式使用多样性度量的效果. 另外, 在第2.1节中总结的从新的角度提出的多样性度量, 它们在实验中都表现出良好的性能. 因此, 对这些度量方法及其与集成精度的关系进行更多的理论分析和实验比较是一项有意义的工作.

2) 将现有多多样性度量与集成系统误差分解得到的多样性项进行比较, 有助于理解哪个度量是对“真实多样性”更好的近似, 甚至能启发提出新的、更好的接近真实多样性的度量.

3) 相对于有标签样例而言, 现实中无标签样例更丰富且更容易得到. 对于连续输出形式下的二分类问题, 已有利用无标签数据度量多样性的方法, 然而对于离散输出形式及多分类问题, 还需要进一步探讨. 另外, 设计通过有效地利用无标签样例来增强集成系统中多样性的新的集成学习方法是一个有意义的方向.

4) 理论研究发现, 增强多样性可以看作在集成学习方法中应用正则化, 将显式多样性正则化用于选择性集成可以得到较好的效果. 因此, 开发显式利用多样性正则化的集成学习方法是一个较好的方向. 另外, 学者们发现集成多样性以多阶相关性的形式存在, 因此, 对于以上多样性理论成果在这种情况下是否仍然成立值得进一步研究.

5) 实际使用过程中, 针对具体问题如何选择使用合适的多样性度量, 目前还没有相应的理论进行说明. 本文认为, 要根据具体问题中是侧重评估分类器之间的多样性还是系统整体的多样性、每种度量方法计算或实现的难易程度、具体问题采用的分类器集成方式来进行选择. 如何更好地确定多样性度量的选择标准需要更多的理论和实验研究.

6) 除了显式使用多样性度量, 研究像 Bagging、AdaBoost 和 Random Subspace 这样以隐式方式产生分类器间多样性的集成学习方法也是有价值的方向.

参考文献(References)

- [1] Dietterich T G. Machine learning research: Four current directions[J]. AI Magazine, 1997, 18(4): 97-136.
- [2] Breiman L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140.
- [3] Schapire R E. The strength of weak learnability[J]. Machine Learning, 1990, 5(2): 197-227.
- [4] Ho T K. The random subspace method for constructing

- decision forests[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844.
- [5] Rodriguez J J, Kuncheva L I, Alonso C J. Rotation forest: A new classifier ensemble method[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2006, 28(10): 1619-1630.
- [6] Hansen L K, Salamon P. Neural network ensembles[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993-1001.
- [7] Dietterich T G. *Ensemble learning*[M]. Cambridge: The MIT Press, 2002: 405-408.
- [8] Kuncheva L I, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. *Machine Learning*, 2003, 51(2): 181-207.
- [9] Yule G U. On the association of attributes in statistics[J]. *Philosophical Trans of the Royal Society of London*, 1900, 194: 257-319.
- [10] Skalak D B. The sources of increased accuracy for two proposed boosting algorithms[C]. *Proc of AAAI-96 Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms*. California: AAAI, 1996: 120-125.
- [11] Ho T K. The random space method for constructing decision forests[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844.
- [12] Giacinto G, Roli F. Design of effective neural network ensembles for image classification purposes[J]. *Image and Vision Computing*, 2001, 19(9/10): 699-707.
- [13] Kohavi R, Wolpert D H. Bias plus variance decomposition for zero-one loss functions[C]. *Proc of the 13th Int Conf on Machine Learning*. San Francisco: Morgan Kaufmann Publishers, 1996: 275-283.
- [14] Fleiss J L. *Statistical methods for rates and proportions*[M]. New York: John Wiley and Sons, 1981: 1644-1645.
- [15] Shipp C A, Kuncheva L I. Relationships between combination methods and measures of diversity in combining classifiers[J]. *Information Fusion*, 2002, 3(2): 135-148.
- [16] Cunningham P, Carney J. Diversity versus quality in classification ensembles based on feature selection[C]. *Proc of the 11th European Conf on Machine Learning*. Berlin: Springer-Verlag, 2000: 109-116.
- [17] Partridge D, Krzanowski W. Software diversity: Practical statistics for its measurement and exploitation[J]. *Information and Software Technology*, 1997, 39(10): 707-717.
- [18] Banfield R E, Hall L O, Bowyer K W. A new ensemble diversity measure applied to thinning ensembles[C]. *Proc of the 4th Int Workshop on Multiple Classifier Systems*. Berlin: Springer-Verlag, 2003: 306-316.
- [19] Giacinto G, Roli F. An approach to the automatic design of multiple classifier systems[J]. *Pattern Recognition Letters*, 2001, 22(1): 25-33.
- [20] Banfield R E, Hall L O, Bowyer K W, et al. Ensemble diversity measures and their application to thinning[J]. *Information Fusion*, 2005, 6(1): 49-62.
- [21] Fan T G, Zhu Y, Chen J M. A new measure of classifier diversity in multiple classifier system[C]. *Proc of the 7th Int Conf on Machine Learning and Cybernetics*. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2008: 18-21.
- [22] Chu R, Wang M, Zeng X Q, et al. A new diverse measure in ensemble learning using unlabeled data[C]. *Proc of the 4th Int Conf on Computational Intelligence, Communication Systems and Networks*. Washington DC: IEEE Computer Society, 2012: 18-21.
- [23] Zhang M L, Zhou Z H. Exploiting unlabeled data to enhance ensemble diversity[C]. *Proc of the 10th IEEE Int Conf on Data Mining*. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2010: 619-628.
- [24] Li Y, Xu L, Wang Y G, et al. A new diversity measure for classifier fusion[C]. *Proc of the 2nd Int Conf on Multimedia and Signal Processing*. Berlin: Springer-Verlag, 2012: 396-403.
- [25] Lofstrom T, Johansson U, Bostrom, H. On the use of accuracy and diversity measures for evaluating and selecting ensembles of classifiers[C]. *Proc of the 7th Int Conf on Machine Learning and Applications*. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2008: 127-132.
- [26] Chandra A, Chen H, Yao X. Trade-off between diversity and accuracy in ensemble generation[M]. Berlin: Springer Berlin Heidelberg, 2006: 429-464.
- [27] Hu Q, Yu D. Diversity measure for multiple classifier systems[C]. *Proc of the 2nd Int Conf on Fuzzy Systems and Knowledge Discovery*. Berlin: Springer-Verlag, 2005: 1261-1265.
- [28] Liu W Y, Wu Z H, Pan G. An entropy-based diversity measure for classifier combining and its application to face classifier ensemble thinning[C]. *Proc of the 5th Chinese Conf on Biometric Recognition*. Berlin: Springer-Verlag, 2004: 118-124.
- [29] Brown G. An information theoretic perspective on multiple classifier systems[C]. *Proc of the 8th Int Workshop on Multiple Classifier Systems*. Berlin: Springer-Verlag, 2009: 344-353.

- [30] Zhou Z H, Li N. Multi-information ensemble diversity[C]. Proc of the 9th Int Workshop on Multiple Classifier Systems. Berlin: Springer-Verlag, 2010: 134-144.
- [31] Yu Y, Li Y F, Zhou Z H. Diversity regularized machine[C]. Proc of the 22nd Int Joint Conf on Artificial Intelligence. Washington DC, 2011: 1603-1608.
- [32] Li N, Yu Y, Zhou Z H. Diversity regularized ensemble pruning[C]. Proc of the European Conf on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2012: 330-345.
- [33] Didaci L, Fumera G, Roli F. Diversity in classifier ensembles: Fertile concept or dead end?[C]. Proc of the 11th Int Workshop on Multiple Classifier Systems. Berlin: Springer-Verlag, 2013: 37-48.
- [34] Brown G, Kuncheva L I. "Good" and "Bad" diversity in majority vote ensembles[C]. Proc of the 9th Int Workshop on Multiple Classifier Systems. Berlin: Springer-Verlag, 2010: 124-133.
- [35] Schapire R E, Freund Y, Bartlett P. Boosting the margin: A new explanation for the effectiveness of voting methods[J]. Annals of Statistics, 1998, 26(5): 1651-1686.
- [36] Tang E K, Suganthan P N, Yao X. An analysis of diversity measures[J]. Machine Learning, 2006, 65(1): 247-271.
- [37] Hsu K W, Srivastava J. Improving bagging performance through multi-algorithm ensembles[C]. Proc of the Int Workshops on PAKDD. Berlin: Springer-Verlag, 2012: 471-482.
- [38] Hsu K W, Srivastava J. Diversity in combinations of heterogeneous classifiers[C]. Proc of the 13th Pacific-Asia Conf on KDD. Berlin: Springer-Verlag, 2009: 923-932.
- [39] Zhou Z H, Wu J, Tang W. Ensembling neural networks: Many could be better than all[J]. Artificial Intelligence, 2002, 137(1/2): 239-263.
- [40] Zhou Z H, Tang W. Selective ensemble of decision trees[C]. Proc of the 9th Int Conf on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Berlin: Springer-Verlag, 2003: 476-483.
- [41] Margineantu D D, Dietterich T G. Pruning adaptive boosting[C]. Proc of the 14th Int Conf on Machine Learning. New York: Association for Computing Machinery, 1997: 211-218.
- [42] Lazarevic A, Obradovic Z. Effective pruning of neural network classifier ensembles[C]. Proc of the 1st Int Joint Conf on Neural Networks. New York: IEEE, 2001: 796-801.
- [43] Fu Q, Hu S X, Zhao S Y. Clustering-based selective neural network ensemble[J]. J of Zhejiang University, 2005, 6(5): 387-392.
- [44] Kuncheva L I, Jain L C. Designing classifier fusion systems by genetic algorithms[J]. IEEE Trans on Evolutionary Computation, 2000, 4(4): 327-336.
- [45] Zhang Y, Burer S, Street W N. Ensemble pruning via semi-definite programming[J]. J of Machine Learning Research, 2006, 7(1): 1315-1338.
- [46] Partalas I, Tsoumakas G, Vlahavas I. Focused ensemble selection: A diversity-based method for greedy ensemble selection[C]. Proc of the 18th European Conf on Artificial Intelligence. Amsterdam: IOS Press, 2008: 117-121.
- [47] Yang L. Classifiers selection for ensemble learning based on accuracy and diversity[J]. Procedia Engineering, 2011, 15(1): 4266-4270.
- [48] Kuncheva L I. That elusive diversity in classifier ensembles[C]. Proc of the 1st Iberian Conf on Pattern Recognition and Image Analysis. Berlin: Springer-Verlag, 2003: 1126-1138.
- [49] Kapp M N, Sabourin R, Maupin P. An empirical study on diversity measures and margin theory for ensembles of classifiers[C]. Proc of the 10th Int Conf on Information Fusion. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2007: 1-8.
- [50] Narasimhamurthy A. Evaluation of diversity measures for binary classifier ensembles[C]. Proc of the 6th Int Workshop on Multiple Classifier Systems. Berlin: Springer-Verlag, 2005: 267-277.
- [51] Hsu K W, Srivastava J. Relationship between diversity and correlation in multi-classifier systems[C]. Proc of the 14th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin: Springer-Verlag, 2010: 500-506.
- [52] Tumer K, Ghosh J. Error correlation and error reduction in ensemble classifiers[J]. Connection Science, 1996, 8(3/4): 385-404.
- [53] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [54] Wang S, Chen H, Yao X. Negative correlation learning for classification ensembles[C]. Proc of the Int Joint Conf on Neural Networks. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2010: 1-8.
- [55] Ko A H R, Sabourin R, Souza Britt A. Combining diversity and classification accuracy for ensemble selection in random subspaces[C]. Proc of the Int Joint Conf on Neural Networks. New Jersey: Institute of Electrical and Electronics Engineers Inc Computer Society, 2006: 2144-2151.
- [56] Ko A H R, Sabourin R. Compound diversity functions for ensemble selection[J]. Int J of Pattern Recognition and Artificial Intelligence, 2009, 23(4): 659-686.