



计算机科学

COMPUTER SCIENCE

一种自适应权重的多分类通用集成方法

魏军胜, 刘琰, 陈静, 段顺然

引用本文

魏军胜, 刘琰, 陈静, 段顺然. 一种自适应权重的多分类通用集成方法[J]. 计算机科学, 2022, 49(11): 212-220.

WEI Jun-sheng, LIU Yan, CHEN Jing, DUAN Shun-ran. [Universal Multi-class Ensemble Method with Self Adaptive Weights](#)[J]. Computer Science, 2022, 49(11): 212-220.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[数据流概念漂移处理方法研究综述](#)

Survey of Concept Drift Handling Methods in Data Streams

计算机科学, 2022, 49(9): 14-32. <https://doi.org/10.11896/jsjcx.210700112>

[基于细粒度语义推理的跨媒体双路对抗哈希学习模型](#)

Fine-grained Semantic Reasoning Based Cross-media Dual-way Adversarial Hashing Learning Model

计算机科学, 2022, 49(9): 123-131. <https://doi.org/10.11896/jsjcx.220600011>

[基于对偶变分多模态注意力网络的不完备社会事件分类方法](#)

Dual Variational Multi-modal Attention Network for Incomplete Social Event Classification

计算机科学, 2022, 49(9): 132-138. <https://doi.org/10.11896/jsjcx.220600022>

[多层注意力机制融合的序列到序列中国连续手语识别和翻译](#)

Sequence-to-Sequence Chinese Continuous Sign Language Recognition and Translation with Multi-layer Attention Mechanism Fusion

计算机科学, 2022, 49(9): 155-161. <https://doi.org/10.11896/jsjcx.210800026>

[基于自注意力模型的本体对齐方法](#)

Ontology Alignment Method Based on Self-attention

计算机科学, 2022, 49(9): 215-220. <https://doi.org/10.11896/jsjcx.210700190>

一种自适应权重的多分类通用集成方法

魏军胜¹ 刘 琰¹ 陈 静² 段顺然¹

1 河南省网络空间态势感知重点实验室 郑州 450001

2 信息工程大学大数据分析教研室 郑州 450001

(1509451874@qq.com)

摘 要 集成学习一直是构建强大和稳定的预测模型的策略之一,它能通过融合多个模型来提升结果的准确性和稳定性。但是,现有的集成方法在权重计算上还存在一定的缺陷,面对多种分类问题时无法自适应地选择集成权重,不具有通用性。针对以上问题,提出了一种自适应权重的多分类通用集成方法(UMEAW)。与通常的集成分类方法只针对一种分类任务不同,UMEAW 面对不同的分类问题,首先根据分类个数计算权重调配系数,然后利用指数函数分布特性,根据模型评价指标与权重调配系数自动计算一次模型融合的权重,最后通过不断迭代的方法自适应地调整融合权重,实现不同分类任务下的模型融合。实验结果表明,UMEAW 在 9 个不同分类个数、不同领域、不同规模的数据集上都能实现模型融合,其融合效果在大部分任务上都优于基线方法。与单个模型相比,用 UMEAW 融合后的结果 F_1 值稳定增加了 3%25%;与其他集成方法相比, F_1 值稳定提升了 1%2%,证明了 UMEAW 的通用性和有效性。

关键词:集成学习;权重;分类;融合;通用方法

中图法分类号 TP391

Universal Multi-class Ensemble Method with Self Adaptive Weights

WEI Jun-sheng¹, LIU Yan¹, CHEN Jing² and DUAN Shun-ran¹

1 Key Laboratory of Cyberspace Situation Awareness of Henan, Zhengzhou 450001, China

2 Department of Big Data Analysis, Information Engineering University, Zhengzhou 450001, China

Abstract Ensemble learning has always been one of the strategies to build a powerful and stable predictive model. It can improve the accuracy and stability of the results by fusing multiple models. However, existing ensemble methods still have certain shortcomings in the calculation of weights. When facing a variety of classification problems, they cannot adaptively select ensemble weights, and they are not universal. In view of the above problems, a universal multi-class ensemble method with self-adaptive weights(UMEAW) is proposed. Different to usual ensemble classification method that only targets one kind of classification task, when facing different classification problems, firstly, UMEAW calculates the weight allocation coefficient according to the number of classification, and then the weights of base classifiers is automatically calculated according to the model evaluation index and the weight allocation coefficient by using the distribution characteristics of exponential function. Finally, the weights is adjusted adaptively through continuous iteration to realize model ensemble under different classification tasks. Experimental results show that UMEAW can achieve model ensemble on 9 datasets with different classification numbers, different fields and different scales, and the effect of UMEAW is better than the baselines in most tasks. Compared with a single model, F_1 value increases by 3%25% after UMEAW fusion. Compared with other ensemble methods, the F_1 value improves by 1%2%. It is proved that UMEAW is universal and effective.

Keywords Ensemble learning, Weight, Classification, Fusion, Universal method

1 引言

集成学习的主要思想是团体决策,其将多个弱学习器的结果通过各种组合机制融合,以获得比任何单独组成算法更好的结果,实现了准确性和稳定性的提升^[1-2]。集成学习作为数据挖掘和机器学习领域的研究重点,其研究热度一直居高不下。从 20 世纪八九十年代少数研究者提出集成的思想以来,大量研究者投身于集成学习的研究中,一些经典的集成

算法逐渐成为该领域的主流方法,包括 Boosting^[3-4], Ada-Boost^[5], Stacking^[6], Weighting average, Majority voting, Weighted majority voting^[7]等。同时,由于集成方法能稳定提升预测性能,研究者们仿佛找到了实现模型预测效果提升的捷径。在进入 21 世纪后,大量利用集成方法解决单一领域问题的方法开始涌现。

过去的研究从不同侧重点出发设计了很多集成算法,为本文开展研究提供了参考和依据。但是,这些方法仍存在着

到稿日期:2021-09-06 返修日期:2022-03-21

基金项目:国家自然科学基金(U1804263,62002386)

This work was supported by the National Natural Science Foundation of China(U1804263,62002386).

通信作者:刘琰(ms.liuyan@foxmail.com)

一些问题,主要表现为以下3个方面:

(1)集成策略容易忽略潜在的较优权重。目前集成方法的集成策略多数采用集体决策,即少数服从多数的原则,导致参与融合的模型表现差的数目多于表现好的数目时,得不到好的融合结果;或者过多地依赖表现好的模型,给好的模型赋予更大的融合权重,忽略了模型融合时,赋予表现差的模型较大权重反而会获得更好的融合结果的可能性。

(2)权重计算收敛速度不够快。有一些集成方法考虑了第一个问题,但这些方法的融合策略过于复杂,导致权重计算收敛速度慢,从而影响模型效率。

(3)多分类问题上的通用性不强。绝大多数方法只适用于二分类或三分类问题,而若要解决多分类问题,则需要从算法层面进行改变。

基于此,本文在权重计算过程中引入指数函数和权重调整系数,提出了一种新的权重计算方法,并以赋权融合为基础,设计了一种自适应权重的多分类通用集成方法(Universal Multi-class Ensemble Method with Self-adaptive Weights, UMEAW),解决了以上3个问题;并在9个不同分类个数、不同领域、不同规模的数据集上验证了该方法的有效性。

本文第2节介绍了相关工作;第3节给出了本文涉及的一些概念,说明了待研究问题的定义;第4节介绍了自适应权重的多分类通用集成方法的理论依据、原理以及算法流程;第5节进行了实验分析,在9个不同的数据集上给出了所提方法的测试结果;最后讨论了自适应权重的多分类通用集成方法的有效性和局限性。

2 相关工作

本文的研究对象是分类集成方法。在过去的研究中,大部分方法聚焦于调整集成过程中的权重以获得好的分类效果,这也是本文研究的主要内容之一。过去的研究从总体上可分为训练过程中的方法,以及训练结束后的方法。

在训练过程中实施融合的研究可以追溯到1990年, Freund^[3]和 Shapire^[4]提出了 Boosting 算法,通过训练多个弱分类器,为每个弱分类器赋予不同的权重,同时调整其权重并进行投票来实现分类,但是 Boosting 算法要求事先知道在最坏的情况下学习器的分类错误率,这很难做到,从而促使了著名集成方法 AdaBoost^[5]的产生,其通过关注错误分类本来调整样本权重,取得了较好的效果。后来 Wolpert^[6]提出了 Stacking 方法,该方法采用多折交叉的方式采样数据,并通过不同的基分类器训练采样数据,然后将训练得到的结果作为训练数据再次训练,从而达到融合多个分类器特征的目的。Yu 等^[8]设计了一种自适应的半监督分类器集成框架,该框架面向高维数据分类器的合成,能自适应地选择模型融合权重。Dong 等^[9]提出了一种 DWES 算法,通过关注分类器的熵值在集成过程中动态计算权重,但该方法只适用于二分类问题。以上方法需要在训练过程中集成,因此每一次集成需要一次训练,会耗费大量的时间。

为了节省训练时间,部分研究者提出了一些模型训练结束后的集成方法。这类方法通常分为两个阶段,第一阶段选择基分类器,第二阶段利用分类器训练后的输出结果,通过合适的融合规则提升结果的准确性。在分类器选择阶段,每一个分类器被视为特征选择的专家,最终选择的结果通常由分类器在特定数据上训练的效果得到^[10-12]。而在第二阶段,需要

有效的融合方法将第一阶段得到的基分类器效果最大化。Kolter 等^[13]提出了一种基于基分类器加权求和算法的通用框架,在分类过程中保持基分类器不变,其权重在不同的数据块上不断地修改,并通过权重与分类器的加权求和进行样本预测。与 Kolter 的研究类似, Mao 等^[14]提出了一种分类权重融合方法,用于平衡分类器的多样性以及准确性。Muhlbaier 等^[15]设计了一种动态加权的咨询和投票机制,以整合分类标签。相比训练过程中的集成方法,该方法能灵活地调整融合策略,便于更好地寻找融合方法。

以上集成方法都属于全集成方法,即在实施集成的过程中融合了所有分类器。但当分类器数目不断增多时,会面临计算和存储开销持续上升,而集成效果却没有提升的问题。为了避免这个问题,部分研究者开展了选择理想分类器进行集成的研究,即选择性集成方法。该方法一般分为聚类法、优化法和排序法3类。聚类法将所有基分类器根据某个准则划分为不同的类别,然后根据这个准则选择用于集成的分类器,例如文献^[16]提出的 CEMF(Clustering Ensembles Based on Multi-classifier Fusion)方法。优化法将分类器选择视为目标优化问题,使用启发式学习得到理想的分类器,例如基于遗传算法的选择性集成算法 GAS-EN(Genetic Algorithm Selective Ensemble)^[17]、基于支持向量机的选择性集成算法 SSE(Support vector machines Selective Ensemble)等^[18],以及最近 Yang 等^[19]提出的基于自适应学习率的自动集成算法(Auto-Ensemble)。排序法通过设定评价指标对基分类器进行排序,将排序靠前的基分类器作为理想分类器^[20-21],该方法在选择分类器的速度和空间利用上具有优势,但排序的评价指标缺乏统一的标准。

在最近几年的研究中,关于集成算法本身的研究逐渐减少,人们开始尝试改进或调整过去的模型,以解决不同领域的分类问题。文献^[22-24]解决了各自领域的实际问题,并取得了很好的效果,从应用层面证明了集成方法的优势以及研究价值,但对通用算法的研究贡献较少。

3 问题描述

F_1 :评价模型优劣的指标,该值结合了精确率和召回率的评价优势,可以较好地评价实验结果。其计算过程如下:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$F_1 = \frac{2 * precision * recall}{precision + recall} \quad (3)$$

其中, TP 是 True Positive, FP 是 False Positive, FN 是 False Negative。

定义 1(归一化的 $F_1 - f^*$) 为了便于计算,本文将模型的 F_1 值进行归一化处理,归一化公式如式(4)所示。为了便于表示,在下面的公式中,记 F_1 值为 f ,并将归一化之后的 F_1 记为 f^* 。

$$f^* = \frac{(f - f_{\min})}{(f_{\max} - f_{\min})} \quad (4)$$

其中, f_{\min} 指所有模型中最小的 F_1 值, f_{\max} 指所有模型中最大的 F_1 值。

定义 2(基模型) 本文将参与模型融合的单个模型称为基模型,记为 B-model。

定义 3(模型融合) 在多个基模型预测结果的基础上,通过融合方法得到集体决策结果。

定义 4(分类标签) 在分类问题中,为了方便地表示不同的类别,通常会赋予每个类别一个标签,本文中用 *label* 表示。

定义 5(多分类问题) 本文将所有分类标签种类大于等于 2 的分类问题统称为多分类问题。

定义 6(标签概率) 分类模型在得到预测结果的过程中,会生成每个分类标签对应的概率值,然后通过标签所对应的概率值大小确定最终的结果。本文将此过程中得到的概率值称为标签概率,并记为 p 。

定义 7(赋权融合) 对每个基模型的预测标签概率赋予相应的权重,记权重为 ω_i 。将所有基模型的标签概率值与权重相乘再求和即可得到赋权融合结果。其计算式如下:

$$p_{\text{weighted}} = \omega_1 * p_1 + \omega_2 * p_2 + \dots + \omega_n * p_n \quad (5)$$

对 x 分类问题而言,每个样本分别对应着 x 个标签概率。假设有 y 个样本,则这 y 个样本就有 $x \times y$ 个标签概率,用矩阵的形式对标签概率进行表示,如式(6)所示:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1y} \\ p_{21} & p_{22} & \dots & p_{2y} \\ \vdots & \vdots & \ddots & \vdots \\ p_{x1} & p_{x2} & \dots & p_{xy} \end{pmatrix} \quad (6)$$

则 n 个模型赋权融合后的标签概率矩阵可以表示为:

$$P' = \sum_i^n P_i \cdot \omega_i = \begin{pmatrix} p'_{11} & p'_{12} & \dots & p'_{1y} \\ p'_{21} & p'_{22} & \dots & p'_{2y} \\ \vdots & \vdots & \ddots & \vdots \\ p'_{x1} & p'_{x2} & \dots & p'_{xy} \end{pmatrix} \quad (7)$$

最终的分类标签结果取每个样本上最大概率值对应的标签,如式(8)所示:

$$\text{label}_k = \arg \max(p'_{k1}, p'_{k2}, \dots, p'_{kx}), k=1, 2, \dots, y \quad (8)$$

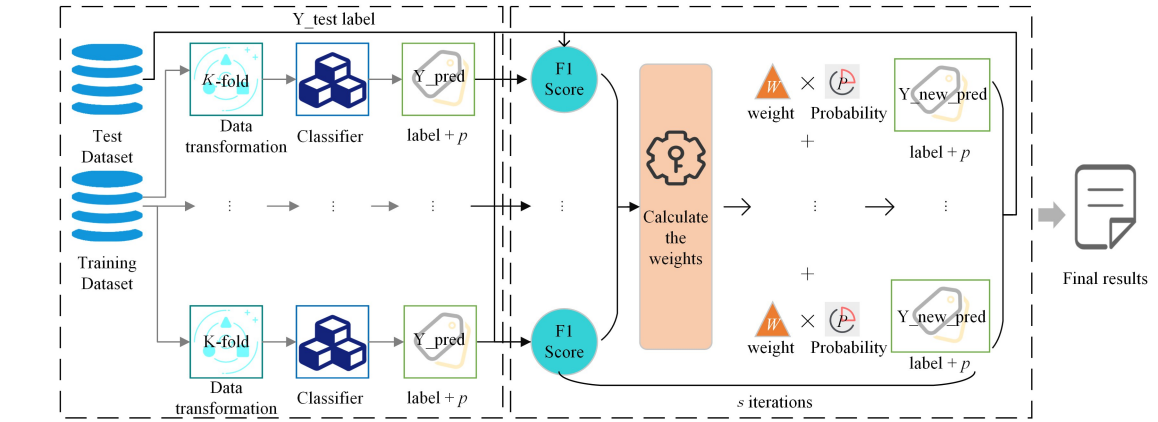
为了便于后续的描述,将所有的样本标签用向量表示为 $L = [\text{label}_1, \text{label}_2, \dots, \text{label}_k, \dots, \text{label}_y]^T$ 。

根据以上概念,本文待研究的问题可以描述为:在假定已经选取了多个分类器作为基模型的前提下,根据基模型的输出结果,以赋权融合方法为基础,设计高效集成方法,解决融合过程中权重的计算,提升分类模型的融合效果。

在提出集成方法的过程中有几个关键的工作:1)基模型如何得到较好的预测结果,方法选择要防止模型在融合后过拟合;2)模型融合过程中的权重该如何计算,权重计算方法影响着模型融合效果;3)模型融合方法设计,主要考虑如何快速有效地使模型融合达到最佳效果。

4 方法描述

本文提出的自适应权重的多分类通用集成方法框架如图 1 所示。



注:左侧虚线框中的内容表示集成准备过程,右侧虚线框中的内容表示模型融合的过程

图 1 自适应权重的多分类通用集成方法总体流程图

Fig. 1 Overall flow chart of Universal Multi-class Ensemble Method with Self Adaptive Weights

该框架包含两个部分:集成准备与模型融合。集成准备表示单个分类器得到预测的分类标签和标签概率的过程,模型融合指单独分类器的结果通过融合策略得到融合结果的过程。该方法大致的过程是:将数据集分为训练数据集和测试数据集,在训练数据集上得到 B-model 预测结果(分类标签和标签概率),然后根据测试数据集上基模型的 F_1 值计算每个基模型的权重,再通过迭代赋权融合得到最终结果,从而达到提升分类预测的效果。

4.1 集成准备

在正式进行模型融合之前,需要得到单个模型预测的分类标签与对应的标签概率。该过程即单个模型的训练和预测过程。为了得到稳定的融合结果,本文将通过 K 折交叉验证的方式训练单个模型。

第 1 节提到,基分类器的选择会影响集成效果,而基分类器训练的处理方式也会影响模型的融合效果。在深度学习的应用上,很多情况下会出现过拟合,表现为学习器在训练过程中参数过度拟合训练数据集,从而影响模型在测试数据集上的泛化性能的现象。究其原因,最根本的问题是训练数据量不足以支持复杂的模型,从而导致模型学习的数据集上出现噪声。但若对结果进行平均,则在一定程度上可以减少过拟合现象。如图 2 所示,由于过拟合,单个模型的决策边界为黄色,但实际上,黑色的决策边界具有更好的泛化能力,因此其效果更好。如果对多个模型进行拟合并对结果进行平均,则将会由于结果的平均而减少对这些噪声点的考虑,决策边界将逐渐向黑线靠拢。

因此,本文将使用 K 折交叉验证来训练模型。首先,将原始

数据平均分为 K 组,并为每个子集数据创建一个验证集,然后将剩余的 $K-1$ 个子集数据用作训练集。这样,将获得 K 组预测结果。对这 K 组预测标签的概率取平均,得到基模型预测结果,即平均后的概率以及对应的标签。在实际操作中, K 值根据分类数量确定,如当面对三分类问题时,一般会将 K 取为 5,这样做既可以避免划分数据时各自为一类的极端情况,也可避免因折数太大而造成计算资源的浪费。

在训练好模型之后,在测试集上得到预测结果,即样本标签 L 和对应的概率矩阵 P ,为下一步进行模型融合提供支撑。集成准备过程如图 3 所示。

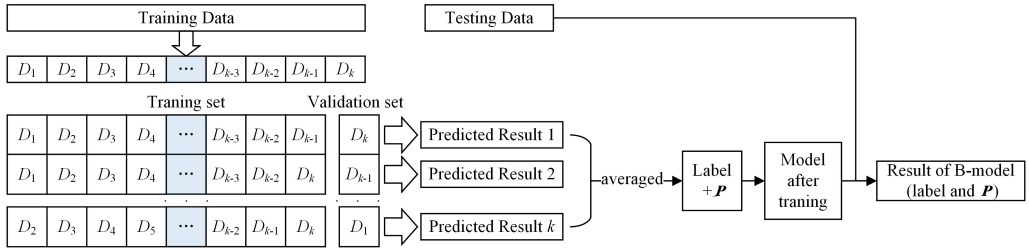


图 3 集成准备过程

Fig. 3 Ensemble preparation process

4.2 模型融合

前文已经介绍过赋权融合在集成学习中有着无可比拟的作用,绝大多数算法以赋权融合为基础。本节也将以赋权融合为基础,设计能自适应于多分类的权重计算方法,用不断迭代的方法得到较优的融合权重,实现模型融合。

4.2.1 自适应权重方法设计

由第 1 节的分析可知,模型融合的权重非常重要,权重分布直接影响模型融合效果,因此如何快速且有效地找到较优的权重组合非常关键。本小节设计了一种自适应权重计算方法(AW),用于计算模型融合的权重,其遵循的基本原则是根据模型的 F_1 值大小赋予模型相应的权重,即 F_1 值较大的模型得到较大权重, F_1 值较小的模型得到较小权重,同时也考虑给 F_1 值较小的模型赋予较大权重的情况。

在描述方法之前,需要说明两个值得注意的问题:1)为了解决权重计算收敛速度不够快的问题,AW 引入了指数函数 $g(x)=e^x$,借助指数函数能够快速增长的特性,计算权重时能做到快速收敛;2)为了解决计算权重容易忽略潜在的较优权重的问题,在计算模型权重时加入了权重调系数 α ,通过改变该系数的值,可以调整模型的权重分布,使得模型能够尽可能多地获得不同的权重。

本文方法设定所有模型的权重和等于 1,且用分数来表示每个权重。假设有 n 个模型,且按照模型 F_1 值大小从小到大排序,则权重的具体表示如下。

(1)分子。将 $e^{f_n^*}$ 作为分子,可以让权重与模型 F_1 值的大小具有正相关的联系。值得注意的是,第 n 个模型的分子还需乘以系数 α ,即 $\alpha \cdot e^{f_n^*}$,其目的在于调整第 n 个模型(F_1 值最大的模型)的权重,当 α 值增大时,该模型也会获得更大的权重,同时其他模型的权重会随之下降。

(2)分母。将所有模型的分子相加作为分母,即 $\alpha \cdot e^{f_n^*} + \sum_{i=1}^{n-1} e^{f_i^*}$ 。这种做法保证了所有分数的求和等于 1,符合

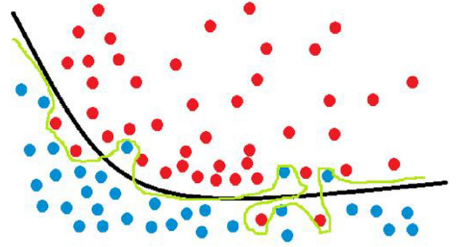


图 2 拟合曲线图(电子版为彩图)

Fig. 2 Curve fitting diagram

原来的设定。

(3)权重计算公式。上述分子与分母组成的分数即为模型对应的权重,其计算式为:

$$\omega_i = \begin{cases} \frac{e^{f_i^*}}{\alpha \cdot e^{f_n^*} + \sum_{i=1}^{n-1} e^{f_i^*}}, & i=1,2,\dots,n-1 \\ \frac{\alpha \cdot e}{\alpha \cdot e^{f_n^*} + \sum_{i=1}^{n-1} e^{f_i^*}}, & i=n \end{cases} \quad (9)$$

由于 $f_n^* = (\frac{f_{\max} - f_{\min}}{f_{\max} - f_{\min}}) = 1$,则 $e^{f_n^*} = e$,则式(9)可以化简为式(10):

$$\omega_i = \begin{cases} \frac{e^{f_i^*}}{\alpha \cdot e + \sum_{i=1}^{n-1} e^{f_i^*}}, & i=1,2,\dots,n-1 \\ \frac{\alpha \cdot e}{\alpha \cdot e + \sum_{i=1}^{n-1} e^{f_i^*}}, & i=n \end{cases} \quad (10)$$

其中, $\omega_i \geq 0, \sum_i \omega_i = 1$ 。

(4)权重调系数 α 的取值范围。由于 α 的取值直接决定权重的计算结果,因此需要对 α 的取值加以讨论。

为了方便讨论 α 的取值范围,首先对式(10)做简化处理。由于 f_i^* 是归一化之后的值,因此它的取值范围为 $[0, 1]$,故 $e^{f_i^*}$ 的取值范围为 $[1, e]$,则 $\sum_{i=1}^{n-1} e^{f_i^*}$ 的取值范围为 $[n-1, (n-1)e]$,为了简化计算量,取 $\sum_{i=1}^{n-1} e^{f_i^*}$ 的近似值为 $\frac{3}{2}(n-1)$,则 ω_i 的值可以简化为式(11):

$$\omega_i = \begin{cases} \frac{e^{f_i^*}}{\alpha \cdot e + \frac{3}{2}(n-1)}, & i=1,2,\dots,n-1 \\ \frac{\alpha \cdot e}{\alpha \cdot e + \frac{3}{2}(n-1)}, & i=n \end{cases} \quad (11)$$

接下来,借助式(11)讨论 α 的取值,主要根据 α 的取值对 w_n (最大权重)的影响开展讨论。首先令 α 的值为 $n-1$ 的倍数,这时可以很轻松地将式(11)-②约分为常数。当 $\alpha=5(n-1)$ 时, $w_n=5e/(5e+3/2)\approx 0.9$, 此时其他模型的权重已经非常小了,如果继续增大 α 的取值,则会使其其他模型在融合过程中的作用微乎其微,因此将 $\alpha=5(n-1)$ 作为 α 的最大取值。

当 $\alpha=1$ 时, w_n 等于根据模型 F_1 值按比例计算得到的值。当 $\alpha<1$ 时, w_n 能够取得更小的值, $w_i (i=1, 2, \dots, n-1)$ 可以取得更大的值。因此,为了扩大权重的取值范围,给较差的模型赋予较大的权重,本文方法加入参数 $t(0<t<1)$ 作为 α 的最小值。同时,为了方便计算 α 的中间取值,将 t 作为 α 的变化步长。

根据以上讨论,可以确定 α 的取值以及个数,其取值范围为 $[t, 2t, 3t, \dots, 5(n-1)]$, 其取值个数为 $5(n-1)/t$, 其中 n 为模型个数。

值得说明的是,式(10)是最终的权重计算公式,而式(11)是为了说明 α 的取值问题而简化的公式。

4.2.2 自适应权重的多分类通用集成方法

为了得到更好的模型融合结果,本节在自适应权重计算方法(AW)的基础上,设计自适应权重的多分类通用集成方法(UMEAW),该方法以集成准备得到的结果为输入数据,通过不断迭代进行多次赋权融合,实现融合效果的提升。UMEAW 的流程总结如算法 1 所示。

算法 1 UMEAW

输入: the testing sample set $D=(d_1, d_2, \dots, d_i)$ and the corresponding label set $Y=(y_1, y_2, \dots, y_i)$; the corresponding predicted label of the n-group B-model after ensemble preparation (Sort in ascending order by the size of F_1): $Y_1'=(y_1', y_2', \dots, y_1')_1, \dots, Y_i'=(y_1', y_2', \dots, y_1')_i, \dots, Y_n'=(y_1', y_2', \dots, y_1')_n$, and the corresponding probability set $P_1'=(p_1', p_2', \dots, p_1')_1, \dots, P_i'=(p_1', p_2', \dots, p_1')_i, \dots, P_n'=(p_1', p_2', \dots, p_1')_n$; the weight function (formula(10)); the F_1 function; the value of step size t ; the Maximum F_1 value $F_{1_max}=0$

输出: the label set after ensemble $Y_e'=(y_1', y_2', \dots, y_1')_e$, and the corresponding F_1 value

1. Calculate the F_1 value of n-group B-model by F_1 function
2. while:
3. for $\alpha \in [t, 2t, 3t, \dots, 5(n-1)]$ do
4. utilize the weight function to calculate the weights of n-group B-model $\{w_1, w_2, \dots, w_n\}$
5. use formula(7) to carry out weighting fusion, obtain probability set $P''=(p_1'', p_2'', \dots, p_1'')$ after weighting fusion and corresponding label set $Y''=(y_1'', y_2'', \dots, y_1'')$, and calculate the corresponding F_1 value by the F_1 function
6. end for
7. Sort the results obtained in step 3-6 in ascending order according to the F_1 value, select the last n results and update $Y_1'=(y_1', y_2', \dots, y_1')_1, \dots, Y_i'=(y_1', y_2', \dots, y_1')_i, \dots, Y_n'=(y_1', y_2', \dots, y_1')_n, P_1'=(p_1', p_2', \dots, p_1')_1, \dots, P_i'=(p_1', p_2', \dots, p_1')_i, \dots, P_n'=(p_1', p_2', \dots, p_1')_n$ with these n results

8. $F_{1_max} \leftarrow$ the largest F_1 value in step 7
9. end while if $F_{1_max} \leq$ the last F_{1_max}
10. Output the predicted label $Y_e'=(y_1', y_2', \dots, y_1')_e$ and the corresponding F_1 value

5 实验

本节将在不同的数据集上测试 UMEAW 的效果,并与多种基线方法进行比较,验证 UMEAW 的有效性以及通用性。

5.1 实验设置

5.1.1 实验数据

本文的实验数据来源于 9 个不同数据集,其中新闻情感分析的数据集(News Sentiment)来源于 2019 CCF BDCI 新闻情感分析赛道¹⁾,其余用于测试方法效果的数据集均是机器学习领域常见的公开数据集。表 1 列出了各数据集的名称、本文中的 ID、分类数、样本总数和数据属性数。

表 1 本文使用的数据集

Table 1 Data sets used in this paper

| Datasets | Dataset ID | Classes | Instances | Attributes |
|-------------------|------------|---------|-----------|------------|
| News Sentiment | D_1 | 3 | 7 357 | 3 |
| Twitter Sentiment | D_2 | 2 | 31 962 | 3 |
| Coverttype | D_3 | 7 | 581 012 | 55 |
| Mushrooms | D_4 | 2 | 8 124 | 21 |
| Spambase | D_5 | 2 | 4 601 | 58 |
| Nursery | D_6 | 5 | 12 960 | 8 |
| Breast Cancer | D_7 | 2 | 699 | 32 |
| Wine | D_8 | 3 | 178 | 13 |
| Iris | D_9 | 3 | 150 | 4 |

由于数据集 $D_2 - D_9$ 是机器学习领域所熟知的数据集,因此本节对这些数据集不再进行详细的分析,只介绍 News Sentiment 数据集的情况。该数据集包含的属性为:新闻 ID 号(id)、新闻标题内容(Title)、新闻正文内容(Content)、新闻情感标签(Label)。标签的正面情感对应 0,中性情感对应 1,负面情感对应 2。

对于每个数据集,按照 7:3 的比例划分训练集和测试集。根据 4.1 节所述,需要将训练集划分为 K 份,如果 K 小于等于分类数 $Classes$,则可能出现划分数据各自为一类的极端情况,为了避免此类情况的发生,将 $D_1, D_2, D_4, D_5, D_7 - D_9$ 数据集的分类数都小于 5 的数据集随机划分为 5 份,将 D_3, D_6 随机划分为 10 份,进行 4.1 节所述的集成准备过程。

5.1.2 评价指标

为了方便多种方法之间的比较,本文使用综合性的评价指标,即 F_1 值。该值结合了精确率和召回率的评价优势,可以较好地评价实验结果。其计算过程已在第 2 节中介绍过,这里不再赘述。值得注意的是,多分类任务计算的是 $Macro-F_1$ 值,通过式(3)计算每一类标签的 F_1 值,再将每一类标签的 F_1 值求平均,即可得到 $Macro-F_1$ 值。

5.1.3 基线方法

(1) 选择基模型

由文献[25]可知,正确率和差异性较大的基模型能帮助集成学习获得更好的泛化精度,因此在选择基模型时需要

¹⁾ <https://www.datafountain.cn/competitions/350/datasets>

兼顾模型的正确率和差异性。

由于 D_1 涉及文学、政治和新闻报道等多个领域,因此在 D_1 数据集的分析任务上放弃传统深度学习模型(如 RNN/CNN)作为基模型,而是选取 Transformer 架构模型。本文选取 BERT^[26],RoBERTa^[27],XLNet^[28] 分别与 softmax 的组合作为基模型。本实验的模型权重文件来自于 Cui¹⁾ 发布在 GitHub 上的项目。

由于 $D_2 - D_9$ 数据集在传统机器学习模型上运用得较为成熟,并且传统的机器学习模型的训练时间相对较短,因此选择逻辑回归(Logistic Regression)、支持向量分类器(Support Vector Classifier)和随机森林(Random Forest)等传统机器学习模型作为基模型,本实验通过调用 sklearn 的包实现以上 3 个模型。

(2)选择基线方法

为了验证所提方法的有效性,本文选择了几种经典的全集成方法(Weighting average(Wavg),Majority voting(Mv),Weighted majority voting(Wmv)^[7],Stacking(Stk)^[6])、经典的选择性集成方法(GAS-EN^[17]),以及最近提出的选择性集成方法(AE^[20])作为基线方法,用于与本文提出的 UMEAW 进行比较。其中 Weighting average, Majority voting, Weighted majority voting, Stacking 用 sklearn²⁾ 提供的库实现, GAS-EN 由 Harry-Mills³⁾ 公布的开源代码实现, AE 根据论文提出的算法实现。

5.1.4 参数设置

对于 BERT, RoBERTa 和 XLNet 这些 Transformer 架构模型,训练过程中的主要参数如表 2 所列。对于 Logistic Regression, Support Vector Classifier 和 Random Forest 这些模型的训练参数,采用 sklearn 库提供的默认参数。

表 2 Transformer 架构模型的主要参数设置

Table 2 Main parameter settings of Transformer architecture model

| parameter | value |
|-------------------------|--------------------|
| max_seq_length | 256 |
| split_num | 3 |
| eval_steps | 200 |
| per_gpu_eval_batch_size | 32 |
| num_train_epochs | 3 |
| learning_rate | 5×10^{-6} |
| lstm_hidden_size | 512 |
| lstm_layers | 1 |
| lstm_dropout | 0.1 |
| train_steps | 1000 |

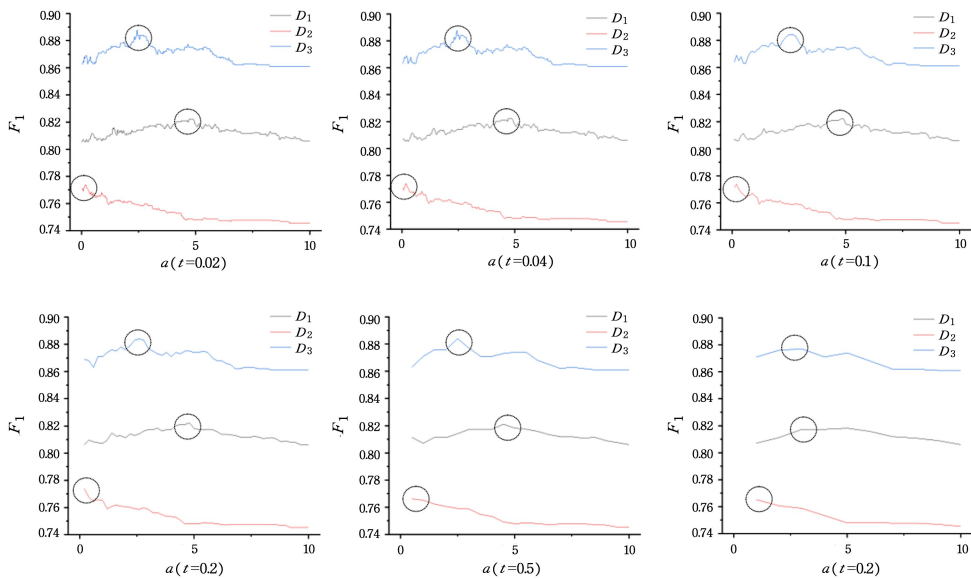
5.2 步长 t 对融合效果的影响分析

由前文可知,权重调配系数 α 直接影响权重的计算结果。由于 $\alpha \in [t, 2t, 3t, \dots, 5(n-1)]$, 且本实验选择的基模型个数为 3, 因此 $\alpha \in [t, 2t, 3t, \dots, 10]$, 故 α 的取值只由 t 确定。因此,对 t 的取值进行深入分析非常有必要。

为了便于分析和作图,本实验选取了 F_1 值相差较大的数据集 D_1, D_2, D_3 来测试 t 的取值对融合结果的影响。

首先需要选取实验中待测试的 t 值。由前文可知,当 t 为 0.02 时, α 的最小值为 0.02, 此时由权重计算式(10)计算可得, F_1 值最大的模型的权重在 0.03 左右, 在模型融合时所占的比重非常小, 当逐步增大 t 的值时, F_1 值最大的模型的权重也随之增大。因此,将 $t=0.02$ 作为本小节实验中的最小值, 然后逐步增大 t 的值开展实验, 取 $t=0.02, 0.04, 0.1, 0.2, 0.5, 1$ 。

为了直观地看到 t 值对模型融合的影响,本节的实验只做 UMEAW 中的第一轮融合,不再迭代多次。在每个数据集上的实验结果如图 4 所示。



注:椭圆标注的部分表示 F_1 值较大的区间

图 4 $t=0.02, 0.04, 0.1, 0.2, 0.5, 1$ 时, D_1, D_2, D_3 上的第一轮融合结果

Fig. 4 When $t=0.02, 0.04, 0.1, 0.2, 0.5, 1$, results of the first round of model fusion on D_1, D_2 and D_3

¹⁾ <https://github.com/ymcui>, 下载的权重文件分别是 Bert-wwm, RoBERTa-wwm-ext-large, XLNet-base

²⁾ <https://scikit-learn.org/stable/modules/ensemble>

³⁾ <https://github.com/HarryMills/GASEN>

由图 4 可以看出,不论 t 取值多大,模型融合效果较好时的 α 取值区间较为固定(图 4 中用椭圆圈住的位置),说明 AW 能较好地捕捉到赋权融合的较优权重。

同时,从图中也可以看出, t 值越小, α 的取值变化越小,模型融合的结果越多,得到的最优结果越好。例如,当 $t=0.02$ 和 $t=0.04$ 时, D_3 的结果中有较为突出的融合结果,而当 $t>0.1$ 时,这些较好的结果丢失了,这说明若想要得到好的融合结果, t 要取较小的值。

但 t 的取值也不宜过小,过小的 t 值既增加了计算量,还会在进一步获取好的权重组合时变得迟缓。观察 $t=0.02$ 和 $t=0.04$ 的两张图可以发现,两幅图的轮廓基本一致, $t=0.02$ 的图中的重要节点并没有在 $t=0.04$ 的图中丢失,这说明当 $t=0.04$ 时,已经能得到较好的权重组合了, t 值继续变小意义不大。基于以上分析,本文将选取 $t=0.04$ 进行后续的实验。

5.3 迭代对模型融合的影响分析

本节在 5.2 节中第一轮融合的基础上完成了整个 UMEAW 的过程,分析迭代对融合结果的影响。

在 5.2 节中,当 $t=0.04$ 时, D_1 , D_2 和 D_3 经过第一轮融合后,最大 F_1 值分别是 0.822, 0.888, 0.773。经过 UMEAW 的融合后,3 个数据集上的最大 F_1 值都有所提升。 D_1 , D_2 和 D_3 的 F_1 值分别经过了 10, 13 和 8 次迭代后停止了迭代,得到了 0.8287, 0.7830, 0.8972 的最终 F_1 值。3 个数据集每一次迭代得到的 F_1 值如表 3 所列。

表 3 D_1 , D_2 和 D_3 每轮迭代得到的 F_1 值

Table 3 F_1 values obtained in each iteration of D_1 , D_2 and D_3

| Number of iterations | D_1 | D_2 | D_3 |
|----------------------|---------------|---------------|---------------|
| 1 | 0.8221 | 0.7730 | 0.8880 |
| 2 | 0.8228 | 0.7748 | 0.8908 |
| 3 | 0.8235 | 0.7751 | 0.8913 |
| 4 | 0.8239 | 0.7765 | 0.8945 |
| 5 | 0.8245 | 0.7771 | 0.8953 |
| 6 | 0.8256 | 0.7777 | 0.8968 |
| 7 | 0.8267 | 0.7782 | 0.8970 |
| 8 | 0.8273 | 0.7790 | 0.8972 |
| 9 | 0.8280 | 0.7794 | 0.8964 |
| 10 | 0.8287 | 0.7801 | |
| 11 | 0.8280 | 0.7809 | |
| 12 | | 0.7824 | |
| 13 | | 0.7830 | |
| 14 | | 0.7819 | |

从表 3 中可以看到,每轮迭代 F_1 值都有小幅的提升,这说明迭代过程对模型融合是有效的。但与第一轮的融合效果相比,后来的每次迭代对 F_1 值的提升幅度相对较小,可能的原因是第一轮直接使用基模型进行赋权融合,每个基模型的差异性较大,融合后 F_1 值的提升较大。而第 2 轮开始的迭代操作,以融合了 3 个基模型的结果为基础,差异性变小,导致融合效果提升较小。

5.4 基模型对融合效果的影响分析

本节通过消融实验探讨每个基模型对模型融合结果的贡献程度。表 4 列出了 D_1 — D_9 数据集上去除不同的基模型后,使用 UMEAW 融合剩余基模型得到的 F_1 值以及使用全部

3 个基模型得到的 F_1 值。其中, D_1 的基模型是 BERT, RoBERTa 和 XLNet, D_2 — D_9 的基模型是 Logistic Regression (LR), Support Vector Classifier (SVC) 和 Random Forest (RF), 表 4 分别用 BM_1 , BM_2 和 BM_3 表示 3 个基模型的结果,用 $No-BM_1$, $No-BM_2$ 和 $No-BM_3$ 分别表示去除对应基模型的结果, All 表示融合全部基模型的结果。

表 4 消融实验的 F_1 值

Table 4 F_1 value of ablation experiments

| Dataset ID | $BM_1/No-BM_1$ | $BM_2/No-BM_2$ | $BM_3/No-BM_3$ | All |
|------------|----------------|--------------------|----------------|-------|
| D_1 | 0.803/0.818 | 0.812/0.813 | 0.802/0.819 | 0.829 |
| D_2 | 0.753/0.759 | 0.749/0.764 | 0.665/0.778 | 0.783 |
| D_3 | 0.697/0.876 | 0.773/0.887 | 0.869/0.853 | 0.897 |
| D_4 | 0.914/0.951 | 0.934/0.943 | 0.931/0.946 | 0.957 |
| D_5 | 0.904/0.947 | 0.697/0.951 | 0.945/0.948 | 0.953 |
| D_6 | 0.931/0.961 | 0.944/0.966 | 0.964/0.946 | 0.983 |
| D_7 | 0.891/0.969 | 0.964/0.959 | 0.942/0.971 | 0.979 |
| D_8 | 0.972/0.989 | 0.98/0.986 | 0.982/0.985 | 0.994 |
| D_9 | 0.976/0.976 | 0.947/0.981 | 0.976/0.980 | 0.989 |

从表中可以看出,对于 D_1 而言, RoBERTa(BM_2) 的 F_1 值最大,并且 $No-BM_2$ 小于 $No-BM_1$ 和 $No-BM_3$, 说明 RoBERTa 在 3 个模型融合中的贡献最大,与自身的 F_1 值成正比。对于 D_2 — D_9 而言, LR, SVC 和 RF 在每个数据集上的表现不一致,例如 LR 在 D_2 上的表现最好,但在 D_3 , D_4 , D_6 等数据集上表现最差,这说明每个基模型在处理不同的数据集上各有所长。而对于 D_5 而言, SVC(BM_2) 的 F_1 值只有 0.697, 但与其他两个基模型分别利用 UMEAW 融合时,还能取得相对较好的结果,这说明表现差的模型参与 UMEAW 的融合时,也能对模型融合有积极的贡献,证明相比其他以集体决策为原则的方法,本文方法具有不可比拟的优势。

5.5 UMEAW 与基线方法的比较

本节在数据集 D_1 — D_9 上分别比较了 Weighting average (Wavg), Majority voting (Mv), Weighted majority voting (Wmv), Stacking (Stk), GAS-EN, Auto-Ensemble (AE) 以及 UMEAW 的融合效果。

针对每个数据集,每种方法执行了 3 次程序,取其中最好的一次结果列入统计,结果如表 5 所列,表 5 中加黑的数字表示所有结果中最好的融合结果。可以看到, UMEAW 在 6 个数据集上取得了最好的成绩,与其他融合方法相比,较为稳定地提升了 1%2%,这说明本文方法与其他融合方法相比具有一定的优越性。

表 5 UMEAW 与基线方法的 F_1 值比较

Table 5 Comparison of F_1 value between UMEAW and baselines

| Dataset ID | Wavg | Mv | Wmv | Stk | GAS-EN | AE | UMEAW |
|------------|--------------|-------|--------------|--------------|--------------|--------------|--------------|
| D_1 | 0.817 | 0.816 | 0.817 | 0.813 | 0.823 | 0.817 | 0.829 |
| D_2 | 0.779 | 0.769 | 0.77 | 0.771 | 0.779 | 0.772 | 0.783 |
| D_3 | 0.862 | 0.868 | 0.875 | 0.834 | 0.892 | 0.897 | 0.897 |
| D_4 | 0.952 | 0.953 | 0.945 | 0.945 | 0.959 | 0.955 | 0.957 |
| D_5 | 0.941 | 0.946 | 0.945 | 0.923 | 0.945 | 0.931 | 0.953 |
| D_6 | 0.986 | 0.977 | 0.978 | 0.980 | 0.981 | 0.986 | 0.983 |
| D_7 | 0.966 | 0.974 | 0.971 | 0.946 | 0.979 | 0.971 | 0.979 |
| D_8 | 0.988 | 0.985 | 0.994 | 0.987 | 0.989 | 0.989 | 0.994 |
| D_9 | 0.979 | 0.989 | 0.985 | 0.989 | 0.986 | 0.985 | 0.989 |

图 5 比较了 D_1 在 BERT, RoBERTa 和 XLNet 这 3 个

基模型和 $D_2 - D_9$ 在 Logistic Regression(LR), Support Vector Classifier(SVC)和 Random Forest(RF)这 3 个基模型上的 F_1 值,以及 3 个基模型用 UMEAW 融合后的结果。可以看到,在不同分类数目的 9 个数据集上,融合后的结果与基模型相比 F_1 值稳定地提升了 3%~25%,证明了本文方法的有效性以及通用性。

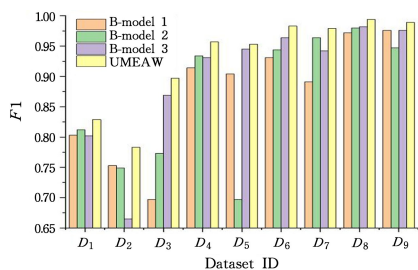


图5 $D_1 - D_9$ 数据集上基模型与 UMEAW 的 F_1 值对比

Fig. 5 F_1 value comparison between baselines and UMEAW on

$D_1 - D_9$ data sets

6 讨论

平均分配权重或者按比例分配权重的方式不同,本文通过改变权重调配系数 α 的值来调整权重的分布,使得模型融合时能选择更多的权重组合,以增加得到较优权重的可能性。同时,由于 UMEAW 借助了指数函数的特性,权重能够在 α 值改变的过程中较快地收敛。

正是由于权重调配系数 α 在 UMEAW 中的关键作用, α 的取值变得非常重要,本文通过引入步长 t 研究了 α 取值变化对融合效果的影响。5.2 节通过实验证明了 t 值越小(即 α 的初始值和变化步长越小),得到的权重分布范围越广,融合效果越好。但当 t 减小到一定的值之后,融合效果提升会变得非常缓慢。这是因为 t 值很小时, α 的变化非常小,从权重计算式(10)可知, α 变化越小,权重变化越小,而相似的权重组合的融合结果几乎相同。因此,取一个合适的 t 值既能减少计算量,也能保证不影响模型融合的效果。

在集成准备过程中,本文使用 K 折交叉验证来增加基模型的预测结果的稳定性,进而增加模型融合取得更好结果的可能性。由于集成准备和模型融合是两个相对独立的过程,因此不使用 K 折交叉验证训练基模型同样也能用本文提出的融合方法进行模型融合。因此,本文提出的模型融合方法能灵活地应用到其他场景,只要使用者提供基模型在测试集上预测的标签以及对应的概率值即可。

此外,UMEAW 还存在着一些不足。由于该方法需要训练多个基模型,并且每个基模型都要通过多折交叉训练得到中间结果,因此相比单个模型,本文方法会耗费较多的时间和计算资源,这也是模型集成方法的通病;另外,由于在权重计算过程中,对表现最优的模型赋予了较大的权重,因此本文方法对最优模型有一定的依赖性。5.3 节通过分析迭代对模型融合的影响发现,迭代过程对融合效果提升的幅度不够大,出现这种情况可能的原因是第 2 轮之后的迭代操作,待融合的模型差异性变小,导致融合效果提升较小。因此,如何降低对最优模型的依赖以及如何提升迭代过程中模型的不同性是

我们下一步的研究方向。

结束语 本文借助指数函数的特性,设计权重计算方法,并以此为基础提出了一种自适应权重的多分类通用集成方法。该方法包括集成准备和模型融合两个部分,以 K 折交叉检验为基础的集成准备为模型融合提供了较好的基础,以自适应权重计算方法(AW)为基础的模型融合能够自适应地、快速地收敛得到较优的权重,从而得到更好的融合结果。

然后,本文在 9 个不同分类个数、不同领域、不同规模的数据集上,选取 3 个基模型进行了单元实验以及对比实验。实验结果表明,UMEAW 在每个数据集上都能取得较为不错的融合效果,证明该方法是有效的,并且能够适用于多种分类模型的集成,可以为其他分类问题的优化提供参考。

此外,我们使用该方法在 2019 年 CCF 大数据与计算智能大赛新闻情感分析赛道中取得了第二名的成绩,也充分证明了该方法的有效性。

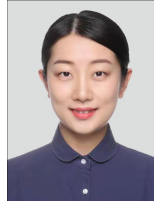
参考文献

- [1] OPITZ D, MACLIN R. Popular Ensemble Methods: An Empirical Study[J]. Journal of Artificial Intelligence Research, 1999, 11: 169-198.
- [2] HANSEN L K, SALAMON P. Neural network ensembles[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 12(10): 993-1001.
- [3] FREUND Y. Boosting a weak learning algorithm by majority[J]. Information and Computation, 1995, 121(2): 256-285.
- [4] SHAPIRE R. The strength of weak learnability[J]. Machine Learning, 1990, 5(2): 197-227.
- [5] JI Z, HUI Z, ROSSET S, et al. Multi-class adaboost[J]. Statistics and its Interface, 2009, 2(3): 349-360.
- [6] WOLPERT D H. Stacked generalization[J]. Neural Networks, 2017, 5(2): 241-259.
- [7] LITTLESTONE N, WARMUTH M K. The weighted majority algorithm[J]. Information and Computation, 1994, 108(2): 212-261.
- [8] YU Z, WANG D, ZHAO Z, et al. Hybrid incremental ensemble learning for noisy real-world data classification[J]. IEEE Transactions on Cybernetics, 2017, 49(2): 403-416.
- [9] DONG M G, ZHANG W, JING C. Dynamic weight ensemble classification algorithm for unbalanced data streams[J]. Miniature Microcomputer Systems, 2020, 41(8): 1649-1655.
- [10] GAO W, ZHOU Z H. Approximation stability and boosting [C]// International Conference on Algorithmic Learning Theory. Berlin: Springer, 2010: 59-73.
- [11] YIN X C, HUANG K, HAO H W, et al. A novel classifier ensemble method with sparsity and diversity[J]. Neurocomputing, 2014, 134: 214-221.
- [12] ZHANG L, SUGANTHAN P N. Random Forests with ensemble of feature spaces[J]. Pattern Recognition, 2014, 47(10): 3429-3437.
- [13] KOLTER J Z, MALOOF M A. Dynamic weighted majority: An ensemble method for drifting concepts[J]. Journal of Machine Learning Research, 2007, 8(Dec): 2755-2790.

- [14] MAO S, JIAO L, XIONG L, et al. Weighted classifier ensemble based on quadratic form[J]. *Pattern Recognition*, 2015, 48(5): 1688-1706.
- [15] MUHLBAIER M D, TOPALIS A, POLIKAR R. Learn++-NC: Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes[J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 152-168.
- [16] HUANG Y, MONEKOSSO D, WANG H. Clustering Ensembles Based on Multi-classifier Fusion[C]// *IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2010)*. Shanghai: IEEE, 2010: 393-397.
- [17] NGUYEN T T, LIEW A W C, TRAN M T, et al. Combining multi classifiers based on a genetic algorithm—a gaussian mixture model framework[C]// *International Conference on Intelligent Computing*. Cham: Springer, 2014: 56-67.
- [18] SONG J Z, GUO C Y, LIU H S. Selective SVM Ensemble Base on Clustering Analysis Apply for Analog Circuit Fault Diagnosis with Small Samples[J]. *Applied Mechanics & Materials*, 2013, 380-384: 841-845.
- [19] YANG J, WANG F. Auto-Ensemble: An Adaptive Learning Rate Scheduling Based Deep Learning Model Ensembling[J]. *IEEE Access*, 2020, 8(12): 217499-217509.
- [20] LU Z, WU X, ZHU X, et al. Ensemble Pruning via Individual Contribution Ordering[C]// *ACM Sigkdd International Conference on Knowledge Discovery & Data Mining*. Washington: ACM, 2010: 871.
- [21] MARTÍNEZ-MUÑOZ G, HERNÁNDEZ-LOBATO D, SUÁREZ A. An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2009, 31(2): 245-259.
- [22] HASAN M A, ABDULLAH N, RAHMAN M, et al. Dental Impression Tray Selection From Maxillary Arch Images Using Multi-Feature Fusion and Ensemble Classifier[J]. *IEEE Access*, 2021, 9(2): 30573-30586.
- [23] BANIA R K, HALDER A. R-HEFS: Rough set based heterogeneous ensemble feature selection method for medical data classification[J]. *Artificial Intelligence in Medicine*, 2021, 114: 102049-102080.
- [24] LAI C M, HUANG H P. A gene selection algorithm using simplified swarm optimization with multi-filter ensemble technique[J]. *Applied Soft Computing*, 2021, 100: 106994-107042.
- [25] ZHOUG, GUO F L. Research on high-dimensional data integration learning method based on feature selection[J]. *Computer Science*, 2021, 48(S1): 250-254.
- [26] DEVLIN J, CHANG M, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv*, 1810. 04805, 2018.
- [27] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for chinese bert[J]. *arXiv*: 1906. 08101, 2019.
- [28] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized Autoregressive Pretraining for Language Understanding[J]. *Advances in Neural Information Processing Systems*, 2019, 32: 5753-5763.



WEI Jun-sheng, born in 1992, postgraduate. His main research interests include big data analysis and so on.



LIU Yan, born in 1979, Ph.D, associate professor, Ph.D supervisor. Her main research interests include network topology discovery and network data analysis.

(责任编辑: 喻黎)