

doi: 10.7690/bgzd.2022.01.009

基于 AHP 的 Stacking 算法基分类器选择

孙 彤, 陈砚桥

(海军工程大学动力工程学院, 武汉 430033)

摘要: 为找到与分类任务最贴切的基分类器组合, 提出一种基于层次分析法 (analytic hierarchy process, AHP) 的 Stacking 算法基分类器选择方法。通过 AHP 确定分类器单一性能指标的权值, 以各指标加权求和的结果作为评价分类器好坏的标准来选择基分类器。通过对不同组合的基分类器进行 Stacking 集成发现, 性能最优的单个分类器集成效果并不是最佳的, 基分类器个数最多的模型也不是最适合任务需求的。结果表明, 该方法在提高分类器性能方面具有可行性和有效性。

关键词: Stacking 算法; 层次分析法; 分类; 基分类器选择

中图分类号: U664.12 **文献标志码:** A

Base Classifiers Selection for Stacking Algorithm Based on AHP

Sun Tong, Chen Yanqiao

(School of Power Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: In order to find the most appropriate combination of base classifiers for the classification task, an AHP-based base classifier selection method for Stacking algorithm is proposed. The weights of the single performance index of the classifier are determined by the AHP, and the base classifier is selected by taking the weighted sum result of each index as the standard for evaluating the quality of the classifier. Through the Stacking integration of different combinations of base classifiers, it is found that the single classifier with the best performance is not the best, and the model with the largest number of base classifiers is not the most suitable for the task requirements. The results show that the method is feasible and effective in improving the performance of the classifier.

Keywords: Stacking algorithm; AHP; classification; base classifier selection

0 引言

单一的机器学习算法难以对大量样本进行精度较高的分类。集成算法可以将不同的机器学习算法结合起来, 提高分类准确率和泛化能力^[1]。Stacking 算法作为一种应用广泛的集成算法, 将不同基分类器产生的结果作为下一层算法的输入, 融合了所有基分类器的特点, 提高了分类准确率^[2]。研究发现: Stacking 算法比 Adabost、Bagging 等集成算法表现要好^[3]。

Stacking 算法发展迅速, 已被广泛应用于各种分类任务中, 有学者对其进行了改进研究。文献[4]立足于网络入侵数据对 Stacking 算法在个体分类器的产生、选择和结合方法上进行改进, 取得了较好效果。文献[5]通过构建 3 层结构的 Stacking 算法提高其在多分类任务上的分类性能。文献[6]提出了用于大数据的自动迭代多层分类器 Stacking 算法, 大大提高了分类准确率。文献[7]—[8]分别采用人工蜂群算法、蚁群算法优化 Stacking 算法, 提高了算法

的分类准确率。文献[9]利用 Stacking 算法对卷积神经网络进行改进, 用主成分分析法对基分类器的输出进行降维, 并在多个数据集上进行验证, 结果证明该方法具有更高的分类准确率和稳定性。Kotsiantis^[10]提出一种代价敏感的 Stacking 算法用于不平衡数据的分类, 提高了少数类数据的识别率。

Stacking 算法通过集成不同类型的基分类器得到一个强分类器, 用于提高 Stacking 算法准确率和泛化能力; 但在准确率提高的同时, 算法的其他性能也存在下降的可能, 这就使得在面对不同的分类任务时, 单纯的堆积 Stacking 算法基分类器如果不对基分类器做出选择就无法得到完成任务的最优解。不同的分类任务对算法的性能要求也有不同, 如在一般的分类任务中, 准确率是评价分类器好坏的重要指标; 但在不平衡数据集的分类任务中, 往往更加关注的是对少数类数据的识别, 准确率就没那么重要。不同的基分类器组合会导致 Stacking 算

收稿日期: 2021-09-30; 修回日期: 2021-10-20

作者简介: 孙 彤(1992—), 男, 辽宁人, 硕士, 从事动力及热力系统的科学管理研究。E-mail: 1346441749@qq.com。

法的性能有所差异。现有的 Stacking 算法研究中，多以性能良好且类型不同作为基分类器的选择标准，这种依靠人们经验的基分类器选择方法很难在众多的基分类器中选择出最优的基分类器组合；因此，有必要针对具体任务的特点，并基于 AHP 的 Stacking 算法研究如何选择出符合具体任务要求的最优基分类器组合。

1 AHP-Stacking 算法

1.1 Stacking 算法介绍

Stacking 算法是一种常用的集成算法，一般分为 2 层。第 1 层通过训练多种分类器将其作为 Stacking 算法的基分类器，用第 2 层的输出结果当作第 2 层的元分类器的输入，训练得到 Stacking 模型。Stacking 算法框架如图 1 所示。

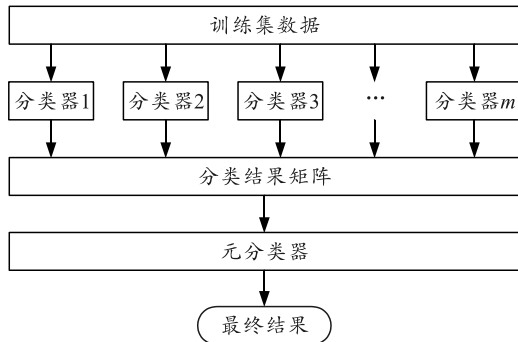


图 1 Stacking 算法框架

1.2 分类器性能评价指标

如表 1 所示，在二分类任务中，可将样本分为正样本和负样本，样本和分类结果共同构成了混淆矩阵。

表 1 混淆矩阵

预测结果	实际为正	实际为负
正	TP	FP
负	FN	TN

上表中：TP 表示实际为正例、预测为正例的个数；FP 表示实际为负例、预测为正例的个数；FN 表示实际为正例、预测为负例的个数；TN 表示实际为负例、预测为负例的个数。

1) 准确率(Accuracy)。

准确率表示被正确分类的样本数量占样本总数的比例；通常来说，准确率越高，分类器越好。准确率的评价模型为：

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

2) 精确率(Precision)。

精确率表示被分类为正样本的数据中实际也为

正样本的比率；精确率越大，说明被分类为正样本的结果更可靠。精确率评价模型为：

$$Precision = TP / (TP + FP) \quad (2)$$

3) 召回率(Recall)。

召回率表示正确分类为正例的样本数占正例样本总数的百分比；召回率越大，说明更多的正样本被正确地识别出来。召回率评价模型为：

$$Recall = TP / (TP + FN) \quad (3)$$

4) F1-Score 值。

精确率和召回率负相关，F1-Score 值是同时考虑了精确率和召回率平衡之后的结果；其值越大，表示分类器越有效。F1-Score 值评价模型为：

$$F1 = 1 / (1/Precision + 1/Recall) \quad (4)$$

1.3 AHP-Stacking 算法

针对不同的分类任务，分类器性能评价指标的重要程度也应有所差异。在一般数据集的分类任务中，Accuracy 往往是评价分类器性能的重要指标，但在不平衡数据集中，Precision 和 Recall 更能体现少数类数据的识别率。笔者采取一种根据分类任务目的，对 Accuracy, Precision, Recall, F1-Score 按其对于任务的重要程度，利用层次分析法(AHP)确定各单一分类性能指标的权重，然后对指标评价值进行加权求和，作为分类性能的综合评价结果。

AHP 是一种定性-定量结合的多目标决策分析及综合评价方法，用来解决多目标问题的定性与定量相结合的问题。决策者利用经验来判断条件相对于任务的重要程度，并合理地给出每个条件的权重。借助这种方法可以将决策者的经验判断量化，从而实现最优化决策。将 AHP 加入到 Stacking 算法的框架中，算法可分为 3 个步骤。具体的 AHP-Stacking 算法流程如图 2 所示。

步骤 1:

- 1) 获取 n 个备选的基分类器；
- 2) 利用穷举法列出所有可能的基分类器组合；
- 3) 利用 pima 数据集和所有可能的基分类器组合构建 Stacking 模型并得到其分类性能指标值。

步骤 2:

- 1) 根据任务需求利用 AHP 确定分类性能指标的权值；
- 2) 对所有训练好的 Stacking 模型分类性能指标值加权求和得到其综合评价指标值。

步骤 3:

选取综合评价指标值最大的 Stacking 模型作为最终的分类模型。

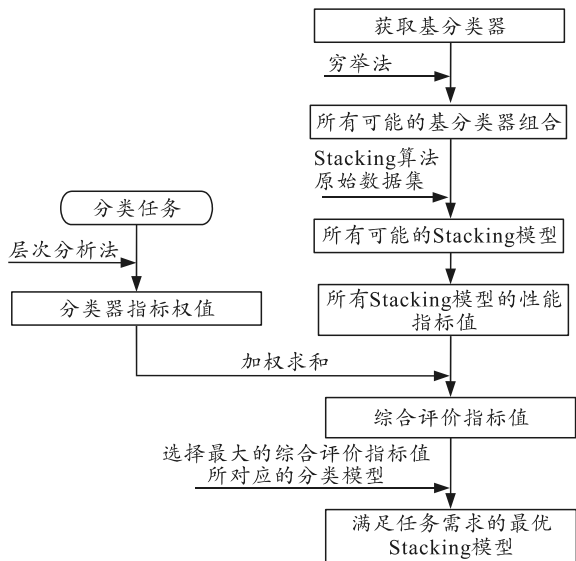


图 2 AHP-Stacking 算法流程

2 实例分析

分类器选用随机森林 (random forest, RF)、逻辑回归 (logistic regression, LR)、人工神经网络 (artificial neural network, ANN)、KNN、决策树 (decision tree, DT)、Adaboost (ADB) 共 6 种，元学习器采用逻辑回归，采用 10 折交叉验证的方法训练弱分类器，训练集数据占比 80%，测试集数据占比 20%，并利用 UCI 数据集中 Pima 数据集 (印第安人糖尿病数据) 对所提方法进行验证。样本数据量为 768，其中少数类数据量为 246 (正例)，多数类数据量为 522 (负例)，不平衡率为 1:2，为不平衡数据集。

2.1 权值确定

在 Pima 数据集中，糖尿病患者数据量远少于

正常样本数据量，即使少数样本被错分，模型的准确率也不会有明显变化，而模型精确率影响却很大；因此，在此分类任务中，精确率比准确率重要得多。精确率越高，表示正常数据被分类为糖尿病患者的少，被分类为糖尿病患者的结果更为可信；召回率越高，表示糖尿病患者被误诊为正常的数据少。糖尿病患者被误诊为正常的代价要高于正常数据被分类为糖尿病患者的代价，故召回率的重要程度要高于精确率。如果分类器的预测结果全为糖尿病患者，则召回率会很高而精确率却很低，此分类结果没有意义；如果分类器的预测结果中，预测为糖尿病患者的结果全部正确但数量很少，则精确率很高而召回率很低，此分类结果也没有意义。F1-Score 是精确率和召回率调和平均值，其值越高代表着分类结果越可信，其重要程度要高于召回率。根据专家打分得到 4 个评价指标的权值判断如表 2 所示。

表 2 权值判断

评价指标	准确率	召回率	精确率	F1-Score
准确率	1	1/5	1/4	1/6
召回率	5	1	2	1/2
精确率	4	1/2	1	1/3
F1-Score	5	2	3	1

利用 AHP 方法，计算得到各指标权值如下： $W=(0.061\ 0, 0.293\ 4, 0.182\ 1, 0.463\ 5)$ 。对得出的权值进行一致性检验，得 $CR=0.041\ 2 < 0.1$ ，结果有效。

2.2 计算结果及分析

根据图 2 的算法流程，利用 Python 软件编程实现该算法。该算法能计算出所有可能的分类器组合的单一性能指标以及综合性能结果。由于分类器组合较多，故在表 3 中只列出部分有价值的分类器组合及其性能指标值。

表 3 部分基分类器组合的 Stacking 算法性能指标

序号	分类器组合	评价指标				综合评价指标值
		准确率	精确率	召回率	F1-Score	
1	RF	0.77	0.77	0.44	0.56	0.61
2	LR	0.81	0.79	0.59	0.68	0.70
3	DT	0.69	0.68	0.17	0.27	0.40
4	ADB	0.79	0.78	0.54	0.64	0.67
5	KNN	0.69	0.55	0.47	0.51	0.57
6	ANN	0.73	0.69	0.35	0.46	0.52
7	LR+ADB	0.81	0.80	0.60	0.69	0.71
8	LR+ADB+ANN	0.82	0.81	0.60	0.69	0.72
9	RF+ LR+ADB+ANN	0.81	0.81	0.56	0.67	0.70
10	RF+ LR+DT+ANN	0.83	0.82	0.60	0.70	0.72
11	RF+LR+DT+ADB+ANN	0.82	0.83	0.62	0.71	0.73
12	RF+LR+DT+ADB+KNN+ANN	0.81	0.81	0.59	0.68	0.71

由上表可看出：

1) 综合评价指标值最高的基分类器组合为 11 号组合，即 RF+LR+DT+ADB+ANN，其值为 0.73；

因此，由 RF、LR、DT、ADB、ANN 这 5 种基分类器集成的 Stacking 算法是满足任务需求的最优解。

2) 准确率最高的基分类器组合为 10 号组合，

即 RF+LR+DT+ANN, 其值为 0.83, 对比 10、11 号 2 种分类器组合: 前者准确率较高, 说明在 Pima 数据集中, 更多的数据被分类正确; 后者的精确率较高, 说明后者对于糖尿病患者这一少数类别的分类结果上更为可信; 后者召回率高于前者, 说明后者可以将更多数量的糖尿病患者分类出来。综合比较以上 3 种指标可以发现, 11 号分类器组合集成的 Stacking 模型对糖尿病患者正确分类的数量多且结果可信度高, 之所以其准确率低于 10 号, 是因为其对正常数据正确分类的数量。F1-Score 值后者高于前者, 说明后者分类的结果更加可信。在 Pima 这种类不平衡的数据集中, 少数类数据往往蕴含了更多的信息, 分类算法关注的重点是有没有将更多的糖尿病患者以较高的准确率分类出来, 故对于此数据集, 11 号分类器组合集成的 Stacking 模型分类效果更好。

3) 比较 9、10 号 2 种基分类器组合, 可以发现 10 号的所有指标值都高于 9 号; 然而, 9 号和 10 号 2 种组合的差异在于 9 号组合选用了 ADB 而 10 号选用了 DT, 比较 ADB 和 DT 可以发现, ADB 的性能指标都高于 DT, 说明性能好的基分类器其构成的 Stacking 模型性能不一定好。

4) 12 号基分类器组合集成的 Stacking 模型的性能指标都低于 11 号, 说明只增加基分类器的数量不一定能提升模型的性能。

综上所述, 笔者提出的基于 AHP 的 Stacking 算法基分类器选择方法是有效的。

3 结束语

笔者立足于具体任务需求提出 AHP-Stacking 算法, 通过此算法可找到针对具体分类任务的最佳基分类器组合, 利用笔者所提方法, 模型准确率、精确率、召回率、F1-Score 值和综合评价指标值分别提高了 0.01、0.02、0.03、0.03 和 0.03。对基分类器和 Stacking 算法性能之间的关系进行分析发

现: 当单个基分类器性能最佳时, 其通过 Stacking 算法集成后的分类模型性能不一定是最佳的; 分类器个数的增加, 有助于 Stacking 算法整体性能的提升, 但并不一定是组合的基分类器个数越多分类性能越好。而且, 当基分类器个数很大时, 计算成本会成指数增加。在未来的研究中, 如何快速有效地找到满足任务要求的最优基分类器组合是重点研究方向。

参考文献:

- [1] ZHANG L, SUGANTHAN P N. Oblique decision tree ensemble via multisurface proximal support vector machine[J]. IEEE Trans Cybern, 2015, 45(10): 2165-2176.
- [2] WOLPERT D H. Stacked Generalization[M]. Spring US, 2011: 20.
- [3] 鲁莹, 郑少智. Stacking 学习与一般集成方法的比较研究[EB/OL]. 北京: 中国科技论文在线[2017-02-09]. <http://www.paper.edu.cn/releasepaper/content/201702-43>.
- [4] DEMIT N, DALKILIC G. Modified stacking ensemble approach to detect network intrusion[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2018, 26(1): 130-143.
- [5] EITAN M, LIOR R, YUVAL E. An improved stacking schema for classification tasks[J]. Information Sciences, 2009, 179(24): 4097-4122.
- [6] ABAWAJY J H, KELAREV A, CHOWDHURY M. Large Iterative Multitier Ensemble Classifiers for Security of Big Data[J]. Emerging Topics in Computing IEEE Transactions on 2014, 2(3): 353-263.
- [7] SHUNMUGAPRIYA P, KANMANI S. Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm[J]. Swarm & Evolutionary Computation, 2013, 12(12): 24-32.
- [8] CHEN Y, MAN L W. An Ant Colony Optimization approach for stacking ensemble[C]//Nature and Biologically Inspired Computing. IEEE, 2011: 7-8.
- [9] 张笑铭, 王志君, 梁利平. 一种适用于卷积神经网络的 Stacking 算法[J]. 计算机工程, 2018, 44(4): 243-247.
- [10] KOTSIANTIS S. Stacking cost sensitive models[C]//Proc of 2008 Panhellenic conference on Informatics, 2008: 217-221.