

基于层次分析与 Adaboosting 的电力用户 信用评价方法

徐宏宽,林顺富,边晓燕,李东东

(上海电力大学 电力工程学院,上海 200090)

摘要:我国电力市场化改革进程不断推进,将作为市场主体的电力用户信用评价问题分析清楚对电力市场的构建具有重要而深远的意义。针对电力用户信用评价问题提出一种基于层次分析与 Adaboosting 的电力用户信用综合评价方法。基于电力用户数据信息,建立了归一化综合评价指标体系;构建了基于层次分析法的信用评价模型,从典型分类算法中遴选出合适的备选分类算法;采用 Adaboosting 算法对备选分类算法进行线性动态赋权,通过对样本数据训练建立组合分类模型。基于实际电力用户信息数据对提出的综合评价方法进行应用验证,结果表明:相比已有的分类算法,所提出的综合评价方法的评价准确率与曲线下面积(AUC)值均较高,具有较好的适用性。

关键词:电力用户;信用评价;层次分析;Adaboosting 算法

中图分类号:TM464 **文献标识码:**A **DOI:**10.19457/j.1001-2095.dqed23520

Power User Credit Classification Method Based on the Analytic Hierarchy Process and Adaboosting Algorithm

XU Hongkuan, LIN Shunfu, BIAN Xiaoyan, LI Dongdong

(College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: With the reform of the electric marketization, it is of great and far-reaching significance for the construction of the electricity market to analyze clearly the credit evaluation problems of power users as the main market. Aiming at the problem of power user credit evaluation, a comprehensive evaluation method based on the analytic hierarchy process (AHP) and Adaboosting algorithm was proposed. Based on the power user data information, a normalized comprehensive evaluation index system was established. A credit evaluation model based on the AHP was constructed, and the suitable candidate classification algorithms were selected out from the typical classification algorithms. The Adaboosting algorithm was adopted to perform the linear dynamic weighting on the candidate classification algorithms, and a combined classification model was established by training sample data. The proposed method was verified with the actual power user information data. The results show that the proposed method has higher evaluation accuracy and area under curve (AUC) value, and has better applicability with the existing classification methods.

Key words: power users; credit evaluation; analytic hierarchy process (AHP); Adaboosting algorithm

电是当今社会发展和人类生活需求的基础性资源,是国家生产力发展的动力。电力企业对所有的电力用户采取一致的营销策略和电费结算策略,这既不利于挖掘更具潜力的电力用户也不利于电力企业的自身管理。这种陈旧的电力营销方式已经不能满足迫切的电力市场改革的

需求。电力用户作为整个电力市场的主角,各售电公司根据对电力用户的信用评价来选择优质电力用户,对存在信用风险的电力用户做好风险预防。电力用户也可以结合自身的信用评价与各大售电公司进行电价和服务上的议价,寻求自身利益最大化。因此,能否有效地、精准地对电

基金项目:国家自然科学基金(51977127);上海市科学技术委员会资助项目(19020500800);

上海市教育发展基金会和上海市教育委员会“曙光计划”资助

作者简介:徐宏宽(1992—),男,硕士研究生,工程师,Email: Xhk920927@163.com

通讯作者:林顺富(1983—),男,博士,教授,博士生导师,Email: shunfulin@shiep.edu.cn

力用户进行信用评价已经显得至关重要。

目前,我国在对电力用户信用评价方面的研究相对较少,且基本上停留在十多年前,很难与国内其他行业的发展相比。文献[1]通过层次分析(analytic hierarchy process, AHP)对信用等级进行评估,指标较少,缺乏科学性和全面性,研究方法粗糙。文献[2]依据电力营销人员的经验,运用主成分分析方法,建立综合函数作为评判标准,确定信用等级的方法主观性比较强,准确度不高。文献[3]既考虑了指标合理性,也考虑了评估者的主观偏好,运用基于期望值的模糊多决策方法进行信用评价。文献[4]采用模糊一致互补判断矩阵计算定性定量指标权重,从模式识别的角度建立信用评价方法。文献[5]充分结合“5C”要素建立了基于区间数和熵权法的信用评价模型。文献[6]针对数据不确定性以及专家评分的主观性问题,提出了基于区间层次分析法(interval analytic hierarchy process, IAHP)和区间熵结合的电力用户评价方法。文献[7]提出了一种基于AHP和主成分分析法(principal component analysis, PCA)的电力用户信用综合评价模型,构建了科学的用户信用评价体系,能较为准确地预测用户欠费风险。文献[8]针对单一电力用户信用评价方法不能完全反映用户的实际情况且传统的组合评估方法之间的兼容性较差,提出基于偏差熵的低压电力用户信用组合评估方法。文献[9]梳理了英、美两国的电力体制改革发展情况,根据国际经验,提出新一轮我国电力用户信用评价体系的完善和借鉴及启示。

上述成果对电力用户信用评价的研究起到了很好的推进作用,而电力用户信用评价方法的准确度需要进一步探讨:当前大多评估的方法采用改进的单一方法,很少用到集成技术,如何组合多个基分类器组建集成模型是未来探索的主要趋势。文章提出了基于层次分析与Adaboosting组合分类器的电力用户信用评价方法,采用经典的层次分析法从8个分类算法中选取4个更适合电力用户信用评价分类的备选分类算法模型,并且采用Adaboosting组合分类器对4个基分类算法器进行组合分类,从而大大提高了信用分类的准确率。该研究聚焦于应用人工智能技术研究^[10-13]电力用户的信用分级技术,具有一定的创新性和实用性,非常适用于电力用户信用评价模型的组合研究,具有重要的实际应用价值。

1 信用评价流程

根据电力用户相关数据构建信用指标体系,并进行数据预处理,从典型分类算法中遴选出合适的备选分类算法构建基分类算法,将构建的基分类算法用Adaboosting算法进行线性动态赋权组合以得到最终的强分类算法,最后将基分类与强分类的分类结果进行比较,从而证明组合分类算法的分类效果较好。基于AHP和Adaboosting组合分类器信用评价流程如图1所示。

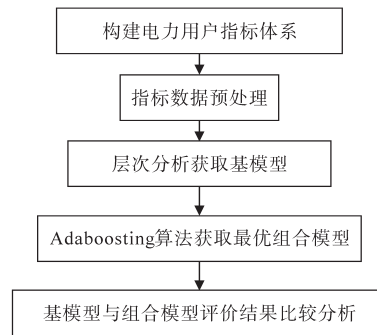


图1 基于AHP和Adaboosting组合分类器信用评价流程图

Fig.1 Credit evaluation flowchart based on AHP and Adaboosting combined classifier

2 评估指标体系建立

2.1 电力用户信用指标选取

影响电力用户信用度的指标很多,前人的研究给出了多种指标体系^[14-15]。笔者在已有研究成果基础上提出了一套电力用户信用评价指标,包含10个特征信息,如表1所示。

表1 特征字段说明

Tab.1 Feature field description

| 字段 | 变量描述 |
|-----------|--|
| 是否配合用电检查 | 配合标识,1表示不配合,0表示配合 |
| 年用电量增长率/% | 用电量增长与上一年用电量的比率 |
| 开户时长/a | 电力用户从建立电力户头迄今的时间 |
| 欠费数/次 | 电力用户累计的欠费次数 |
| 用户节能改造程度 | 电力用户节能改造分为高、中、低,节能改造程度高会降低生产成本,增加用户竞争力,提高电力用户信用等级,2代表高,1代表中,0代表低 |
| 月电费/元 | 每月缴纳的电费 |
| 账户余额/元 | 电力用户在电力账户上余额 |
| 违约用电数/次 | 电力用户各种被查处的违约用电累计的次数 |
| 催费数/次 | 电力用户被催收电费的次数 |
| 窃电数/次 | 电力用户累计的窃电次数 |

2.2 指标数据预处理

电力用户信用评价系统中涉及的指标比较

多,而各个指标所对应的属性又不尽相同,各指标所对应的量值绝对值大小可能相差巨大。当指标属性量化后,进行相关整合处理时绝对值相差大的指标属性会出现绝对值大的覆盖绝对值小的现象,这种干扰会使指标体系模型严重失真,有必要对各指标进行无量纲化处理。假设考虑对 n 个电力用户进行评估,若每个用户有 m 种评估指标,将每个指标记为 $x_{fg}(f=1,2,\dots,n;g=1,2,\dots,m)$ 。采用极值法进行指标的无量纲化处理,如下式:

$$\bar{x}_{fg} = \frac{x_{fg} - \min(x_g)}{\max(x_g) - \min(x_g)} \quad (1)$$

式中: x_{fg} 为第 f 个用户的第 g 个指标; $\max(x_g),\min(x_g)$ 分别为所有用户第 g 个指标最大值、最小值。

3 基于层次分析的信用评价模型

层次分析法最早应用于运筹学,由美国运筹学家——匹茨堡大学教授萨蒂于20世纪70年代初首次提出。层次分析法是将一个复杂的多目标决策问题作为一个系统,将目标分解为多个目标或准则,进而分解为多指标的若干层次,通过定性指标模糊量化方法算出层次单排序和总排序,以作为目标、多方案优化决策的系统方法。层次分析法通过定性与定量相结合的系统化、层次化的分析方法^[16-17],是进行权值计算时常用的工具。

3.1 信用评价分类模型备选集的确定

在众多分类模型中选择若干模型作为信用评价的备选集,即随机森林模型、决策树模型、BP神经网络模型^[18]、K最近邻模型、朴素贝叶斯模型、支持向量机模型^[19]、遗传模型、逻辑回归模型等。这些模型均为分类应用的经典模型,具备成熟的理论研究。

3.2 信用评价分类模型选取原则

4种评价模式的选取原则如下:

1) 准确性。在信用评价中,模型的准确性总是最好或最有效的。

2) 可扩展性。在公开研究中,对模型进行扩展研究的数量或程度。

3) 可行性。模型公式易于理解和实现且工具支持,模型评估结果易于评估和确认。

4) 实践性。在公开研究中,模型实际应用的数量或程度。

3.3 层次分析模型建立

模型选取的层次分析模型共有3层:1)目标层:不同数据集下的信用评级模型的综合权重。

2)准则层:四项模型选取原则;3)方案层:所有备选模型。AHP模型如图2所示。

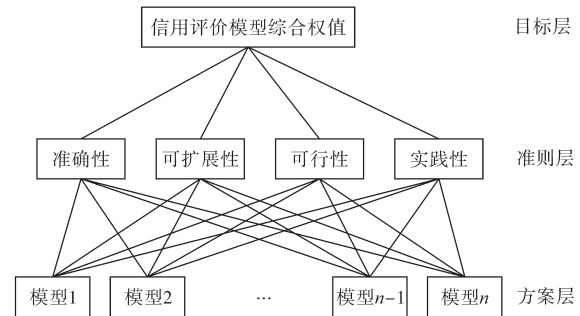


图2 信用评价选取的AHP模型

Fig.2 AHP model selected by credit evaluation

3.4 信用评价模型选取方法

1) 构造两两比较矩阵。一般来说,对于 n 个指标 A_1, A_2, \dots, A_n 进行两两的比较,可以使用成对比较矩阵。成对比较矩阵定义为

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix} \quad (2)$$

其中,数值 a_{ij} 是指标 A_i 与指标 A_j 比较相对重要性的结果,且 $a_{ij}=1/a_{ji}, a_{ii}=1$ 。上述指标比较尺度在1~9之间,如表2所示。

表2 AHP评估尺度

Tab.2 AHP evaluation scale

| 成对比较标准 | 定义 | 内容 |
|---------|------|----------------------|
| 1 | 同等重要 | 两个准则具有同等重要性 |
| 3 | 稍微重要 | A_i 比 A_j 略显重要 |
| 5 | 重要 | A_i 比 A_j 重要 |
| 7 | 明显重要 | A_i 比 A_j 明显重要 |
| 9 | 绝对重要 | A_i 比 A_j 重要很多 |
| 2,4,6,8 | | 用于上述标准中的折中值 |
| 上述数值的倒数 | | A_j 和 A_i 的重要性比较 |

2) 一致性检验。为了保证系统中使用的两两比较矩阵的有效性,需进行一致性检验。在成对比较矩阵 A 中,若 $a_{ik} \cdot a_{kj} = a_{ij}$,则称 A 为一致阵。若成对比较矩阵是一致阵,取对应于最大特征根 n 的归一化特征向量 $w=[w_1, w_2, \dots, w_n]$ 作为权向量。若成对比较矩阵不是一致阵,则应使用与其最大特征根 λ 对应的归一化特征向量作为权向量 w ,且 $Aw=\lambda w$ 。为了确定两两比较矩阵 A 的可用性,需要对其进行一致性检验,使用指标是一致性比率。

定义一致性指标如下式:

$$CI = \frac{\lambda - n}{n - 1} \quad (3)$$

通过比较 CI 和随机一致性指标 RI (见表3),可以

得到一致性比率 CR :

$$CR = \frac{CI}{RI} \quad (4)$$

表3 随机一致性指标 RI 的数值
Tab.3 Values of RI

| n | RI | n | RI |
|-----|------|-----|------|
| 1 | 0 | 6 | 1.26 |
| 2 | 0 | 7 | 1.36 |
| 3 | 0.52 | 8 | 1.41 |
| 4 | 0.89 | 9 | 1.46 |
| 5 | 1.12 | 10 | 1.49 |

如果 $CR < 0.1$, A 被认为是可以接受的一致阵,其归一化特征向量可以作为权向量。否则,应重新构造对比较矩阵,即调整 A 中各元素的取值。

4 基于 Adaboosting 的动态赋权组合建模

本节介绍如何利用 Adaboosting 算法将基于 AHP 选取的若干信用评价分类模型组合,获得线性动态赋权组合模型。Adaboosting 是一种自适应算法,其适应性主要体现在分类器的分类结果上。主要特点是降低正确分类样本权重,提高错误分类样本权重,将更改过的权重应用于下一次的迭代过程中;当进入一个新的迭代时,会添加一个新的基分类器,通过不断设置迭代次数来训练,以确定最强的分类器^[20],如图3所示。

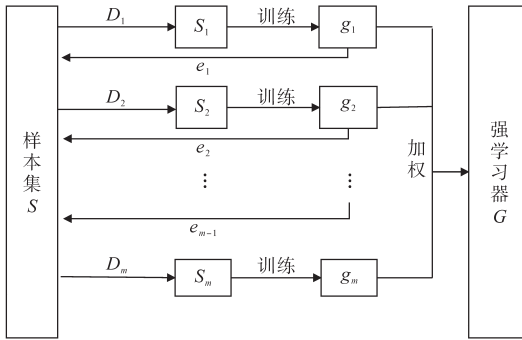


图3 Adaboosting 方法流程

Fig.3 Workflow of Adaboosting algorithm

步骤1:首先,初始化训练数据的权值分布,每一个训练样本最开始时都被赋予相同的权值,即

$$\begin{cases} D_1 = (w_{11}, w_{12}, \dots, w_{1i}, \dots, w_{1N}) \\ w_{1i} = \frac{1}{N} \quad i = 1, 2, \dots, N \end{cases} \quad (5)$$

步骤2:进行多轮迭代, m 表示第 m 轮迭代, $m=1, 2, \dots, M$ 。

选取一个当前误差率最低的弱分类器 g 作为第 m 个基本分类器 G_g , 并计算弱分类器 g_m , 该弱

分类器在分布 D_m 上的误差为

$$e_m = \sum_{i=1}^N w_{mi} 1[G_m(x_i) \neq y_i] \quad (6)$$

其中, $1[G_m(x_i) \neq y_i]$ 表示当预测结果和实际结果不一样时取值1, 否则取值0。由式(6)可知, $G_m(x)$ 在训练数据集上的误差率 e_m 就是被 $G_m(x)$ 误分类样本的权值之和。

计算弱分类器在最终分类器中所占的权重(弱分类器权重用 b 表示)如下:

$$b_m = \frac{1}{2} \ln \frac{1 - e_m}{e_m} \quad (7)$$

更新训练样本的权值分布 D_{m+1} , 因为权重更新依赖于 b , 而 b 又依赖于误差率 e , 所以可以直接将权重更新公式用 e 表示:

$$D_{m+1} = \frac{D_m(i)}{z_m} \exp[-b_m y_i G_m(x_i)] \quad (8)$$

归一化常数 z_m 如下式:

$$z_m = 2\sqrt{e_m(1 - e_m)} \quad (9)$$

1) 当样本分错时, $y_i G_m(x_i) = -1$, 错误分类样本的权值更新如下式:

$$D_{m+1} = \frac{D_m(i)}{z_m} \exp(b_m) = \frac{D_m(i)}{2e_m} \quad (10)$$

2) 当样本分对时, $y_i G_m(x_i) = 1$, 正确分类样本的权值更新如下式:

$$D_{m+1} = \frac{D_m(i)}{z_m} \exp(-b_m) = \frac{D_m(i)}{2(1 - e_m)} \quad (11)$$

步骤3:最后按弱分类器权重 b_m 组合各个弱分类器:

$$f(x) = \sum_{m=1}^M b_m G_m(x) \quad (12)$$

通过符号函数 sign 的作用, 得到一个强分类器:

$$\begin{aligned} G_{\text{final}} &= \text{sign}[f(x)] \\ &= \text{sign}\left[\sum_{m=1}^M b_m G_m(x)\right] \end{aligned} \quad (13)$$

5 电力用户信用度评价实例

依据 AHP 方法从8个经典分类模型中选出排名靠前的4个备选模型。采用 Adaboosting 算法组合备选, 并将组合模型与4个经典分类模型的结果进行对比, 验证组合建模方法的有效性与可行性。

5.1 数据预处理

选取了浙江省某1000名电力用户的数据进行信用评价。用户原始数据如表4所示, 预处理数据如表5所示。

表4 电力用户信用指标数值

Tab.4 Original value of the indexes

| 用户序号 | 是否配合用电检查 | 年用电量增长率/% | 开户时长/a | 欠费数/次 | 用户节能改造程度 | 月平均电费/万元 | 账户余额/万元 | 违约用电数/次 | 催费数/次 | 窃电数/次 |
|------|----------|-----------|--------|-------|----------|----------|---------|---------|-------|-------|
| 001 | 1 | -17.11 | 3 | 5 | 1 | 11.24 | -5 | 3 | 12 | 2 |
| 002 | 0 | 5.52 | 4 | 0 | 1 | 22.76 | 30 | 0 | 0 | 0 |
| 003 | 0 | 1.87 | 5 | 1 | 0 | 45.72 | 50 | 0 | 2 | 0 |
| 004 | 0 | 8.56 | 5 | 0 | 1 | 2.66 | 3 | 0 | 0 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 999 | 0 | 4.45 | 5 | 0 | 1 | 3.72 | 5 | 0 | 0 | 0 |
| 1000 | 1 | 3.56 | 5 | 2 | 0 | 3.28 | 1 | 1 | 3 | 0 |

表5 电力用户信用指标数据预处理

Tab.5 Data preprocessing of power user credit index

| 用户序号 | 是否配合用电检查 | 年用电量增长率/% | 开户时长/a | 欠费数/次 | 用户节能改造程度 | 月平均电费/万元 | 账户余额/万元 | 违约用电数/次 | 催费数/次 | 窃电数/次 |
|------|----------|-----------|--------|-------|----------|----------|---------|---------|-------|-------|
| 001 | 1 | 0 | 0.200 | 1 | 1 | 0.199 | 0 | 1 | 1 | 0.250 |
| 002 | 0 | 0.881 | 0.267 | 0 | 1 | 0.466 | 0.636 | 0 | 0 | 0 |
| 003 | 0 | 0.723 | 0.333 | 0.2 | 0 | 1 | 1 | 0 | 0.167 | 0 |
| 004 | 0 | 0.558 | 0.333 | 0 | 1 | 0 | 0.145 | 0 | 0 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 999 | 0 | 0.843 | 0.333 | 0 | 1 | 0.026 | 0.182 | 0 | 0 | 0 |
| 1000 | 1 | 0.809 | 0.333 | 0.4 | 0 | 0.014 | 0.109 | 0.333 | 0.252 | 0 |

5.2 基于AHP的备选模型选取

各层次的两两比较矩阵结合多名专家学者根据多年的研究经验给出打分情况。表6、表7给出了专家对电力用户信息分析后给出选用适合电力用户信用分类的各个基分类算法的评分表。表6给出准则层打分,表7给出准确性打分。可扩展性、可行性与实践性的打分情况类似,这里不做赘叙。据此可以得到各对应的成对比较矩阵。

表6 准则层打分

Tab.6 Scores for criterion layer

| 准则层 | 准确性 | 可扩展性 | 可行性 | 实践性 |
|------|-----|------|-----|-----|
| 准确性 | 1/1 | 9/1 | 2/1 | 4/1 |
| 可扩展性 | 1/9 | 1/1 | 1/4 | 1/3 |
| 可行性 | 1/2 | 4/1 | 1/1 | 2/1 |
| 实践性 | 1/4 | 3/1 | 1/2 | 1/1 |

表7 准确性打分

Tab.7 Scores for accuracy

| 准确性 | 随机森林 | 逻辑回归 | 决策树 | BP神经网络 | K最近邻 | 遗传模型 | 支持向量机 | 朴素贝叶斯 |
|--------|------|------|-----|--------|------|------|-------|-------|
| 随机森林 | 1/1 | 1/2 | 1/5 | 1/6 | 2/1 | 1/3 | 1/7 | 1/4 |
| 逻辑回归 | 2/1 | 1/1 | 1/4 | 1/5 | 3/1 | 1/2 | 1/6 | 1/3 |
| 决策树 | 5/1 | 4/1 | 1/1 | 1/2 | 6/1 | 3/1 | 1/3 | 2/1 |
| BP神经网络 | 6/1 | 5/1 | 2/1 | 1/1 | 7/1 | 4/1 | 1/2 | 3/1 |
| K最近邻 | 1/2 | 1/3 | 1/6 | 1/7 | 1/1 | 1/4 | 1/8 | 1/5 |
| 遗传模型 | 3/1 | 2/1 | 1/3 | 1/4 | 4/1 | 1/1 | 1/5 | 1/2 |
| 支持向量机 | 7/1 | 6/1 | 3/1 | 2/1 | 8/1 | 5/1 | 1/1 | 4/1 |
| 朴素贝叶斯 | 4/1 | 3/1 | 1/2 | 1/3 | 5/1 | 2/1 | 1/4 | 1/1 |

表6对应的成对比较矩阵为

$$A = \begin{bmatrix} 1 & 9 & 2 & 4 \\ 1/9 & 1 & 1/4 & 1/3 \\ 1/2 & 4 & 1 & 2 \\ 1/4 & 3 & 1/2 & 1 \end{bmatrix} \quad (14)$$

表7的对应矩阵类似,经计算各个打分的CR值如表8所示。

表8 各层打分CR值

Tab.8 CR value of each layer

| 名称 | CR值 | 名称 | CR值 |
|------|---------|-----|---------|
| 准则层 | 0.006 1 | 可行性 | 0.026 8 |
| 准确性 | 0.029 6 | 实践性 | 0.028 3 |
| 可扩展性 | 0.036 2 | | |

由表8可知,所有的CR值均小于0.1,都能通过一致性检验。

各模型权值如表9所示。从表9可以看出,综合评估值排在前4位的模型依次是决策树模型、BP神经网络模型、支持向量机模型以及朴素贝叶斯模型,而随机森林模型、逻辑回归模型、K最近邻模型以及遗传模型在研究中的综合权值

表9 模型权值

Tab.9 Weight of models

| 模型 | 权值 | 模型 | 权值 |
|----------|-------|---------|-------|
| 随机森林模型 | 0.033 | K最近邻模型 | 0.023 |
| 逻辑回归模型 | 0.050 | 遗传模型 | 0.073 |
| 决策树模型 | 0.158 | 支持向量机模型 | 0.327 |
| BP神经网络模型 | 0.228 | 朴素贝叶斯模型 | 0.108 |

较低。该结果与各个模型在现实中的应用结果类似,因此,依据提出的AHP模型选择法,可有效选出最为合适的4个备选模型,即决策树模型、BP神经网络模型、支持向量机模型以及朴素贝叶斯模型。

5.3 Adaboosting最优组合模型

选取800条数据作为训练集,剩下200条作为测试集,设定训练次数为 $M=50$,Adaboosting最优模型组合(Adaboosting combinatorial model, $AMCM_{best}$)过程如下:

1)初始化训练数据的权重 $D_m(i) = 1/800$ 。

2)取已知的4个模型中误差率比较小的支持向量机作为第一个基本分类器,它的误差率为 $e_1=0.231$,根据误差率计算 G_1 的权重 $a_1=0.601$,这个值代表 G_1 在最终分类器中所占的权重。然后更新训练样本数据的权值分布,对于分类正确训练样本的权值更新为 $D_2=0.000813$,错误训练样本的权值更新为 $D_2(i)=0.00271$ 。依次根据该过程迭代。

3)迭代完成后组合基本分类器形成一个强分类器模型的函数构建完成如下:

$$F(x)=\text{sign}[0.601G_1(x)+0.647G_2(x)+0.424G_3(x)+\dots] \quad (15)$$

5.4 分类模型评估

分类模型的评价指标有很多,使用准确率、曲线下面积(area under curve, AUC)和受试者操作曲线(receiver operator characteristic curve, ROC)来评估。4个备选模型与 $AMCM_{best}$ 的准确率与AUC值如表10所示,其中加粗的数字表示最佳结果。

表10 分类结果比较

Tab.10 Comparison of classification results

| 模型 | 准确率/% | AUC值 |
|---------------|--------------|----------------|
| 决策树模型 | 83.28 | 0.866 9 |
| BP神经网络模型 | 84.89 | 0.868 0 |
| 支持向量机模型 | 88.61 | 0.917 3 |
| 朴素贝叶斯模型 | 79.35 | 0.833 2 |
| $AMCM_{best}$ | 93.38 | 0.925 0 |

由表10可以发现,提出的 $AMCM_{best}$ 模型获得的准确率与AUC值都比较高。这表明 $AMCM_{best}$ 分类效果较好,可以有效改进单个备选模型的准确率与AUC值,这是因为AdaBoosting算法可以反复训练这些备选模型,从而在最终组合模型AMCM中动态最优化它们权值。为了更加直观地比较5个模型的分​​类效果,用ROC曲线将预测

结果显示出来,如图4所示。

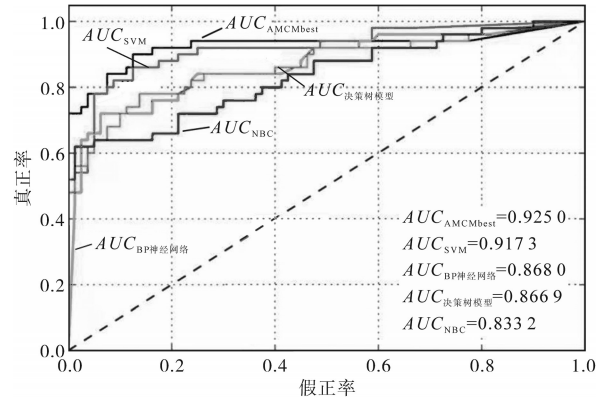


图4 ROC曲线

Fig.4 ROC curves

6 结论

采用基于AHP的电力用户信用评价模型选取方法选出最为合适的备选模型,进而依据Adaboosting算法,对4个备选模型进行组合建模,经过多次组合试验进行学习训练,从而确定最优的 $AMCM_{best}$ 模型。最后将 $AMCM_{best}$ 模型与4个备选模型进行实例应用分析,可以得出 $AMCM_{best}$ 模型准确率最高且AUC值高于4个备选模型。实验结果表明,基于Adaboosting算法组合建模确定的 $AMCM_{best}$ 模型在电力用户信用评价问题上分类效果更佳。

参考文献

- [1] 陈婷婷,姜健,张丽华. 基于Fuzzy AHP的电力客户信用等级评估[J]. 科技与产业, 2013, 13(5): 107-109.
CHEN Tingting, JIANG Jian, ZHANG Lihua. Evaluation of customer credit based on fuzzy AHP evaluation model[J]. Science Technology and Industry, 2013, 13(5): 107-109.
- [2] 瞿斌,李存斌,田惠英. 工业用电客户信用综合评估指标体系的构建方法[J]. 电网技术, 2007, 31(1): 75-79.
QU Bin, LI Cunbin, TIAN Huiying. Construction and methodology of comprehensive evaluation system for credit of industrial electricity customers[J]. Power System Technology, 2007, 31(1): 75-79.
- [3] 任保瑞. 基于模糊综合评判的电力客户信用评估[J]. 现代电力, 2011, 28(4): 90-94.
REN Baori. Credit evaluation of electricity customers based on fuzzy synthetic evaluation method[J]. Modern Electric Power, 2011, 28(4): 90-94.
- [4] 李江,卢毅勤. 电力客户信用等级评估体系构建研究[J]. 华北电力大学学报, 2007, 4(4): 9-12.
LI Jiang, LU Yiqin. Research on the construction of credit rating evaluation system of electricity clients[J]. Journal of North Chi-

- na Electric Power University, 2007, 4(4):9-12.
- [5] 许东辉. 电力客户信用评估研究[D]. 北京:华北电力大学, 2011.
XU Donghui. Research on credit evaluation of electric power customers[D]. Beijing: North China Electric Power University, 2011.
- [6] 王炼,孙和平,谢振平,等. 一种基于缴费积极性的电力客户信用评估模型[J]. 计算机工程与应用, 2016, 52(22): 253-259.
WANG Lian, SUN Heping, XIE Zhenping, et al. Electricity consumer credit evaluation model based on charge payment proactivity[J]. Computer Engineering and Applications, 2016, 52(22):253-259.
- [7] 杨铮宇,田园,李申章. 基于主成分分析和层次分析的高压电力用户信用评价模型研究[J]. 云南大学学报(自然科学版), 2020, 42(S2):6-12.
YANG Zhengyu, TIAN Yuan, LI Shenzhang. The research on comprehensive credit evaluation model of high voltage electric power users based on principal component analysis and analytic hierarchy process[J]. Journal of Yunnan University (Natural Science Edition), 2020, 42(S2):6-12.
- [8] 马文,杨铮宇,张梅. 基于偏差熵的低压电力用户信用组合评价研究[J]. 云南大学学报(自然科学版), 2020, 42(S2):1-5.
MA Wen, YANG Zhengyu, ZHANG Mei. The research on credit combination evaluation for low-voltage power users based on deviation entropy[J]. Journal of Yunnan University (Natural Science Edition), 2020, 42(S2):1-5.
- [9] 张龙,曾鸣,沈红宇,等. 不同电力体制改革背景下国际标杆电网评价体系研究及借鉴[J]. 电力建设, 2015, 36(11):148-153.
ZHANG Long, ZENG Min, SHEN Hongyu, et al. International benchmarking grid assessment systems research and reference under different electricity system reform background[J]. Electric Power Construction, 2015, 36(11):148-153.
- [10] 李明峰,贾修一. 基于多分类器集成学习的中文反语识别技术[J]. 计算机与数字工程, 2018, 46(9):1790-1795.
LI Mingfeng, JIA Xiuyi. Chinese irony recognition based on ensemble learning of multiple classifiers[J]. Computer & Digital Engineering, 2018, 46(9):1790-1795.
- [11] 黄颖坤,金炜东,余志斌,等. 基于深度学习和集成学习的辐射源信号识别[J]. 系统工程与电子技术, 2018, 40(11): 2420-2425.
HUANG Yinkun, JIN Weidong, YU Zhibin, et al. Radar emitter signal recognition based on deep learning and ensemble learning[J]. Systems Engineering and Electronics, 2018, 40(11): 2420-2425.
- [12] 李婉婉. 基于卷积神经网络和集成学习的材质识别和分割方法研究[D]. 北京:北京交通大学, 2018.
LI Wanwan. Ensemble learning for material recognition and segmentation with convolutional neural networks[D]. Beijing: Beijing Jiaotong University, 2018.
- [13] COSTA E O, SOUZA G, POZO A, et al. Exploring genetic programming and boosting techniques to model software reliability[J]. IEEE Transactions on Reliability, 2007, 56(3):422-434.
- [14] 胡亚红,方豪达,江大川,等. 基于AHP和k-means算法的电力用户信用度评估[J]. 浙江工业大学学报, 2018, 46(5): 515-521.
HU Yahong, FANG Haoda, JIANG Dachuan, et al. Credit evaluation of power consumers based on AHP and k-means algorithms[J]. Journal of Zhejiang University of Technology, 2018, 46(5):515-521.
- [15] 周嵩. 浅谈电力用户用电信息采集系统大数据在用户行为研究中的应用[J]. 华东科技(学术版), 2014(8):271-271, 283.
ZHOU Song. Talking about the application of big data in the research of user behavior in the electricity user information collection system[J]. East China Technology (Academic Edition), 2014(8):271-271, 283.
- [16] 张炳江. 层次分析法及其应用案例[M]. 北京:电子工业出版社, 2014.
ZHANG Bingjiang. Ceng ci fen xi fa ji qi ying yong an li[M]. Beijing: Publishing House of Electronics Industry, 2014.
- [17] 李海英,李媛媛,宋建成. 基于模糊控制和AHP的矿井变频调速通风系统[J]. 电气传动, 2012, 42(1):17-21.
LI Haiying, LI Yuanyuan, SONG Jiancheng. Mine frequency conversion speed regulation ventilation system based on fuzzy control and AHP[J]. Electric Drive, 2012, 42(1):17-21.
- [18] 向楠,张向文. 电动汽车再生制动模糊神经网络控制策略研究[J]. 电气传动, 2020, 50(7):86-91.
XIANG Nan, ZHANG Xiangwen. Research on fuzzy neural network control strategy for regenerative braking of electric vehicles[J]. Electric Drive, 2020, 50(7):86-91.
- [19] 张育炜,石琦,武恩光. 基于支持向量机GIS局放小波包能量谱故障诊断[J]. 电气传动, 2020, 50(9):99-105, 114.
ZHANG Yuwei, SHI Qi, WU Enguang. Fault diagnosis based on support vector machine GIS partial discharge wavelet packet energy spectrum[J]. Electric Drive, 2020, 50(9):99-105, 114.
- [20] 康铎,许继平,赵峙尧,等. 基于Adaboost算法的水质组合预测方法研究[J]. 计算机测量与控制, 2018, 26(8):41-45.
KANG Duo, XU Jiping, ZHAO Zhiyao, et al. Research on water quality combination forecasting method based on Adaboost algorithm[J]. Computer Measurement & Control, 2018, 26(8):41-45.

收稿日期:2021-05-19

修改稿日期:2021-05-30