

一种基于改进 RBF 神经网络的学生信用评价模型与算法研究

Research on a Student Credit Evaluation Model and Algorithm Based on Improved RBF Neural Network

张米

Mi Zhang

绵阳职业技术学院 中国·四川 绵阳 621000

Mianyang Vocational and Technical College, Mianyang, Sichuan, 621000, China

摘要: 针对高校学生的信用评价问题, 提出一种基于改进 RBF 神经网络的学生信用评价模型。首先, 基于学生的角色特征建立多维度细粒度的静态评价指标体系。其次, 通过粒子群算法对径向基 RBF 神经网络进行优化, 建立高校学生信用评价模型。最后, 通过真实的数据集验证, 提出的 PSO-RBF 模型收敛速度更快、泛化能力更强, 具有更好的预测效果, 可以有效地对高校学生的信用进行评估, 在人才培养、用人招聘等方面有着较高的应用价值。

Abstract: Aiming at the problem of credit evaluation of college students, this paper proposes a student credit evaluation model based on improved RBF neural network. Firstly, a multi-dimensional and fine-grained static evaluation index system is established based on the role characteristics of students. Secondly, the radial basis RBF neural network is optimized by particle swarm algorithm to establish a credit evaluation model for college students. Finally, through the verification of real data sets, the proposed PSO-RBF model has faster convergence speed, stronger generalization ability, and better prediction effect. It can effectively evaluate the credit of college students, and has high application value in talent training and recruitment.

关键词: 网络; 信用评价; 算法研究

Keywords: network; credit evaluation; algorithm research

课题项目: 绵阳市哲学社会科学重点研究基地——绵阳职业教育研究中心资助“高职院校学生信用评价系统的研究与应用” (项目编号: MZY21B10)。

DOI: 10.12346/sde.v4i8.7010

1 引言

诚信是支撑社会道德的支点, 也是维护社会良好秩序的规矩。孔子曾曰: “人而无信, 不知其可也。” 诚信是亘古不变的追求, 也是个人自我的约束。目前无论从国家还是社会层面, 社会信用体系已建立得相对完善, 如国家、省级、市级层次的“信用中国”“信用四川”“信用绵阳”等, 而建立健全规范的个人信用体系是建立完善社会信用体系的基础, 其中高校学生个人信用评价体系是社会信用体系下的重要组成部分, 不仅关乎学生的健康成长和人格的健全, 而且还关系到国家的建设, 关系到“中国梦”的实现。“小胜

靠智, 大胜靠德”, 而诚信是最基础的道德标准, 进行科学合理的学生信用评估, 有利于培养学生的信用意识和诚信品质, 提高学生的综合素质; 有利于学校的学风、校风建设, 提高学校的管理质量, 使不良现象得到有效遏制; 有利于社会上用人单位招聘工作的开展, 以及国家助学贷款工作的良性运转。因此, 随着中国个人信用体系建设的逐渐深入, 研究高校学生信用评价体系也愈显其重要性。

目前, 中国对于个人信用评估的研究主要集中在不同模型在个人信用评估中的应用、模型算法的优化以及组合模型的应用等方面。姜明辉先后利用分类树、线性判别分析法等

【作者简介】张米 (1992-), 女, 中国江苏徐州人, 硕士, 助教, 从事数据挖掘与人工智能研究。

方法对建立个人信用评估模型进行了探索。石庆焱在前人研究的基础上,利用国巧某商业银行的信用卡数据,比较了多种分类方法,包括判别分析、决策树、逻辑回归、神经网络算法以及线性规划法。郭春香考察了两种朴素贝叶斯分类器在信用评估中的适用性,并通过实证研究将朴素贝叶斯与神经网络进行对比。

以上研究主要针对个人征信体系,缺少高校学生角色特征,构建出符合高校学生特点的信用指标体系。此外,BP神经网络在构建网络结构时,存在收敛速度慢、易陷入局部极小值等固有缺陷,而RBF神经网络具有逼近精度高、收敛速度快等特点,因此作为本研究内容的主要算法,但其网络的初始值是随机生成的,且对训练的效果具有较大的影响,因此优化初始值十分必要。所以,论文根据大学生信用评价指标体系的特点,利用粒子群算法(Particle Swarm Optimization, PSO)在RBF神经网络进行了改进,提出了一种基于改进RBF神经网络的学生信用评价模型。

2 静态评价指标体系的选取

评价指标体系的选取是影响评价模型准确度的关键因素之一。从高校和学生的实际情况和特点出发,本着全面性、层次性、可操作性、定性与定量指标相结合的原则,特制定了如表1所示的信用历史、行为偏好、学习状况以及身份特征4个一级指标,一级指标下又细分为若干个二级指标,共计14个二级指标组成了高校学生信用评价指标体系。

表1 高校学生信用评价指标体系表

一级指标	二级指标
信用历史	图书归还, 学费缴纳, 助学贷款, 考试作弊, 违纪处理
行为偏好	一卡通消费, 图书借阅行为偏好, 出勤率
学习状况	专业成绩, 竞赛获奖, 荣誉证书
身份特征	政治面貌, 是否贫困生, 担任职务

依据评价理论和方法的相关内容,高校学生信用评价等级选取极好、优秀、良好、中等、较差5个等级。

3 RBF神经网络的信用评价模型构建

3.1 RBF基本原理

RBF神经网络是一种具有单隐藏层的前馈神经网络,如图1所示,包括输入层、隐藏层和输出层。其三层前馈网络结构中,输入层包含 $X=[x_1, x_2, \dots, x_n]^T$ 的输入向量;输入层到隐藏层为非线性映射,采用的基函数是高斯函数 a_j ,如式(1)所示,其中 $C_j=[C_{j1}, C_{j2}, \dots, C_{jm}]^T$ 为网络的第 j 个节点中心向量, $\sigma=[\sigma_1, \sigma_2, \dots, \sigma_m]^T$ 为基宽向量;隐藏层到输出层为线性映射,如式(2)所示,其中 $W=[w_1, w_2, \dots, w_m]^T$ 为网络的输出层连接权重向量, y_m 为学生信用评价指标经过RBF神经网络的预测输出,与训练集对应标签 y 比较,定义预测误差如式(3)所示。

$$\alpha_j = \exp\left(-\frac{|x-C_j|^2}{2\sigma_j^2}\right) (j=1,2,\dots,m) \quad (1)$$

$$y_m(k) = W^T H \quad (2)$$

$$E(k) = y(k) - y_m(k) \quad (3)$$

在网络的构建过程中,根据梯度下降原则,对连接权值、节点中心以及节点基宽等参数进行迭代调整。

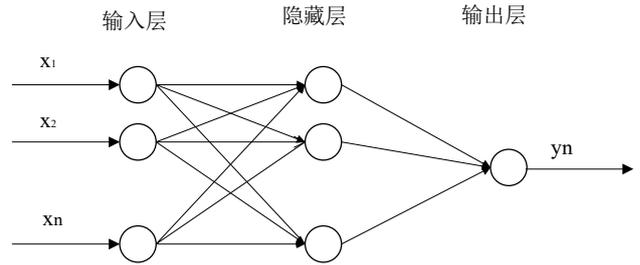


图1 RBF神经网络结构

3.2 基于粒子群优化RBF算法 PSO-RBF

3.2.1 RBF神经网络存在的问题

在前馈多层神经网络的RBF算法中,基宽向量初始值 σ ,节点中心向量初始值,输出层连接权值初始值是在一个固定范围内按均匀分布随机产生的。初始值的选择对于模型质量和网络收敛速度均有一定程度的影响,如果初始值范围选择不当,学习过程一开始就可能进入“假饱和”现象,甚至进入局部极小点,影响RBF网络性能。所以,初始值的选择随网络结构模式和训练样本不同而有所差别,一般应视实际情况而定。基于此,为了提高RBF神经网络算法的准确率和泛化能力,论文从RBF网络的初始值着手进行优化改进^[1]。

3.2.2 粒子群优化RBF神经网络算法

上一小节简要阐述了RBF神经网络算法中初始值对整个网络模型的影响,为了提高模型质量和收敛速度,论文采用粒子群优化算法对模型进行改进,寻找RBF神经网络的最优初始值。粒子群优化算法实现过程如表2所示。

表2 粒子群优化算法实现过程

PSO的实现过程如下:

- (1) 初始化 PSO 的种群规模 N , 种群粒子位置 $x_{id}(t)$ 以及速度 $v_{id}(t)$
- (2) 通过优化函数计算每个粒子的适应度值 $fit(i)$
- (3) 比较每个粒子的适应度值 $fit(i)$ 和 p_{best} , 若 $fit(i) > p_{best}$, 那么用 $fit(i)$ 代替粒子的个体极值 p_{best}
- (4) 然后再比较每个粒子的适应度值 $fit(i)$ 和全局极值 p_{best} , 若 $fit(i) > g_{best}$, 那么用 $fit(i)$ 代替粒子的全局极值 g_{best}
- (5) 通过式(4)和式(5)更新粒子位置及速度
- (6) 如果误差足够小或者达到算法最大循环次数就退出, 否则返回步骤(2)

$$v_{id}(t+1) = w \cdot v_{id}(t) + c_1 r_1 (p_{best} - x_{id}(t)) + c_2 r_2 (g_{best} - x_{id}(t)) \quad (4)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (5)$$

在利用 PSO 算法优化 RBF 神经网络时，通过定义一定数量的粒子个数，即种群数量，其中每个粒子表示一个 RBF 神经网络，每个粒子中包含 RBF 神经网络中的节点中心、节点基宽以及输出层连接权值。因论文是学生信用分类问题，所以以给定训练样本集的网络输出误差 $E(k)$ 作为粒子的适应度函数 $fitness$ ，误差越小则表明粒子在搜索中具有更好的性能，粒子在权值空间内移动搜索使得网络输出层的误差最小，通过不断迭代，改变粒子的速度也就是更新网络参数，最终找到最优粒子的最优位置，即 RBF 网络的最优初始值。

3.3 基于 PSO-RBF 算法的信用评价模型

信用评价模型以学生的信用评价指标作为样本输入，论文根据高校学生信用评价的特点，将 $(x_1, x_2, \dots, x_{14})$ 14 个经预处理后的评价指标规范值，作为网络输入层的节点输入；信用评价等级类别为输出，对应高校学生信用评价的五个评

价等级取值概率。利用这些已知类别的样本对 PSO-RBF 神经网络进行训练和参数确定，最后利用训练好的神经网络预测一个新样本的类别。具体流程如图 2 所示。

4 实验论证

4.1 数据获取与处理

论文的信用数据来自高校教务系统、学工管理系统、一卡通消费系统、图书借阅系统和助学贷款平台等多处来源，在数据采集过程中注重学生的个人隐私性保护，做相应脱敏处理，不存储与个人信用无关或法律法规禁止存储的信息^[2]。

在 PSO-RBF 算法中，输入层到隐藏层的映射函数采用的是高斯函数，因此为满足高斯函数的输入要求，需要对高校学生信用评价的样本指标数据进行预处理，包括指标属性值规范以及标准化处理（见表 3）。

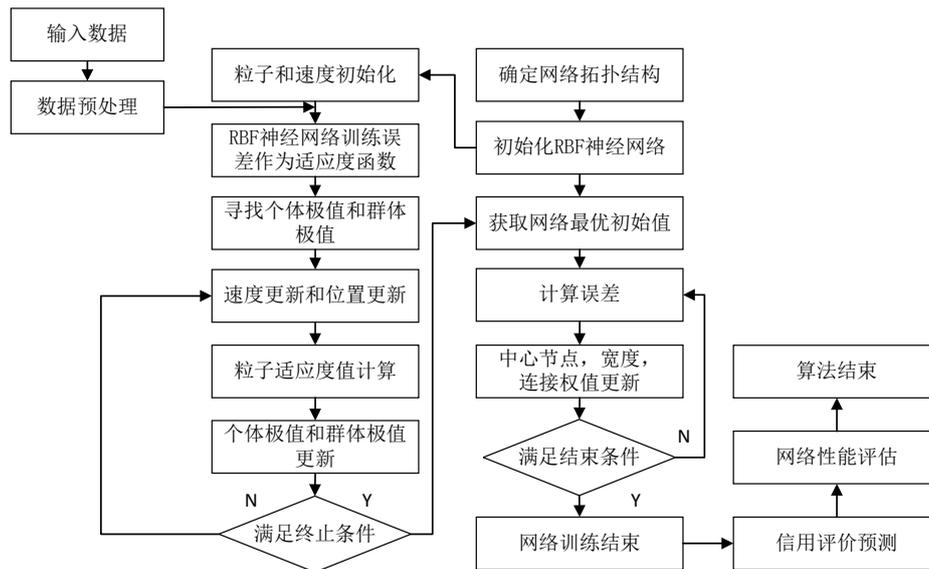


图 2 基于 PSO-RBF 算法的信用评价模型工作流程图

表 3 指标属性值规范表

指标名	属性值规范要求
图书归还	按时 =8, 偶尔 =4, 无记录 =6
学费缴纳	按时 =8, 积极补缴 =5, 恶意拖欠 =0
助学贷款还贷	按时 =8, 偶尔 =4, 拖欠 =0, 无记录 =6
考试作弊	警告处分 =-1, 记过 =-2, 留校查看 =-3, 开除 =-4
违纪处理	警告处分 =-1, 记过 =-2, 留校查看 =-3, 开除 =-4
一卡通消费	经常消费 =8, 偶尔消费 =4, 无记录 =0
图书借阅行为偏好	经常借阅 =8, 偶尔借阅 =4, 从不借阅 =0
出勤率	全勤 =8, 偶尔缺勤 =4, 严重缺勤 =0
专业成绩	优 =10, 良 =8, 中 =6, 较差 =4, 差 =2
竞赛获奖	国家级竞赛 =4, 市级竞赛 =3, 校级竞赛 =2, 无 =0
荣誉证书	国家级证书 =4, 市级证书 =3, 校级证书 =2, 无 =0
政治面貌	党员 =4, 预备党员 =3, 团员 =2, 其他 =1
是否贫困生	是 =4, 否 =2
担任职务	校级干部 =4, 院级干部 =3, 班级干部 =2, 无 =0

通过专家标注,选取 5000 个已知信用等级测试样本,对信用评价模型进行训练,确定网络模型的参数;选取其他 1000 组样本利用训练好的 PSO-RBF 神经网络进行预测。

4.2 实验结果与分析

为了比较粒子群优化后的 PSO-RBF 神经网络与传统 RBF 神经网络的性能差异,构建了网络结构相同、训练样本一致的两个网络。经实验,PSO-RBF 神经网络与 RBF 神经网络的训练误差对比曲线如图 3 所示。

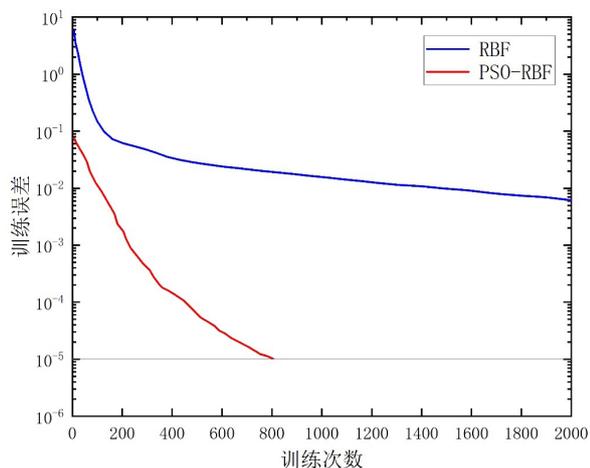


图 3 PSO-RBF 神经网络与 RBF 神经网络训练误差对比图

从图 3 可以看出,RBF 神经网络的收敛速度较快,曲线没有出现局部振荡,即没有出现局部极小点。经过 801 次迭代,达到目标训练误差的精度,其中最大误差为 0.00001,最小误差为 0,误差均小于 10^{-5} ,满足精度要求,网络训练结果较好。与此同时,RBF 神经网络在训练的前 300 次中,下降速度相较于 PSO-RBF 神经网络较为缓和,且在后续的训练过程中,趋于平缓,训练误差趋于稳定不变,由此可见,PSO-RBF 神经网络的收敛效果和训练精度均优于 RBF 神经网络^[3]。

为多方面评价 PSO-RBF 神经网络的性能,将带标签的 350 个测试样本按照逐次递增 50 个测试样本的规律依次输入到训练好的 PSO-RBF 网络、RBF 网络、BP 神经网络进行测试,输出对应的测试样本的正确率,绘制成图 4 所示的折线图。

从图 4 可以看出,PSO-RBF 网络随着测试样本的增加,正确率急剧提高,且正确率最高可达 95.8%,相反地,RBF 网络、BP 神经网络随着测试样本的增加,正确率缓慢提高,整体维持在窄幅区间震荡。因此,PSO-RBF 神经网络的泛

化能力更强,更适用于较大规模人群的信用评价。

综上所述,实验结果表明经过粒子群算法优化的 PSO-RBF 神经网络的个人信用评价模型的收敛速度更快,泛化能力更强,精确度更高。

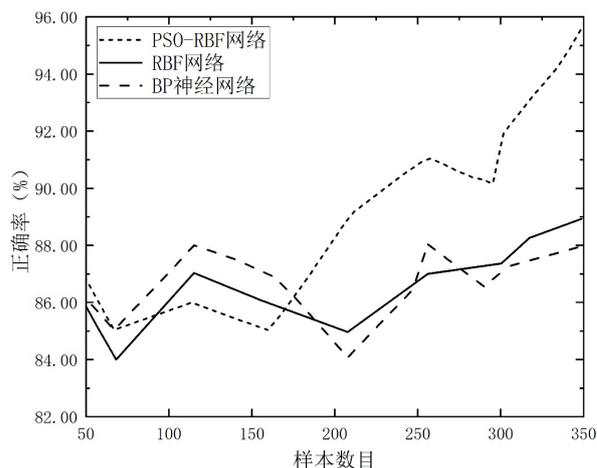


图 4 几种网络正确率对比图

5 结语

随着高校学生信用评价模型的广泛研究,论文提出了一种粒子群优化 RBF 神经网络的信用评价模型 PSO-RBF。从高校和学生的实际情况和特点出发,建立了多维度细粒度的静态评价指标体系,通过真实的数据集验证,PSO-RBF 模型的收敛速度、泛化能力都优于 RBF 模型,达到了较好的训练效果^[4]。同时,由于高校学生信用评价指标体系还未有统一标准,在数据取得、指标建立等方面还可以继续深入研究,且论文模型中 RBF 网络和 PSO-RBF 网络的隐含层数均设置为单层,可以考虑设置多层对模型进一步优化,以此提高模型的预测精度。

参考文献

- [1] 蒋维.基于改进 PSO-BP 神经网络的个人信用评价模型及算法研究[D].成都:电子科技大学,2017.
- [2] 王磊.基于 BP 神经网络的 P2P 网络借贷个人信用评价模型[J].现代商贸工业,2019,40(3):3.
- [3] 张晓航,李政,朱晓明,等.基于 RBF 神经网络的可信加密流量分类方法[J].计算机与现代化,2022(2):7.
- [4] 张启凡,王永忠,马俊逸.基于 RBF 神经网络的短时降水量预测方法研究[J].计算机与数字工程,2022,50(4):5.