

构建基于 PSO-BP 网络的电信客户 信用度评估模型

张 奇, 黄卫东

(南京邮电大学, 江苏 南京 210003)

摘 要:随着中国电信行业的迅速发展,提高电信客户信用评估的准确度和科学性及其重要。文中针对 BP 神经网络的不足之处,研究了将粒子群算法应用于 BP 神经网络的优化问题,主要是将 PSO 算法运用于优化 BP 神经网络的权值,并通过对电信客户行为属性的统计分析,将其作为客户信用度预测评估的依据,建立了信用度评估模型,并用 Matlab 软件及其神经网络工具进行仿真和计算。实验表明,新模型采用的算法具有收敛速度快,预测精度高的优点,是一种有效的电信用户信用度评估模型。

关键词:粒子群优化;BP 神经网络;信用度

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2012)12-0146-03

Construction of Credit Evaluation Model for Telecommunication Clients Based on PSO-BP Neural Network

ZHANG Qi, HUANG Wei-dong

(Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: With the rapid development of telecoms in China, it is very important to promote the accuracy and scientific of telecommunication clients' credit rating. In this paper, the deficiency of Back Propagation (BP) neural network is optimized based on Particle Swarm Optimization (PSO) algorithm is studied aimed at the deficiency of BP neural network, PSO algorithm is used for optimizing parameters and threshold values of BP neural network. A credit evaluation model is built according to the statistic analysis of telecommunications clients' behavior character. Matlab software and its neural network box is used to simulate and compute. The experiment results show that PSO-BP algorithm works with quicker convergence rate and the higher forecast precision. This model is a effective one which is content with the demand of credit evaluation of telecommunications clients.

Key words: PSO; BP neural network; credit evaluation

0 引 言

随着中国电信业的迅猛发展,中国电信行业把“客户满意度、经营收益和经营风险”视作其运营管理的核心,这就意味着各个运营商希望能在得到最大的经营效益的同时将风险降到最低,同时也希望能够吸引更多的客户^[1]。所以电信运营商希望能够根据客户的信用度及其消费行为特征对于不同的客户采取不同的话费控制措施,例如对于低信用度用户,优先纳入系统进行话费监控等调控手段,以降低企业总体欠费水平,并且采用一些手段吸引高信用度有价值的客户。

为了实现对不同客户采取不同方式的个性化服务,对每个客户的信用状况进行科学的分析非常必要。

随着计算数学的发展,统计学及运筹学等计量方法越来越多地被运用到信用评分领域,大部分的信用评分模型都使用其中的一种方法,或将几种方法结合起来使用。但这些大多从国外金融业信用系统改进而来,姚琦云^[2]建立用户信用度等级通过多途径提高电信欠费催缴率的方面,但该思路计算的倾向是比较注重缴费情况;陈大峰^[3]利用模糊数学中的综合判定方法对移动系统中客户信用度评定进行了研究,但该文并没有进行实际数据的数值实验,所以该方法的有效性有待验证。

文中的研究思路是针对基本 BP 算法收敛速度慢和易陷入局部极值的局限,将 PSO 算法引进 BP 神经网络,用以修正神经网络的权值和阈值,并挖掘隐含在电信客户属性中的信用要素,进行统计分析的基础上,

收稿日期:2012-04-16;修回日期:2012-07-20

基金项目:江苏省自然科学基金资助项目(BK2010524);南京邮电大学青蓝基金资助项目(NY210059)

作者简介:张 奇(1980-),女,江苏无锡人,讲师,研究方向为人工智能、网络舆情等;黄卫东,教授,研究方向为应急管理、网络舆情。

建立电信用户的信用度模型。

1 建立信用度评估指标体系

文中在建立指标体系时,借鉴了一些比较成熟的个人信用的指标体系,并结合电信用户的行为特性,总结了一套基于全面、科学、公正、可操作的评价指标体系^[4],如表 1 所示。

表 1 电信用户信用风险评估指标体系

总目标	子目标	编号	指标
电信用户信用度属性评估指标体系	用户基本属性	x1	客户所属地(本、外地)
		x2	客户类别
		x3	职业类型
		x4	客户资料真实性
		x5	黑名单
	用户业务属性	x6	入网时长
		x7	缴费方式
		x8	客户级别
	用户欠费属性	x9	欠费金额
		x10	欠费时长
		x11	用户状态
		x12	月平均停机次数
		x13	平均停机时长
	用户贡献度属性	x14	用户月平均消费额
		x15	用户累计销账总额
		x16	用户每次缴费金额

2 PSO-BP 神经网络模型的建立

2.1 BP 神经网络模型

BP(Back Propagation)^[5]神经网络算法由输入层、隐含层和输出层组成,目的是解决多层网络中隐含单元连接权学习的问题。BP 神经网络的输入层从外界接受信息,输入信号从输入节点依次传过各隐含层,然后传到输出节点,输出层把网络处理后的信息传向外界。但是传统的 BP 神经网络还存在一些不足人意的地方,主要表现在以下三个方面:

(1)较慢的收敛速度。BP 算法的收敛过程是反向传播误差信号,从而不断修改各层神经元的权值、阈值直至达到设定的目标,所以说,误差减小会消耗大量的时间。

(2)经常搜寻到局部极小值,BP 算法是采用梯度下降法调整网络权值、阈值,所以常常搜寻到局部极小点就结束算法。

(3)BP 网络中隐含层及隐含层节点数的设定都要靠不断的试算。

基于上述缺点,文中模型采用 PSO 算法对神经网络的权值进行优化。

2.2 粒子群优化(PSO)算法

1995 年,PSO 算法^[6]的提出是基于鸟群觅食行为的特性,这是一种思路比较新颖的全局优化进化的人工智能算法。有关 PSO 算法的基本原理及算法实现过程详见文献[7~9]。

2.3 用 PSO 优化后的 BP 神经网络

将 PSO 算法应用于优化 BP 神经网络^[10]的算法大体流程是:传统 BP 算法中的梯度下降法用 PSO 算法进行替代,以此来迭代进化 BP 神经网络模型的各个权值参数组合,直到目标的适应度趋于平缓为止^[11],接着,对此时得出的网络参数用 BP 算法再次精确优化,训练结束后,得到的就是最优网络参数,此时认为该 BP 神经网络达到最优参数组合^[12]。

PSO-BP 算法步骤:

(1)确立基础的 BP 神经网络结构,设定输入层、隐含层、输出层的神经元个数。

(2)对粒子群算法进行初始化,包括粒子的初始位置、初始速度、种群大小、惯性因子 w 的最大值和最小值、最大迭代次数和学习因子 c_1 、 c_2 ,并设定每一个粒子的个体极值和全局最优值等。

(3)设定适应度函数。粒子搜索性能的评价指标设定为神经网络的最小均方差 MSE,即适应度,用于评估种群的搜索结果。

(4)根据其他输入、输出样本值,求出每个粒子适应度。

①先对输入的一个粒子,其中的一个样本通过 BP 神经网络计算出一个网络的输出值,然后根据式(1)得出其误差;同理接着得出所有样本的误差;该粒子的适应度根据式(2)就是所有样本的均方差。

$$f_i = \sum_{j=1}^J (y_{ij} - Y_{ij})^2 \tag{1}$$

$$f_m = \frac{1}{N} \sum_{i=1}^N f_i \tag{2}$$

②返回到步骤①,用其他粒子作为输入,计算出所有粒子的适应度。

(5)根据各粒子的适应度更新个体极值和全局最优值。

(6)更新粒子的速度和位置。

(7)计算出该算法的误差。

$$E = \frac{\sum_{i=1}^k f(P_g^{(i)})}{k} \tag{3}$$

其中, k 为当前迭代次数, $f(P_g^{(i)})$ 为第 i 次迭代的全局最优值的适应度。

(8)迭代过程停止的条件。

算法结束的标准有两个,一个是误差达到设定的精度,或者迭代的次数达到了设定的最大次数,那此时认为算法收敛,最后一次的全局最优值 P_g 中每一维的权值和阈值就是最优解;若迭代次数未达到预设的最大值,算法返回步骤(4)继续进行迭代。

3 电信用户信用度数据实验

3.1 数据归一化处理

文中所使用的电信用户实验数据来自某市电信固网用户 2007 年 1 月~7 月中的随机 2000 个客户数据进行信用度评分建模实验。其中取 95% 即 1900 客户数据作为训练数据,余下的 5% 即 100 个客户数据作为测试数据。可用的样本数据属性共有 16 个,首先对训练数据和测试数据进行归一化处理,对于模型所采用的 16 个指标,先将其量化后数字化表示,然后将其分成离散型变量和连续型变量两组。

对于离散型变量,例如客户所属地、客户资料真实性、缴费方式等采用最小-最大归一化方法进行处理,即

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

其中 $y \in [0, 1]$ 表示归一化后的变量值, x_{\min} 和 x_{\max} 分别表示变量的最大值和最小值。

对于连续型变量,例如入网时长、客户级别欠费金额、欠费时长等通过对变量值的分布状况分析,用如下归一化方法处理:

$$y = \varphi\left(\frac{x - \mu}{\sigma}\right) \quad (5)$$

其中 $\varphi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 表示累计标准正态分布概率。

3.2 模型结构设计

该电信客户信用度预测的 PSO-BP 模型采用三层网络结构,分别为属性输入层、隐含层、评估值输出层,相邻层之间通过连接函数进行完全连接。输入层节点个数为 16 个,采用表 1 中归一化后的信用度指标值,输出层为该电信用户的信用分值。输出层的神经元个数设为 1,隐含层节点个数根据大拇指规则,然后根据试凑法来调节,最后确定为 $H=6$ 。

3.3 模型网络参数设置

粒子的维数:本模型的粒子维数大小为 $D=99$;种群大小 $N=40$,预设精度设为 10^{-4} ;学习因子习惯性取 $c_1=2, c_2=2$,最大迭代次数我们预设 100,最大速度 $V_{\max}=0.5$,惯性因子的极值为: $W_{\max}=0.90, W_{\min}=0.30$,粒子位置的初始化范围设为 $(a, b), a=-1, b=1$;粒子速度初始化范围设为 $(m, n), m=-1, n=1$ [13]。

BP 神经网络输入层与隐含层之间的传递函数采用“tansig”;而隐含层与输出层之间的传递函数采用“logsig”。根据电信用户信用度评估标准,欠费用户与正常用户的区分阈值为 20,则该模型输出结果为一个 $[0, 40]$ 区间内的数值,可以以此来区分信用度。

利用归一化后的样本对网络进行训练,经过多次

迭代,网络误差达到设定的最小值。

对于已经训练好的网络,将余下的 100 个样本作为测试数据,得出所对应的输出作为信用度分值分布图如图 1 所示:

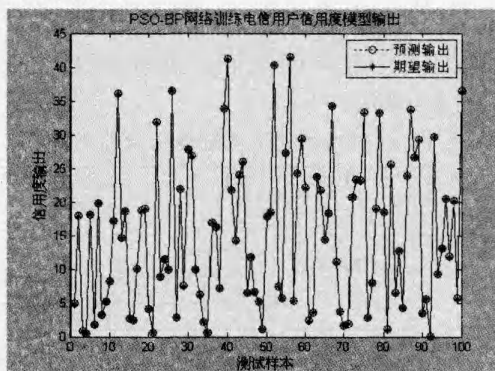


图 1 PSO-BP 网络训练样本信用度值输出

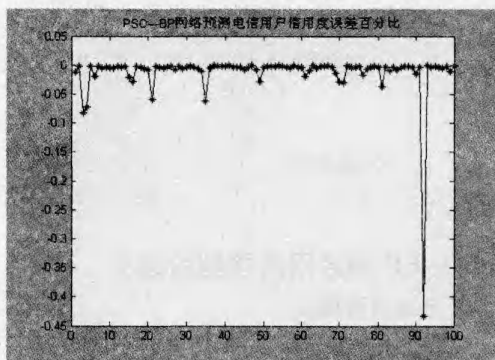


图 2 PSO-BP 网络预测信用度误差百分比

从图 1、图 2 看出,该信用度模型与原始数据有较高的重合度,误差较小。100 个用户的评估精度为 82.74%,其预测效果是比较理想的,说明该算法建立的模型有很高的准确性。

4 结束语

文中通过运用粒子群优化算法改善 BP 神经网络,并采用实际数据集通过网络训练得到电信客户的信用度评估模型,同时在 Matlab 工具箱中实现了该模型。实验结果显示,该模型简洁、准确率高。该算法建立的电信客户信用评估模型是电信用户信用控制和防欺诈的基础,这样有助于对客户群行为特征的区分,更好地预测客户未来的信用风险及企业的损益程度从而建立预警机制,并在一定程度上提高客户的满意度。

参考文献:

- [1] 杨奕. 电信企业如何建立信用度管理机制[J]. 电信科学, 2002(2): 51-52.
- [2] 姚琦云. 多途径提高电信欠费催缴率[J]. 邮电企业管理, 2001(22): 47-48.

(下转第 152 页)

组合这些子模块内容偏好业务进行推广销售。基于改进算法,可以求出更多内容偏好的关联组合,对于 $XY \Rightarrow Z$ 形式的强关联规则,其结果见表 2。

表 2 形如 $XY \Rightarrow Z$ 的内容偏好强关联结果

偏好 1	偏好 2	偏好 3	置信度
新闻	求职	投资	100.00%
投资	科技	新闻	100.00%
新闻	投资	科技	87.50%
新闻	科技	投资	87.50%
求职	投资	新闻	80.00%
新闻	投资	求职	50.00%

注:表中关联规则为:偏好 1 \cup 偏好 2 \Rightarrow 偏好 3

表 2 中组合 {新闻、求职、投资} 和 {投资、科技、新闻} 也都跨越大模块内容偏好分类。由此可见,不仅可以组合各相关内容偏好进行销售,而且可以根据客户已经具有的内容偏好类别,推荐该客户另一由这两个偏好推导出的偏好业务。

通过对大模块内容偏好强关联分析可知,客户互联网访问内容偏好模块,即音乐、资讯、商务、娱乐之间存在着很强的关联性,可以在大方向上进行业务组合的销售;通过进一步对子模块内容偏好强关联分析,得到更细化关联分析结果,这不仅可以进行业务的组合销售,还可以挖掘客户的潜在内容偏好,进行业务的推荐销售。

4 结束语

综上所述,文中是基于关联规则挖掘移动客户互联网访问内容数据间的关联,在已细分的各个偏好内容模块中寻找内在的关联。从而针对特定的用户提供相应的精确业务,赢得营销机会。该方法相对于以往

凭借客户基础资料和客户通信行为的分析有较大进步,更能有效命中目标客户群的偏好访问内容,提高营销效率,增加客户满意度。

参考文献:

- [1] 中国移动通信集团福建有限公司. 基于客户偏好研究,开展数据及信息业务内容深度营运[EB/OL]. [2010-09-20]. <http://www.doc88.com/p-703879803187.html>.
- [2] Witten L H, Frank E, Hall M A. Data Mining: Practical Machine Learning Tools and Techniques[M]. 3rd ed. [s. l.]: Morgan Kaufmann, 2011.
- [3] 安淑芝. 数据仓库与数据挖掘[M]. 北京:清华大学出版社, 2005.
- [4] 储文静, 奉国和. 基于 Weka 读者借阅行为分析[J]. 情报科学, 2010, 28(3): 424-429.
- [5] 付 锋. 移动互联网访问行为分析研究[EB/OL]. [2011-09-06]. <http://labs.chinamobile.com>.
- [6] 王 静, 张春海. 基于布尔关联规则挖掘的加权阈值分析[J]. 计算机技术与发展, 2009, 19(12): 13-16.
- [7] 丁艳辉, 王洪国, 高 明, 等. 一种基于矩阵的关联规则挖掘新算法[J]. 计算机科学, 2006, 33(4): 188-189.
- [8] Lu Juemin, Ma Guodong, Zheng Yu. The Association Analysis for Library Circulation Data Based Mining Technique[J]. Journal of Modern Information, 2009, 29(9): 108-110.
- [9] 邵峰晶, 于忠清. 数据挖掘原理与算法[M]. 第 2 版. 北京: 科学出版社, 2009.
- [10] 王爱平, 王占凤, 陶嗣干, 等. 数据挖掘中常用关联规则算法[J]. 计算机技术与发展, 2010, 20(4): 105-108.
- [11] 中国互联网监测研究机构 & 数据平台. DCCI 2011 中国移动互联网用户调查报告[EB/OL]. [2011-06-10]. <http://www.dcci.com.cn>.
- [12] Han Jiawei, Kamber M. Data Mining Concepts and Techniques[M]. San Francisco: Morgan Kaufmann Publishers, 2001.

(上接第 148 页)

- [3] 陈大峰, 万洛楷. 移动 BOSS 系统中客户信用度综合评定的研究[J]. 南京审计学院学报, 2006(4): 99-103.
- [4] 王丽平, 李多全. 基于 AHP 方法计算电信用户信用度[J]. 计算机工程与应用, 2008(32): 232-239.
- [5] 李祚泳, 彭荔红. BP 网络学习能力与泛化能力满足的不确定关系式[J]. 中国科学(E 辑), 2003, 33(10): 887-895.
- [6] 彭 宇, 彭善元, 刘兆庆. 微粒群算法参数效能的统计分析[J]. 电子学报, 2004(2): 209-213.
- [7] Kennedy J, Eberhart R. Particle Swarm Optimization[C]//IEEE International Conference on Neural Networks (ICNN 95). Perth, Australia: [s. n.], 1995.
- [8] Parsopoulos K E, Vrahatis M N. Particle Swarm Optimizer in Noisy and Continuously Changing Environments[C]//Proceeding of the IASTED International Conference on Artificial Intelligence and Soft Computing. ICanun Mexico: IASTED/

ACTA Press, 2001: 289-294.

- [9] Parsopoulos K E, Vrahatis M N. Particle Swarm Optimization Method for Constrained Optimization Problems[C]//Proceedings of the Euro-international Symposium on Computational Intelligence. Konica, Slovakia: IOS Press, 2002.
- [10] 吴小红, 金炳尧. 前馈神经网络的一种优化 BP 算法[J]. 计算机科学, 2004, 31(10A): 240-241.
- [11] 潘 昊, 侯清兰. 基于粒子群优化算法的 BP 网络学习研究[J]. 计算机工程与应用, 2006(16): 41-43.
- [12] 曾万里, 危韧勇, 陈红玲. 基于改进 PSO 算法的 BP 神经网络的应用研究[J]. 计算机技术与发展, 2008, 18(4): 49-51.
- [13] 唐 俊. PSO 算法原理及应用[J]. 计算机技术与发展, 2010, 20(2): 213-216.