

基于自适应遗传神经网络的银行客户分类研究

汤亚玲, 黄 华, 程泽凯

(安徽工业大学 计算机学院, 安徽 马鞍山 243002)

摘 要: 银行产品的营销行为都是针对广大客户的。若能提前分辨出哪些是优质客户, 再为其定制合理的营销策略, 那银行就能获得更大的竞争力。文中将遗传算法与 BP 神经网络结合用于对银行客户分类进而预测客户是否会购买银行产品。该方法有效地克服了 BP 神经网络容易陷入局部极小值和收敛速度慢的问题, 并且针对其中遗传算法的计算时间和精度问题提出了一种新的自适应遗传算法。实验结果表明, 基于这种自适应的遗传神经网络的方法用更短的计算时间达到了更高的预测精度, 可以准确地为银行客户分类。

关键词: 遗传算法; 自适应; 神经网络; 客户分类

中图分类号: TP183

文献标识码: A

文章编号: 1673-629X(2014)07-0192-04

doi: 10.3969/j.issn.1673-629X.2014.07.048

Research on Classification of Bank Customers Based on Adaptive GA-BP Algorithm

TANG Ya-ling, HUANG Hua, CHENG Ze-kai

(School of Computer Science, Anhui University of Technology, Ma'anshan 243002, China)

Abstract: The products in bank marketing are faced to the majority of customers. If tell in which are high-quality customers in advance and then develop reasonable marketing strategy for them, bank will be able to achieve greater competitiveness. It combines genetic algorithm with BP network for bank customers classification to predict whether the customers will buy the bank marketing products. It can effectively overcome the shortcomings of BP network, such as trapping to the local minimum and slowness in training speed. Aiming at the computation time and accuracy of genetic algorithm, a new adaptive GA-BP algorithm is proposed. Experimental results show that the adaptive GA-BP algorithm can reach a higher prediction accuracy with a shorter calculation time and it can classify bank customers accurately.

Key words: genetic algorithm; adaptive; neural network; customer classification

0 引言

随着外资银行在我国的大量植入, 国内银行面临着巨大的竞争挑战, 随即客户便成为了各银行竞争的焦点。因为高价值的客户越多, 银行的竞争力就越大^[1], 所以为了合理制定营销策略, 减少营销成本, 银行必须首先对客户进行分类, 识别出哪些是优质客户, 再为其分配特定的服务资源以求获得更大的收益。

目前对银行客户分类有定性和定量两种分类方法^[2], 早前用的是基于定性的分类法, 也就是根据决策者的经验对客户进行类别划分, 划分结果粗糙, 准确率低。而随着时代信息化的不断发展, 面对银行的海量

客户数据, 数据挖掘作为一个能从数据中发现和提取知识的新技术被应用进来, 也就是之后所用的基于定量的分类法^[3]。其中 BP 神经网络是一种非常受青睐的分类方法^[4], 因为它具有联想记忆、自适应、自组织、自学习能力, 并且它可以无限逼近非线性系统。但是由于它是建立在梯度下降算法的基础上的, 所以它具有学习时间长、收敛速度慢、容易陷入局部极小等缺点。而遗传算法是一种基于生物自然选择和遗传机理的并行随机搜索最优化方法, 将两者相结合, 用 GA 来优化 BP 神经网络的初始权值和阈值^[5], 可以有效地使神经网络的学习跳出局部最优^[6]。而用 GA 来改

收稿日期: 2013-09-26

修回日期: 2013-12-30

网络出版时间: 2014-04-24

基金项目: 国家自然科学基金资助项目(60473142); 安徽省高校重点项目(KJ2010A051, KJ2011A039); 安徽省高校省优秀青年人才基金项目(2009SQRZ076)

作者简介: 汤亚玲(1974-), 男, 安徽庐江人, 副教授, 硕士生导师, 研究方向为智能化信息处理、数据挖掘及网络数据库系统; 黄 华(1989-), 女, 江苏南通人, 硕士研究生, 研究方向为人工智能、数据挖掘。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140424.0824.072.html>

善 BP 算法的同时又不得不考虑计算时间的问题^[7], 所以文中提出采用自适应遗传算法, 使得交叉算子和变异算子随着适应度值的改变而自适应变化^[8]。结果表明, 自适应遗传神经网络模型用在银行客户分类上有比其他模型更高的分类精度和更短的计算时间。

1 自适应遗传神经网络模型及算法

1.1 编码

采用实数编码^[9]。因网络权值范围不明确, 利用实数编码可以避免难以进行编码的困难; 取消了编码和译码过程, 提高了学习算法的速度; 便于较大空间的遗传搜索, 提高了遗传算法的精度要求。

以三层神经网络为例, 染色体形式表示为 (W_1, b_1, W_2, b_2) , 其中 W_1 表示输入层与隐藏层之间的权值, b_1 表示隐藏层的阈值, W_2 表示隐藏层与输出层之间的权值, b_2 表示输出层的阈值。染色体长度为:

$$l = W_1 + b_1 + W_2 + b_2 \quad (1)$$

染色体就由 l 个 0~1 之间的随机数组成。

1.2 适应度函数

根据个体得到 BP 神经网络的初始权值和阈值, 用训练数据训练网络后预测系统输出, 预测输出和期望输出的误差绝对值和的倒数作为染色体的评价函数。即

$$f = 1 / \sum_{i=1}^n \sum_{j=1}^k |y_{ij} - o_{ij}| \quad (2)$$

其中, i 为训练样本数目; j 为神经网络输出层的神经元个数; y_{ij} 为第 i 个样本输入后输出层第 j 个神经元的实际输出值; o_{ij} 为其期望输出值。

1.3 选择算子

利用轮盘赌选择法, 保证各个个体被选中的概率与其适应度大小成正比。并且采用最优个体保存策略, 使得当前群体中适应度最高的个体不参与交叉和变异运算, 并且用它来替换本代群体中适应度最低的个体, 此方法可以保证迄今为止得到的最优个体不被破坏, 保证了算法的收敛性^[10]。

1.4 交叉算子和变异算子

相对于早前 Srinivas 等提出的自适应遗传算子^[11], 文中提出一种新的自适应交叉算子和变异算子, 使得当个体适应度值小于平均适应度值的时候, 以较大的概率进行交叉和变异, 使该解被淘汰; 当个体适应度值大于平均适应度时, 以递减的概率进行交叉和变异, 使该解得以保留进入下一代^[12]。公式如下:

$$p_c = \begin{cases} p_{cl} * \frac{1}{\exp(\frac{f - f_{avg}}{f_{max} - f_{avg}})} & f \geq f_{avg} \\ p_{cl} & f < f_{avg} \end{cases} \quad (3)$$

$$p_m = \begin{cases} p_{ml} * \frac{1}{\exp(\frac{f - f_{avg}}{f_{max} - f_{avg}})} & f \geq f_{avg} \\ p_{ml} & f < f_{avg} \end{cases} \quad (4)$$

其中, f 为两个要交叉的个体中较大的个体的适应度值; f_{avg} 为群体的平均适应度值; f_{max} 为群体中最大的适应度值; f 为要变异的个体的适应度值; p_{cl} 和 p_{ml} 分别为最大的交叉概率和变异概率。

这种交叉算子和变异算子的自适应调整曲线如图 1 所示。

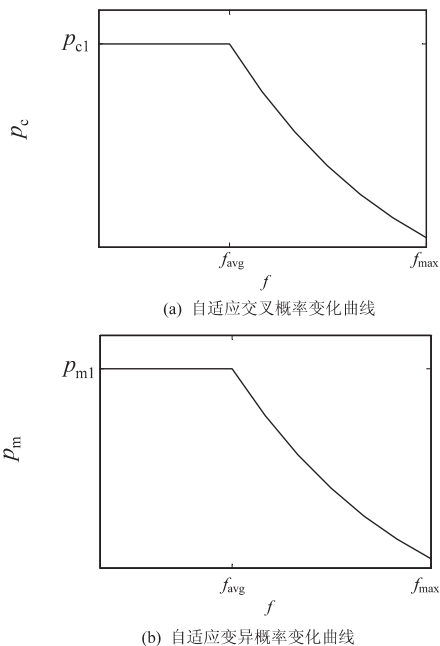


图 1 自适应交叉和变异概率调整曲线

1.5 算法流程

(1) 种群初始化。随机产生 N 组初始网络权值, 每组作为一个染色体。

(2) 选择操作。分别计算每个染色体的适应度值和相应比例, 采用轮盘赌法和最优个体保存策略选择 N 个个体进入下一代。

(3) 交叉操作。

(4) 变异操作。若到达迭代步数, 则转向(5), 否则转向(2)。

(5) 把最优染色体作为 BP 神经网络的初始权值和阈值代入, 用训练集继续训练, 直到达到误差精度或者最大迭代数, 训练结束。

图 2 为算法流程图。

2 仿真实验与结果分析

2.1 数据准备和预处理

实验数据来源于 UCI 机器学习数据库中的一组银行客户数据, 这组数据是关于银行的营销产品最终是否能成功被客户购买。一共从中取了 2 000 组数据

样本,每个样本由 16 个属性和 1 个类别构成,前面 16 个属性可以分为三类:

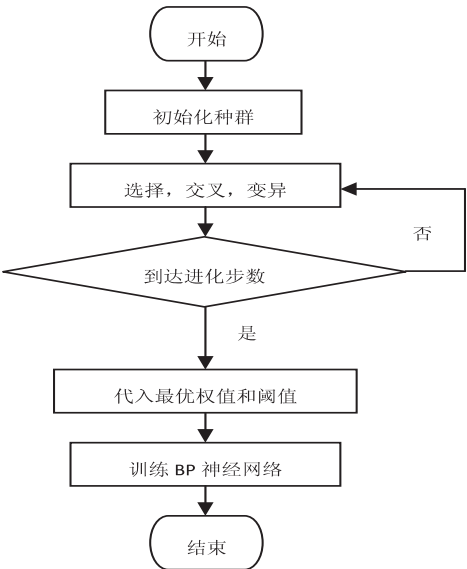


图 2 自适应遗传神经网络算法流程图

人口统计变量:描述客户的一些基本信息,包括 age, job, marital, education, default, balance, housing, loan。

银行与客户对于目前的营销活动做的最后一次交流的一些变量,包括 contact, day, month, duration。

银行与客户就之前的活动交流的一些变量,包括 campaign, pdays, previous, poutcome。

后面这个类别说明这个客户是否会购买银行这次的营销产品。

预处理主要进行以下两步^[13]:

(1)非数值属性数值化。把非数值属性中每个属性类别都用一个数值表示。

(2)连续属性离散化。把连续属性中的每个数据范围都用一个数值表示^[14]。

2.2 模型的建立和参数的选取

首先设定网络的拓扑结构,网络的输入层节点数为客户的属性个数 16,输出层节点数为 1,表示客户的两个类别,隐含层节点数根据经验公式(5)计算,最终确定为 10。构成 16-10-1 的网络。

$$n = \sqrt{a + b} + c \tag{5}$$

其中, a 和 b 分别为输入层和输出层的节点个数; c 为 1~10 之间的某个实数。

GA 的参数如下:染色体由网络的所有权值和阈值构成,根据公式(1)得出长度为 $16 \times 10 + 10 + 10 \times 1 + 1 = 181$,群体规模取 40,最大进化代数取 200,交叉概率中的 p_{cl} 取 0.9,变异概率中的 p_{ml} 取 0.1。

2.3 算法性能比较分析

文中将数据集随机分为两组,一组为训练集 1 900 例,用于网络的建模,另一组为测试集 100 例,用于检

测网络的泛化性能。

2.3.1 分类准确率比较

图 3 为对 100 组测试集进行分类的误差结果图。其中 GABP1 表示一般的遗传神经网络,GABP 表示自适应遗传神经网络,可以看出后者误差更向 0 值靠拢,表明它用于银行客户分类的精确度更高。若把误差处于 0.01 范围内的表示为分类准确,那么两者用于分类的准确率如表 1 所示。

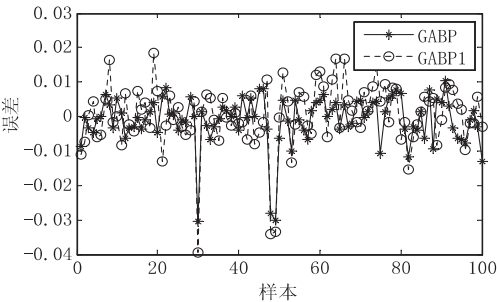


图 3 测试集分类误差结果图

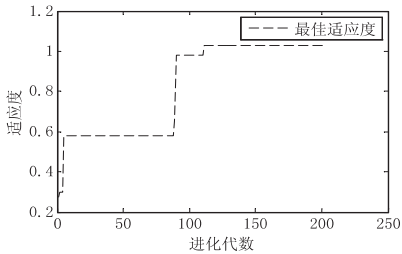
表 1 改进前后神经网络分类准确率

神经网络类型	分类准确率/%
一般遗传神经网络	74
自适应遗传神经网络	91

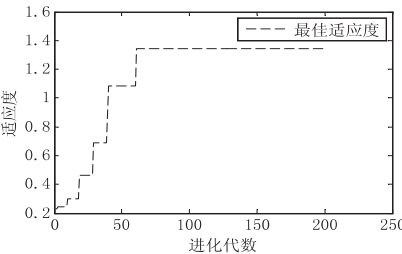
数据显示后者比前者分类准确率高了 17%。

2.3.2 运行时间比较

图 4 为改进前后 GABP 的适应度曲线变化图,图 a 为一般遗传神经网络的遗传迭代图,图 b 为自适应遗传神经网络的遗传迭代图。设定终止代数均为 200,图 a 显示大约在 120 代时趋于收敛,图 b 显示大约在



(a) 改进前适应度曲线变化图



(b) 改进后适应度曲线变化图

图 4 改进前后 GABP 的适应度曲线变化图 70 代时趋于收敛,而且收敛值高于图 a。这说明如果设定一个固定的收敛值,那么后者的迭代数一定远远

小于前者,所以后者的运行效率要比前者高很多。

3 结束语

文中采用自适应遗传神经网络对银行客户购买营销产品的行为进行分类,以此来预测客户是否会购买本次营销产品,有利于银行能够充分掌握客户的购买行为,这样,银行就可以把营销策略重点放在优质客户身上以获取更大的利润空间。

通过实例仿真,验证了自适应遗传神经网络比简单遗传神经网络在银行客户分类问题上有更高的分类准确率和更高的分类效率。

参考文献:

[1] 张国政. 客户关系管理中基于数据挖掘的客户细分研究[J]. 商业研究,2006(13):153-155.

[2] 樊宁. K 均值聚类算法在银行客户细分中的研究[J]. 计算机仿真,2011,28(3):369-372.

[3] Ngai E W T, Li Xiu, Chau D C K. Application of data mining techniques in customer relationship management: a literature review and classification[J]. Expert Systems with Applications, 2009, 36: 2592-2602.

[4] Xu E, Shao Liangshan, Gao Xuedong, et al. An algorithm for predicting customer churn via BP neural network based on rough set[C]//Proceedings of the 2006 IEEE Asia-Pacific

conference on services computing. Guangzhou: IEEE, 2006: 47-50.

[5] Zhou Li, Wu Qizong. Application of GA-BP algorithm in customer classification[C]//Proceedings of 2011 IEEE 2nd international conference on software engineering and service science. Beijing: IEEE, 2011: 255-257.

[6] 张庆红,程国建. 基于遗传算法的神经网络性能优化[J]. 计算机技术与发展,2007,17(12):125-127.

[7] 沈亮,吴克坚. 混合遗传算法对 BP 神经网络算法的改进[J]. 机械设计与研究,2004,20(2):10-12.

[8] 曹道友,程家兴. 基于改进的选择算子和交叉算子的遗传算法[J]. 计算机技术与发展,2010,20(2):44-47.

[9] 范睿,李国斌,景韶光. 基于实数编码遗传算法的混合神经网络算法[J]. 计算机仿真,2006,23(1):161-163.

[10] 赵振勇,王力,王保华,等. 遗传算法改进策略的研究[J]. 计算机应用,2006,26(12Z):189-191.

[11] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Transactions on Systems, Man and Cybernetics, 1994, 24(4): 656-667.

[12] 李延梅. 一种改进的遗传算法及应用[D]. 广州:华南理工大学,2012.

[13] 李延梅. 基于 ASR 的呼叫中心实现及客户分类研究[D]. 大连:大连理工大学,2008.

[14] 王萍. 运用数据挖掘技术预测客户购买倾向-方法与实证研究[J]. 情报科学,2005,23(5):738-741.

(上接第 128 页)

4 结束语

综合层次路由与语义路由机制设计一种层级型无线传感网基于语义路由的数据转发方法,通过融入语义,提出一种新的路由数据转发算法,提高路由效率,达到延长网络生命周期的目的。但需要解决的问题仍很多,例如路由能量平衡、路由器的轮替等。在后期工作中,继续加大研究,更希望能够对能耗和路由可靠性进行验证。

参考文献:

[1] 戴世瑾,张翼德,李乐民. 无线传感器网络的路由协议研究与分析[J]. 计算机应用研究,2006,23(12):294-297.

[2] Li J Z, Li J B, Shi S F. Concepts, issues and advance of sensor networks and data management of sensor networks[J]. Journal of Software, 2003, 14(10): 1717-1727.

[3] Kulik J, Heinzelman W R, Balakrishnan H. Negotiation based protocols for disseminating information in wireless sensor networks[J]. Wireless Networks, 2002, 8(2-3): 169-185.

[4] 张婷婷,周鸣争. 一种基于语义的无线传感器网络路由算法[J]. 计算机技术与发展,2007,17(12):103-106.

[5] 李章华,刘高嵩,张怡. 一种基于语义的结构化 P2P 网络

模型[J]. 计算机时代,2006(11):17-19.

[6] 翟晓波,杨放春. 基于语义的网络自适应上下文资源感知框架设计[EB/OL]. 2005. [http://www. paper. edu. cn/releasepaper/content/200503-102](http://www.paper.edu.cn/releasepaper/content/200503-102).

[7] 向永生,张颖,陈曦. 自组织聚类的 P2P 语义路由算法[J]. 计算机工程,2010,36(9):123-125.

[8] 邵鹰,刘业. 结构化 P2P 路由协议 Chord 的分析和实现[J]. 计算机工程,2007,33(19):122-124.

[9] 郭龙江,李建中,李贵林. 无线传感器网络环境下时-空查询处理方法[J]. 软件学报,2006,17(4):794-805.

[10] 叶春,葛燧和,熊齐邦. 基于语义路由的 P2P 信息检索[J]. 计算机仿真,2004,21(10):143-145.

[11] Madden S, Franklin M J, Hellerstein J M, et al. TAG: a tiny aggregation service for ad-hoc sensor networks[C]//Proc of the 5th symposium on operating systems design and implementation. New York, NY, USA: ACM, 2002: 131-146.

[12] Beaver J, Sharaf M A, Labrinidis R, et al. Power-aware in-network query processing for sensor data[C]//Proc of the 2nd Hellenic data management symposium. [s. l.]: [s. n.], 2003.

[13] Xu Y, Lee W C. Window query processing in highly dynamic sensor networks: issues and solutions[C]//Proc of workshop on GeoSensor networks. [s. l.]: [s. n.], 2003.

作者：[汤亚玲](#)，[黄华](#)，[程泽凯](#)，[TANG Ya-ling](#)，[HUANG Hua](#)，[CHENG Ze-kai](#)
作者单位：[安徽工业大学 计算机学院, 安徽 马鞍山, 243002](#)
刊名：[计算机技术与发展](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2014(7)

参考文献(14条)

1. [张国政](#) [客户关系管理中基于数据挖掘的客户细分研究](#) 2006(13)
2. [樊宁](#) [K均值聚类算法在银行客户细分中的研究](#) 2011(03)
3. [Ngai E W T;Li Xiu;Chau D C K](#) [Application of data mining techniques in customer relationship management:a literature review and classification](#) 2009
4. [Xu E;Shao Liangshan;Gao Xuedong](#) [An algorithm for predicting customer churn via BP neural network based on rough set](#) 2006
5. [Zhou Li;Wu Qizong](#) [Application of GA-BP algorithm in cus-tomer classification](#) 2011
6. [张庆红;程国建](#) [基于遗传算法的神经网络性能优化](#) 2007(12)
7. [沈亮;吴克坚](#) [混合遗传算法对BP神经网络算法的改进](#) 2004(02)
8. [曹道友;程家兴](#) [基于改进的选择算子和交叉算子的遗传算法](#) 2010(02)
9. [范睿;李国斌;景韶光](#) [基于实数编码遗传算法的混合神经网络算法](#) 2006(01)
10. [赵振勇;王力;王保华](#) [遗传算法改进策略的研究](#) 2006(12z)
11. [Srinivas M;Patnaik L M](#) [Adaptive probabilities of crossover and mutation in genetic algorithms](#) 1994(04)
12. [李延梅](#) [一种改进的遗传算法及应用](#) 2012
13. [李延梅](#) [基于ASR的呼叫中心实现及客户分类研究](#) 2008
14. [王萍](#) [运用数据挖掘技术预测客户购买倾向-方法与实证研究](#) 2005(05)

引用本文格式：[汤亚玲](#)，[黄华](#)，[程泽凯](#)，[TANG Ya-ling](#)，[HUANG Hua](#)，[CHENG Ze-kai](#) [基于自适应遗传神经网络的银行客户分类研究](#)[期刊论文]-[计算机技术与发展](#) 2014(7)