

基于优化 GA-BP 神经网络的金融产品营销预测

金鑫, 潘宜安, 吴靖

(中央财经大学 信息学院, 北京 100081)

摘要: 传统 BP 神经网络存在着网络结构参数确定过于依赖经验、易于陷入局部解等缺陷, 为了改进 BP 神经网络模型的应用缺陷, 提出优化 GA-BP 算法, 通过 GA 算法优化 BP 神经网络拓扑结构和网络参数初始值的选取过程, 并且为了验证模型的可行性, 以某银行短期理财产品营销的客户历史数据作为实证研究对象, 并通过与 BP 神经网络模型的对比实验, 验证该模型可以更精确地预测银行理财产品的客户营销结果。实验结果表明将该模型用于对金融产品营销数据的仿真计算, 可以更精确地预测未来营销结果。

关键词: GA-BP; 银行营销; 优化

中图分类号: TP312

文献标识码: A

文章编号: 1000-436X(2014)Z2-0020-06

Financial production marketing prediction based on optimization GA-BP neural network

JIN Xin, PAN Yi-an, WU Jing

(School of Information, Central University of Finance and Economics, Beijing 100081, China)

Abstract: The traditional BP neural network has some application problems. For example, the network structure parameter is too dependent on experience and easy to fall into local solution. In order to improve the application defects of BP neural network model, the optimization GA-BP algorithm to optimize BP neural network topology and the selection process of network initial parameter value is proposed. In order to verify the feasibility of the model, marketing customer historical data of a bank short-term financial products as the research object is used to validate the model which could more accurately predict the customer compared with BP neural network model. The test results show that the model could be applied to analysis financial product marketing data and more accurately predict the future marketing results.

Key words: GA-BP; bank promotion; optimization

1 引言

随着市场竞争愈演愈烈, 商业银行除了经营传统的存贷款业务外, 现在更创新了多种金融产品来提高银行的盈利水平。精确定位目标客户群体, 合理预测产品销售结果, 有利于更合理地分析产品的盈利性。目前用于商业银行客户市场细分、销售预测的方法主要包括神经网络、决策树以及遗传规划等, 其中神经网络是商业预测方法的热点。对数据的逻辑关系要求不严格, 不必详细表述自变量与因变量之间的函数关系, 使 BP 神经网络成为应用最

广泛的神经网络模型^[1]。BP 神经网络存在着网络结构参数确定过于依赖经验、易于陷入局部解等缺陷, 而遗传算法 GA 可以在全局范围搜索最优点, 弥补了 BP 算法的不足。本文从这点出发, 探索将 GA 算法与 BP 神经网络模型相结合, 运用于金融产品的销售预测。

2 相关工作

人工神经网络是模拟人的思维的一种算法, 最早的研究工作可以追溯至 20 世纪 40 年代, 心理学家 W Mcculloch 和数理逻辑学家 W Pitts 在分析、总

收稿日期: 2014-10-10

基金项目: 国家自然科学基金资助项目 (61273293)

Foundation Item: National Natural Science Foundation of China (61273293)

结神经元基本特性的基础上提出神经元的形式化模型。后面的学者在此基础上，不断对神经元联结的更新机制进行探索和发展。如 D Hebb 提出 Hebb 学习规则^[2]、B Widrow 和 M Hoff 等提出 Adaline 理论等。然而有效更新隐含层连接权值的算法始终匮乏。上世纪 80 年代，以 McClelland J L, Rumelhart D E 为首的美国 PDP 小组在神经网络中成功应用误差反向传播算法^[3]，最终提供了有效调整非线性连续函数多层前馈神经网络的权重的算法。

BP 神经网络具有给定足够多的隐藏单元和训练样本，可以逼近任何函数的优点，这使其得到越来越广泛的应用。但标准的 BP 算法在应用中暴露出不少内在的缺陷：如 BP 算法的神经网络模型参考自适应控制实时性差、精度不高、收敛慢，对此，李红、杨剑锋提出了基于改进的双向权值调整 BP 算法的神经网络模型参考自适应控制^[5]。对于 BP 神经网络收敛速度慢的缺陷，早在 1996 年，我国学者席道瑛、张涛就证明利用模拟退火算法与 BP 神经网络相结合计算可以具有良好的收敛速度^[6]。

对于网络参数选取易于陷入局部极值的缺陷，最常用的是应用粒子群算法、遗传算法等进行优化。张卫东、韩云昊进行的银行信用风险评估^[7]，苏晨、李成义进行的服装销售预测^[8]等，是将遗传算法与 BP 神经网络相结合在商用领域预测的研究应用。但他们对遗传算法的结合应用只限于对网络初始权值、阈值参数的优化，侧重于利用 GA 的全局搜索功能突破权值、阈值局部极值的缺陷，并没有涉及网络结构参数确定经验化的问题。

本文以某银行短期理财产品营销的客户历史数据作为实证研究对象，使用 GA 算法优化 BP 神经网络参数，突破 BP 神经网络局部最小值以及网络拓扑结构选择的主观性的限制，建立实现全局最优的拟合模型。该模型可以预测新客户的营销状况，为银行的定向营销策略提供指导。

3 GA 算法与 BP 神经网络

3.1 GA 算法

遗传算法模拟自然选择，将自然界的遗传进化理念运用到解决问题中。所谓自然选择通过优胜劣汰的方式使逐代进行遗传进化，从而最终存活下来的物种具有最好的适应性。遗传算法实质上把问题可行解经过遗传编码操作变成向量（即染色体），而向量中的每一个元素就是染色体中的遗传基因。

由此产生初始种群后，按照自然选择的原理，逐代演化产生出越来越好的近似解。在每一代，根据问题域中个体的适应度大小选择个体，并借助遗传算子进行组合交叉和变异，产生出代表新的解集的种群。末代种群中的最优个体即为问题的近似最优解。

在具体实现上，先进行遗传算法种群初始化编码，将问题的解集初始化，接着对种群进行遗传算法基本操作，计算个体适应度值。最后循环操作直至找到相对最优适应度值对应的个体即为问题的最优解。遗传算法基本操作包括选择、交叉、变异。

遗传算法的优点是在使用概率搜索技术的同时，进行解空间的多点搜索，避免了陷入局部最小值，增加了搜索的灵活性，提高了系统的计算速度。

3.2 BP 神经网络

BP 神经网络是在非线性变换单元组成的多层前馈网络中应用后向传播误差算法进行学习，其工作机制是通过将误差回溯，然后利用梯度下降法对连接的参数进行更新，不断减少误差^[9]。

整个 BP 网络的运算过程主要包括前向传播输入和后向传播误差。由训练元组以及相关目标值组成的数据集 D ，对 D 中每个训练元组进行操作。

在前向传播输入的过程对输入层单元 j ，输出值等于输入值 $O_j=I_j$ ；对于隐含或输入层单元 j ， $I_j=\sum_i W_{ij}O_i+\theta_j$ ， $O_j=1/(1+e^{-I_j})$ ，其中， W_{ij} 是上一相邻层单元 i 到单元 j 的连接权重， θ_j 是单元 j 的偏倚值，用来改变单元的活性。

后向传播误差包括误差计算和权值与偏倚值更新。对每个输出层单元 j 的误差 $Err_j=O_j(1-O_j)(T_j-O_j)$ ，其中， T_j 表示 j 基于给定训练元组的已知目标值；对每个隐含层单元 j 的误差 $Err_j=O_j(1-O_j)\sum_k Err_k W_{jk}$ 。 K 单元是下一相邻层单元。将误差值回溯，更新权值和偏倚值。对网络中的每个权值 W_{ij} ，计算权重增量 $\Delta W_{ij}=lErr_j O_j$ ， $W_{ij}=W_{ij}+\Delta W_{ij}$ ；偏倚值增量 $\Delta \theta_j=lErr_j$ ， $\theta_j=\theta_j+\Delta \theta_j$ ， l 表示学习率。循环上述过程直至满足终止条件^[10]。

BP 神经网络根据所提供的数据，通过学习和训练找出输入与输出之间的内在联系从而求解问题的解，而不是完全依据对问题的经验知识，因而具有自适应特点，这对于弱化权重确定中的人为因素是十分有益的。但算法中网络的隐含层层数和单元数主要依靠经验确定，影响了模型预测结果的客观性。而由于 BP 神经网络的本质上是梯度下降法，

所要最小化的目标函数又比较复杂,因此容易陷入局部极小点且学习过程收敛速度慢。

4 优化 GA-BP 神经网络

神经网络参数的初始值,即神经网络的权值和阈值的初始值,决定了 BP 神经网络训练在误差曲面的开始点及隐含层权值在训练初期的调整量,值得大小会对计算时间及 BP 算法的预测质量产生影响,因此获取合适的网络参数至关重要。

神经网络结构的设计要考虑设几个隐含层和每个隐含层设几个隐含层节点的问题。对于隐含层数的设计,理论认为至多 2 层隐含层的前馈网可以映射所有连续函数,而节点数个数的设计会影响到网络从样本中获取信息的能力。隐含层节点的作用是从样本中提取并储存其规律,每个节点有若干个权值,每个权值都是增强网络映射能力的一个参数。节点数量太少,网络从样本中获取信息的能力就差,不足以概括和体现训练集中的样本规律;节点数量过多,有可能把样本中非规律性的内容如噪声等也学会记牢,从而出现所谓“过度拟合”问题,反而降低了泛化能力。所以,神经网络结构的设计也是运用 BP 神经网络进行预测前的准备要点。

利用 GA 优化神经网络,主要是借用了试凑法的思想,在全局范围内寻找最优网络结构与网络参数,使网络误差达到最小。GA 的全局寻优能力与 BP 神经网络的非线性拟合能力使 GA-BP 的结合应用备受研究人员关注。

4.1 优化算法设计

GA-BP 算法运用 GA 先优化神经网络拓扑结构再优化网络参数,最后回归 BP 预测模型。算法过程如图 1 所示。先对网络拓扑结构进行 GA 优化,即确定网络隐含层数、隐含层节点数的最优个体,接着,将最优个体代入网络参数的 GA 优化过程,优化网络连接权值与阈值。

在计算过程中,GA 与 BP 是紧密衔接的,在优化层数、节点数及权值、阈值时须反复调用 BP 用于适应度值计算;在得到最优的网络结构、网络参数后,需要回归到 BP 算法用于数据的预测。

4.2 算法操作

在 GA-BP 神经网络中,算法的操作主要包括遗传算法中包含的种群编码、遗传操作、适应度计算、以及 BP 学习训练等。

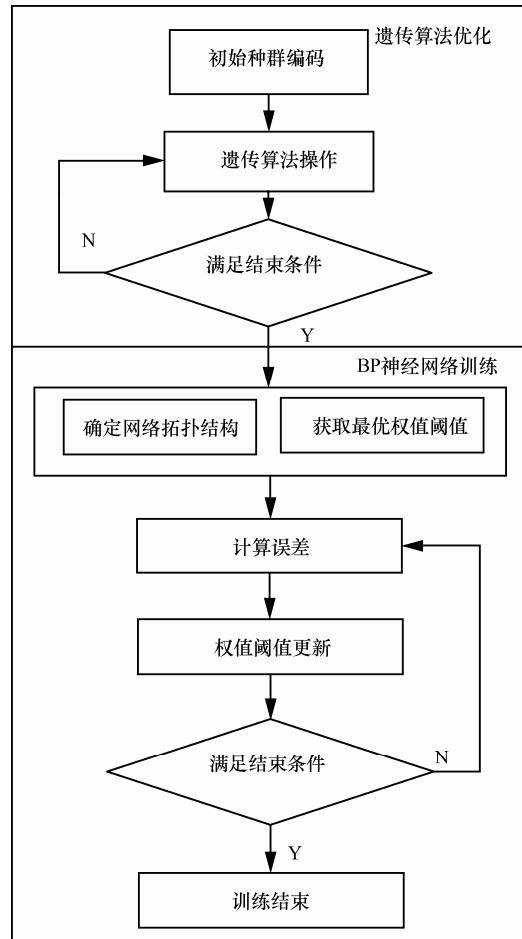


图 1 GA-BP 算法流程

4.2.1 遗传种群编码

在 GA-BP 神经网络中,遗传算法操作主要作用于网络拓扑结构及网络参数 2 个对象的优化。相应地,种群的编码也分为 2 种:网络拓扑结构编码及网络参数编码。网络拓扑结构的个体编码结构包含 3 部分:隐含层层数标识、第一隐含层的单元数、第二隐含层的单元数(包含 2 个隐含层的多层神经网络可以近似拟合所有函数)。网络参数的个体编码由输入层与隐含层连接权值、隐含层阈值、隐含层与输出层连接权值以及输出层阈值 4 部分组成^[1]。假设 BP 神经网络结构为 33-10-1,即输入层 33 个节点,隐含层 7 个节点,输出层 1 个节点,共 $33 \times 7 + 7 \times 1 = 238$ 个权值, $7 + 1 = 8$ 个阈值,所以遗传算法网络参数个体编码长度为 $238 + 8 = 246$ 。

4.2.2 遗传算法操作

遗传算法操作主要包括:选择、交叉和变异。

选择操作:采用轮盘赌法,基于适应度比例确定选择的策略,每个个体 i 的选择概率为

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (1)$$

其中, $f_i = k/F_i$, F_i 为个体 i 的适应度值, 由于适应度越小越好, 所以在个体选择前对适应度值求倒数; k 为系数; N 为种群个体数目。

交叉操作: 第 k 个染色体 a_k 和第 l 个染色体 a_l 在 j 位的交叉操作方法如下。

$$\left. \begin{aligned} a_{kj} &= a_{kj}(1-b) + a_{lj}b \\ a_{lj} &= a_{lj}(1-b) + a_{kj}b \end{aligned} \right\} \quad (2)$$

式(2)中, b 是[0,1]间的随机数。

变异操作: 选取第 i 个体的第 j 个基因 a_{ij} 进行变异的操作方法如下。

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max})f(g), r \geq 0.5 \\ a_{ij} + (a_{\min} - a_{ij})f(g), r < 0.5 \end{cases} \quad (3)$$

式(3)中, a_{\max} 为基因 a_{ij} 的上界; a_{\min} 为基因 a_{ij} 的下界; $f(g) = r_2(1 - g/G_{\max})$; r_2 为一个随机数; g 为当前迭代次数; G_{\max} 为最大进化次数; r 为[0,1]间的随机数。

4.2.3 适应度计算

进行训练时, 将期望输出设为 $y(k)$, 将调用 BP 神经网络算法产生的网络当前输出设为 $y_i(t)$, 目标函数为

$$F = \frac{1}{2} \sum_{k=1}^q \sum_{t=1}^s (y(k) - y_i(k))^2 \quad (4)$$

其中, q 为训练样本数, s 为网络输出节点数。

4.2.4 BP 神经网络学习训练

BP 神经网络的训练过程包括以下几个步骤: 网络初始化、隐含层及输出层输出计算、误差计算、权值及阈值更新。

网络初始化是根据系统输入输出序列(X ; Y) 确定网络输入层节点数 n 、隐含层节点数 l , 输出层节点数 m , 初始化输入层、隐含层和输出层神经元之间的连接权值 w_{ij} , w_{jk} , 初始化隐含层阈值 a , 输出层阈值 b , 给定学习速率和神经元激励函数。

隐含层、输出层输出计算是根据输入向量 X , 连接权值 w_{ij} 以及阈值 a , 计算隐含层输出 H , 并根据 w_{jk} 和阈值 b , 计算 BP 神经网络预测输出 O 。

误差计算是根据网络预测输出 O 和期望输出

Y , 计算网络预测误差 e 。权值、阈值更新是根据网络预测误差 e 更新网络连接权值 w_{ij} , w_{jk} 以及网络节点阈值 a , b 。

5 实证分析——基于金融产品营销预测

本文使用 GA 算法优化 BP 神经网络结构、参数, 以某银行短期理财产品营销的客户历史数据作为实证研究对象, 验证该模型可以更精确地预测银行理财产品的客户营销结果。

5.1 数据预处理

5.1.1 行为特征的选取

最初从银行得到的数据不乏冗余而对预测结果没有帮助的属性。一组经过量化后的营销行为属性, 可以区别 2 类营销行为具有显著的差异, 这样的属性称其为行为特征。从一组特征中挑选出最有效的特征以达到降低特征空间维数的目的, 这个过程叫特征选择。为了更精确的客户购买行为的预测结果, 需要对初始特征值进行选择。

特征选取的方法: 在给定维数 D 的特征空间内, 选择这样的 d 个特征($d \leq D$), 使营销成功与失败的 2 种类型能够尽可能的分隔开。对于成功营销的情况(A)和营销失败的情况(B)2 类, A 的样本中的任何一个属性与 B 中相对应的属性值都有一个距离, 把所有这些距离相加求平均, 则可求出在对应属性下 A、B 2 类的类距离。类距离分为类内距和类间距。类间距的计算方法采用的是以其中一类(A 类)作为基准点, 对于该类中的每个样本的有无贷款属性的取值与另一类(B 类)的每个样本进行比较, 如果取值相同则定义距离为零, 不同则定义距离为一个单位。同时采用上述方法, 计算 2 类的类内距。类内距是以同一类作为计算基准点。使类间距和类内距差别明显的属性就是一种重要的营销行为属性。

令 X_A , X_B 分别为 X 属性下, 成功营销和营销失败 2 种类型的样本, $d(x_A, x_B)$ 为这 2 个样本之间的距离, 则类距离的数学表达式为

$$L(x) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n d(x_{A_i}, x_{B_j}) \quad (5)$$

其中, m 为 A 类的样本数, n 为 B 类的样本数^[12]。

以是否有个人贷款属性为例, 计算 A、B 类的营销记录数据在该属性下的类距离。最终的计算结果如表 1 所示。

表 1 人贷款特征选择

贷款状态	营销结果比例/%		类距离	
	成功	失败	类间距	类内距
有	26.21	73.79	0.63	0.14
无	13.35	86.65		

显然，是否有个人贷款这一属性的类间距明显高于类内距。由此可以推断，有无贷款这一用户属性可以有效地将 2 类营销结果区分开。对于营销结果预测，这就是一种重要行为特征。

根据特征选择的方法，最终从众多行为属性中选取了近 10 个特征属性，如表 2 所示。其中， y 的取值为 1 或 0，表示购买与不购买。

5.1.2 样本的设计

训练样本的数量是影响网络分析能力的一个重要因素，样本数量少则网络的健壮性差、泛化性能低，而样本数量过多，训练时间则会过长。本文所采用的实证数据共计 20 200 条，包含客户基本信息、营销信息和最后的营销结果 3 类信息。将得到数据随机选择 20 000 条数据作为训练数据集进行网络的训练，剩余的 200 条数据作为测试数据集进行模型拟合精度的测算。

表 2 为特征属性描述

字段	含义
age	年龄
job	工作类型
marital	婚姻状况
education	教育状况
balance	年末个人收支状况
loan	是否有个人贷款
contact	营销方式
campaign	单客户营销次数
duration	单客户平均营销时长
pdays	距最近一次客户营销的天数
poutcome	此前营销结果
Y	最终是否购买

5.1.3 数据预处理

对样本数据进行归一化处理，先将模型的各个属性量化数字化表示，然后将其分成离散型变量和连续型变量 2 组。

对于离散型变量，例如客户工作类型、是否有贷款、此前采用的营销方式等采用最小—最大归一化方法进行处理。

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{6}$$

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

对于连续型变量，例如单一客户营销时长、年末个人收支状况等通过对变量值的分布状况分析，用如下归一化方法处理。

$$y = \varphi\left(\frac{x - \mu}{\sigma}\right) \tag{7}$$

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。

5.2 经网络拓扑结构的确定

神经网络拓扑结构的确定主要包括隐含层数和隐含层节点数。神经网络的结构对最后预测结果的影响很大，为了确定合适的神经网络结构，最好的办法是在网络的学习过程中根据环境要求自组织地学习、调整网络结构，最后得到一个大小适合的网络结构。

使用 GA 遗传算法循环迭代可以实现神经网络结构的自发调整。遗传算法的迭代次数是 100 次，种群规模为 10，交叉概率为 0.3，变异概率为 0.1。适应度值可以反映模型的拟合情况，优化过程中最优个体的适应度值变化曲线如图 2 所示。

得到的最优网络结构个体为“001000011110”，根据网络结构编码规则将码串进行解码操作，最终选定满足网络最小误差要求的隐含层层数为 1，隐含层节点数为 16。

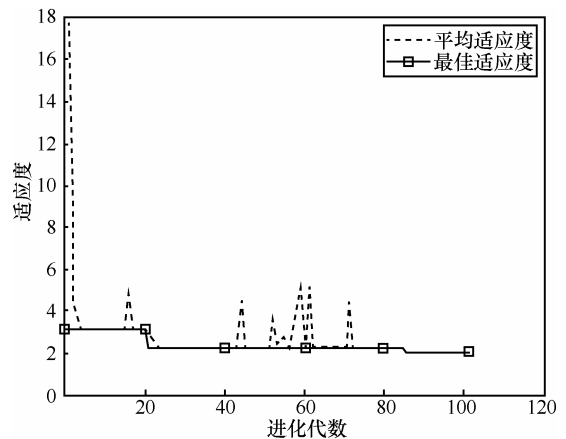


图 2 适应度曲线(100 代)

5.3 BP 与 GA-BP 实验结果对比

仿真实验过程首先针对同一组数据，先不借助

遗传算法，直接对营销结果进行仿真预测如图 3 所示；随后，使用遗传算法优化神经网络的结构和参数初始值。预测结果如下图 4 所示。

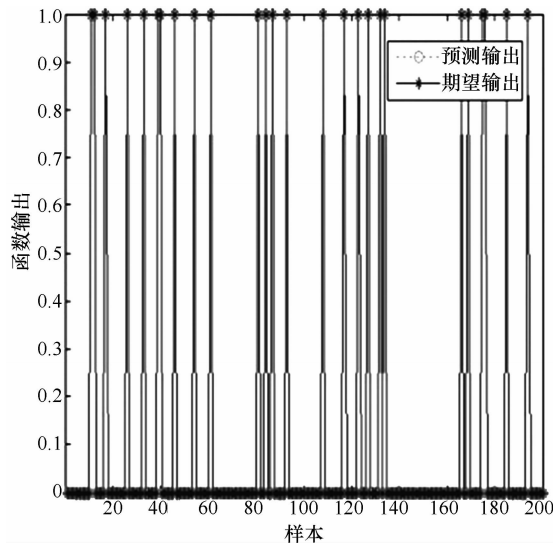


图 3 BP 神经网络输出

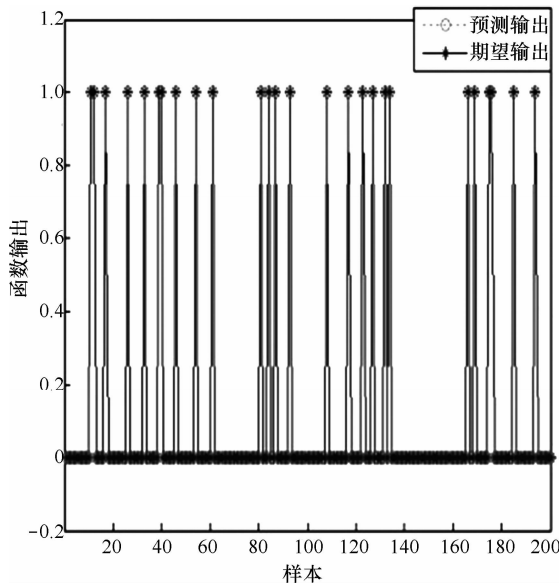


图 4 GA-BP 神经网络输出

在程序设计中，以 0.5 为分界，概率值大于 0.5，则购买决策为 1，即购买，反之，则为 0，不购买。多次运行算法显示 2 种方法的正确率均可达到 90% 以上。

通过 BP 与 GA-BP 误差对比图（如图 5 所示）可以看出，遗传算法优化的 BP 网络以近似于 0 的概率误差预测出 200 次中每次 y 取的确切值。与直接的 BP 神经网络相比，达到了较小的误差率。

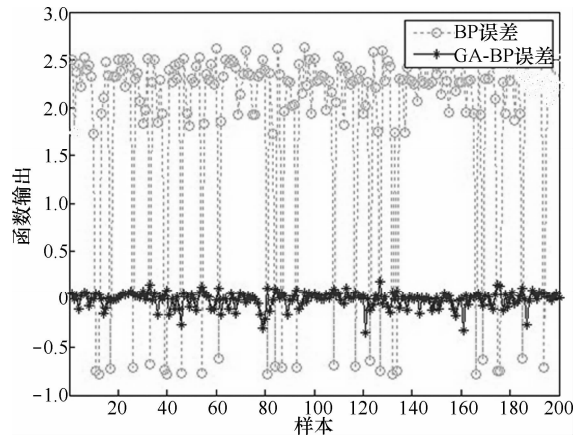


图 5 BP 与 GA-BP 预测输出误差对比

6 结束语

神经网络技术主要根据所提供的数据，通过学习和训练，找出输入与输出之间的内在联系，从而求取问题的解。把 GA 算法引入 BP 神经网络结构和参数的确定过程，减少了网络结构及参数确定的主观性。实验使用银行理财产品销售数据验证了 GA-BP 神经网络算法具有更小的误差率。本模型的缺陷在于对隐含层节点数进行优化时，需要循环套用 BP 算法，运行时间长，对系统性能要求高。在实际应用中，对客户营销新数据运用该模型进行计算，可以得到更为精确的预测结果。

参考文献：

- [1] 董槐林, 郭阳, 郑宇辉. 基于 PSO-BP 神经网络的企业信用风险评估模型研究 [J]. 计算机与现代化, 2009, 4(4): 123-126.
DONG K L, GUO Y, ZHENG Y H. Research on evaluation model of enterprise credit risk based on PSO-BP neural networks[J]. Modern Computer, 2009,4(4):123-126.
- [2] HEBB D O. The Organization of Behavior: A Neuropsychological Approach[M]. John Wiley & Sons, 1949.
- [3] MCCLELLAND J L, RUMELHART D E. Parallel distributed processing[J]. Psychological and biological models, 1987, 2:76-82.
- [4] WIDROW B, HOFF M E. Adapting switching circuits[A]. IRE WESCON Conv Record[C]. 1960. 96-104.
- [5] 李红, 杨剑锋. 基于改进的 BP 神经网络模型参考自适应控制[J]. 兰州交通大学学报, 2011, 30(1):37-41.
LI H, YANG J F. Reference adaptive control based on improved BP neural network[J]. Journal of Lanzhou Jiaotong University. 2011, 30(1): 37-41.
- [6] 张卫东, 韩云昊, 米阳. 基于 GA-BP 模糊神经网络的商业银行信用风险评估[J]. 工业工程与管理, 2006, 5: 80-84.
ZHANG W D, HAN Y H, MI Y. A study on credit risk assessment of commercial banks based on GA BP fuzzy neural network[J]. Industrial Engineering and Management. 2006,5:80-84.

(下转第 32 页)

[11] CANETTI R, HALEVI S, KATZ J. A forward-secure public-key encryption scheme[A]. Cryptology—Eurocrypt 2003[C]. Springer Berlin Heidelberg, 2003.255-271.

[12] OKAMOTO T, TAKASHIMA K. Fully secure functional encryption with general relations from the decisional linear assumption[A]. Cryptology—CRYPTO 2010[C]. Springer Berlin Heidelberg, 2010. 191-208.

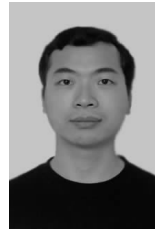
[13] LEWKO A, OKAMOTO T, SAHAI A, *et al.* Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption[A]. Cryptology—EUROCRYPT 2010[C]. Springer Berlin Heidelberg, 2010.62-91.

[14] BONEH D, GOH E J, NISSIM K. Evaluating 2-DNF Formulas on Ciphertexts[M]. Theory of Cryptography. Springer Berlin Heidelberg, 2005.325-341.

[15] BEIMEL A. Secure Schemes for Secret Sharing and Key Distribution[D]. PhD Thesis, Israel Institute of Technology, Technion, Haifa, Israel, 1996.



马建峰 (1963-), 男, 陕西西安人, 博士, 西安电子科技大学教授、博士生导师, 主要研究方向为密码学、计算机网络与信息安全。



熊金波 (1981-), 男, 湖南益阳人, 福建师范大学讲师, 主要研究方向为访问控制技术与结构化文档安全。

作者简介:



李琦 (1989-), 男, 江苏淮安人, 西安电子科技大学博士生, 主要研究方向为信息安全与基于属性的密码学。



刘西蒙 (1988-), 男, 陕西西安人, 西安电子科技大学博士生, 主要研究方向为公钥密码学与信息安全、安全网络编码及其应用。

(上接第 25 页)

[7] 席道瑛, 张涛. BP 网络的改进及模拟退火神经网络在地学中应用[J]. 物探化探计算技术, 1996, 18(3):219-224.

XI D Y, ZHANG T. Application of improved BP neural network and simulated annealing neural network in Earth Science[J]. Computing Techniques for Geophysical and Geochemical Exploration. 1996, 18(3): 219-224.

[8] 苏晨, 李成义. 基于遗传算法和 BP 神经网络的服装销售预测[J]. 经营与管理, 2012, 2: 145-158.

SU C, LI C Y. Garment sales prediction of based on genetic algorithm and BP neural network[J]. Operation and Management, 2012, 2: 145-158.

[9] TAN P N, MICHAEL S, VIPIN K. Introduction to Data Mining[M]. Beijing: Machine Industry Press, 2010:246-255.

[10] HAN J, KAMBER M, PEI J. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann, 2006.

[11] 傅荟璇, 赵红等. MATLAB 神经网络应用设计[M]. 机械工业出版社, 2010.

FU H X, ZHAO H, *et al.* Neural network application design based on MATLAB[M]. Beijing: Machine Industry Press, 2010.

[12] HECTOR C. Practical data analysis[M]. Beijing: Machine Industry Press, 2014.

作者简介:



金鑫 (1973-), 男, 内蒙古乌海人, 博士, 中央财经大学教授, 主要研究方向为商务智能。

潘宜安 (1991-), 女, 福建福州人, 中央财经大学硕士生, 主要研究方向为数据挖掘。

吴靖 (1957-), 女, 北京人, 中央财经大学教授, 主要研究方向为信息经济。