

# BP神经网络在福利彩票预测中的应用

涂晔, 车文刚

昆明理工大学, 信息工程与自动化学院, 云南, 昆明, 650051

**摘要:** 采用BP网络的三种改进算法, 对福彩双色球(2008073期~2008133期)的历史中奖数据进行分析, 建立了基于BP神经网络的双色球预测模型, 并对三种改进算法的训练结果进行了比较。与现存的彩票分析方法不同的是, 本方法并非基于频数分析缩小样本集来产生预测号码, 而是基于BP神经网络对历史数据进行学习, 挖掘其中隐含的规律, 生成用来预测号码的神经网络节点的连接参数, 再根据这些参数产生下一期预测号码。

**关键词:** BP神经网络 改进算法 福彩预测

## 0 引言

彩票业在我国是一个发展空间极大的新兴产业, 彩票业不仅在特定领域发挥了重要作用, 还影响到出版、印刷、电视、广播、交通等多种行业的发展。积极发展我国的彩票业, 将对我国经济建设和社会发展产生更加积极的影响。目前在中国, 福利彩票双色球以每星期开奖三次, 最高奖额五百万元人民币, 每期奖池奖金累计过亿的雄厚实力, 吸引着庞大数量的彩民, 而彩民朋友们最为关心的就是如何预测当期的开奖号码, 因此对福利彩票双色球奖号的预测具有十分重要的研究意义。

## 1 福彩双色球预测依据

首先必须强调的是, 只要摇奖机的构造能够保证任何一个球的出现都完全随机, 那么, 每个号码的出现都具有相等可能性, 即出现的机率都是一样的。据此, 根据福利彩票双色球的游戏规则, 数字完全自由组合后, 将可能产生一千七百万组不同号码, 任何一个彩民购买一组号码, 他的中奖机率都只有一千七百万分之一, 无论什么“专家”或者“大师”都无法准确预测所谓的中奖号码。

然而, 把每期出现的中奖号码进行统计, 积累

到一定量之后, 就能发现各数字总共出现的频率特性。如奇偶比, 遗漏期K值, 区段出号码数字, 数出频率的次数, 二数同出数组等。例如“上海风采”福利彩票网站根据上海风采发售以来的历史资料, 认为数字分布具有一定的规律性: ①奇数与偶数出现的次数应各占总数的1/2; ②1-18数组, 19-35数组的出现次数各占总数的1/2; ③1-12区段, 13-24区段, 25-35区段, 三段数的出现应各占总数的1/3。

既然奖号的出现有了规律可循, 各种各样的彩票预测方式也应运而生, 这些方式都是通过数字统计与图表分析, 运用概率论理论来判断冷热号码出现的周期, 分析号码可能出现的区段, 缩小精选号码范围, 为新一期选择投注号码提供依据, 从而达到排除某些号码组合, 减小号码排列样本集, 从而提高中奖率的目的。虽然号码是根据科学依据推算出来的, 但这些选号方式明显带有个人喜好与主观意愿。这是因为从统计学的大数规律讲, 再大的样本集也是估计, 都有不中的可能。于是, 研究其它科学有效的测算途径就显得迫切需要了。

双色球号码预测过程可认为是一种复杂非线性函数关系的逼近过程, 要对其进行较为准确的预测, 就必须采用能捕捉非线性变化规律的预测方法。BP神经网络具有逼近任意连续函数和非线性映射的能力, 并可进行高维非线性的精确映射。它是由大量的, 同时也是很简单的神经元通过广泛的互相连接而形成的复杂网络系统, 它反映了人脑功能的许多基本特性。其信息处理是通过学习, 动态修改各神经元之间的连接权值使系统不断产生“进化”来实现。

本文所要讨论的方法就是基于BP神经网络对福彩双色球历史数据进行学习, 挖掘中奖号码中隐含的规律, 生成用来预测号码的神经网络节点的连接参数, 然后再根据这些参数生成下一期预测号码。

## 2 BP神经网络及其改进学习算法

### 2.1 BP神经网络原理

BP神经网络是一种具有三层或三层以上阶层型结

构的神经网络。层间各种神经元实现全连接，即下层的每一个神经元与上层的每个神经元都实现权连接，而每层各神经元之间不连接。网络中每一层神经元的连接权值都可以通过学习来调整。当给定一个输入节点数为 $n$ 、输出节点数为 $s$ 的BP网络，输入信号由输入层到输出层传递，通过非线性函数的复合来完成从 $n$ 维欧氏空间到 $s$ 维欧氏空间的映射，该过程是向前传播的过程；如果实际输出信号存在误差，网络就转入误差反向传播过程，并根据误差的大小来调节各层神经元之间的连接权值。当误差达到人们所希望的要求时，网络的学习过程就结束。

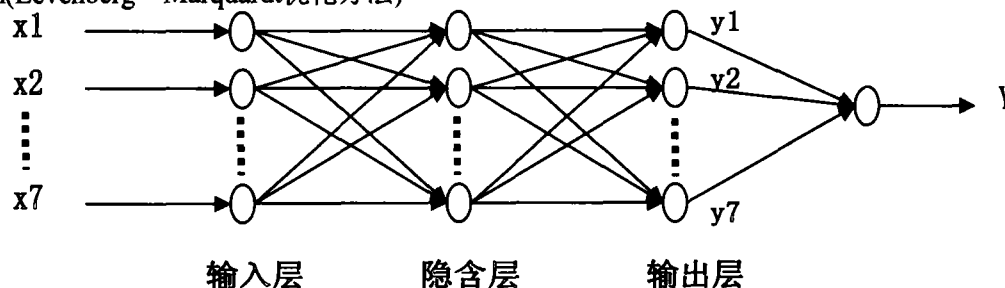
## 2.2 BP网络改进学习算法

传统 BP 网络的学习算法实质上是一种简单的最速下降静态寻优算法，在修正权值  $W(k)$  时，只是按照  $k$  时刻的负梯度方式进行修正，而没有考虑到以前积累的经验，即以以前时刻的梯度方向，从而常常使学习过程发生振荡，通常具有收敛速度慢、易陷入局部极小值和网络推广能力不强等方面的缺陷。本文采用 MATLAB 神经网络工具箱提供的三种改进快速学习算法，有效克服了传统 BP 网络学习算法的缺陷。

### 2.2.1 traingdm(动量梯度下降反向传播算法)

该算法在对权重和阈值更新时不仅考虑当前的梯度方向，而且还考虑了前一时刻的梯度方向，从而降低了网络性能对参数调整的敏感性，有效地抑制了局部极小，但收敛速度较慢。

### 2.2.2 trainlm(Levenberg—Marquardt优化方法)



## 3.2 BP 网络设计中的数据处理

3.2.1 输入数据的预处理。由于BP神经网络的隐层一般采用Sigmoid转换函数，所以，需要对输入数据进行归一化处理，这里采用如下公式：

Levenberg—Marquardt优化方法能够根据网络训练误差变化情况，自动调节网络训练参数 $\mu$ ，使网络实时采取适宜的训练方法。当 $\mu$ 较小，网络训练过程主要依据Gauss—Newton法；当 $\mu$ 较大，网络训练过程主要依据梯度下降算法。该训练函数的效率优于梯度下降法，但占用内存较大，适用于规模较小的网络。

### 2.2.3 trainbr(Levenberg—Marquardt 优化方法与 Bayesian正则化方法)

trainbr函数采用Levenberg-Marquardt优化方法进行网络权值和阈值的最优化搜索，并采用Bayesian正则化方法在网络训练过程中自适应地调节性能函数比例系数的大小，使其达到最优，并且采用trainbr函数训练后BP网络具有较好的推广能力。

## 3 BP 神经网络的双色球预测模型的确定

### 3.1 BP 网络模型确定

输入层和输出层确定。福彩双色球的备选号码为 01-33 号红球加上 01-16 号蓝球，其中从 33 个红球中选择 6 个号码，从 16 个蓝球中选择 1 个号码，这样组成 6+1 的组合模式。每组历史中奖号码均为 7 个数字，因此，输入层神经元个数定为： $n=7$ 。网络的输出，即为预测的双色球中奖号码，所以输出层神经元个数为： $s=7$ 。

BP 神经网络预测模型如下图所示：

$$\bar{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \times d_1 + d_2 \quad \text{式中：} X_{\max}、X_{\min} \text{ 分别为输入}$$

样本 $X_i$ 的最大值和最小值， $\bar{X}_i$ 为处理后的样本输入值， $d_1$ 、 $d_2$ 分别取0.998和0.001。

3.2.2 输出数据的反变换处理。为了防止部分神经元达到过饱和状态, 对输出样本亦需作如上变换。这样实际上将输入输出样本变换到[0.001, 0.999]区间, 避免了0和1两个值。因为0和1恰好是Sigmoid型函数的极小值和极大值, 要求联接权足够大才能使网络的输出值与其匹配, 从而需要相当多的训练次数来不断修正权值, 导致训练速度缓慢。因此网络输出的结果要进行反变换才能得到预测值。

### 3.3 双色球预测神经网络模型 MATLAB 仿真实现

#### 3.3.1 创建改进的 BP 网络

使用 MATLAB 利用如下代码创建一个符合上述要求的 BP 网络: threshold=[0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1]; net=newff(threshold,[隐层节点数,7],{'tansig','logsig'},'训练函数'); 其中, threshold 设定了网络输入向量的取值范围[0, 1]和输入神经元个数为 7。输出层节点数为 7, 输入层到隐含层的变换函数为 tansig, 隐含层到输出层的变换函数为 logsig, 网络所用的训练函数将分别采用三种改进的 BP 学习算法, 分别为: traindm, trainlm, trainbr。

#### 3.3.2 BP 网络的训练与测试

①隐含层节点的确定。分别选取节点数为 15, 18, 20 来对网络进行训练, 设定预测精度为 0.01, 综合考虑网络性能和训练速度, 发现隐含层节点数为 18 三种改进算法的拟合结果如下表:

Traindm函数		Trainlm函数		Trainbr函数	
有效性系数	相对误差	有效性系数	相对误差	有效性系数	相对误差
0.745	0.211	0.998	0.015	0.992	0.019

测试样本为2008130期的数据, 即(01 12 18 28 30 32+06), 期望输出(实际值)为2008131期数据, 即(02 12 15 18 22 32+08), 分别运用三种改进BP算法进行预测。预测值与实际值比较如下:

	实际值	预测值
Traindm 函数	02 12 15 18 22 32+08	01 11 16 18 22 33+06
Trainlm 函数		02 12 14 19 21 32+08
Trainbr 函数		02 11 14 18 22 33+08

2008131 期各媒体专家工作室预测选登 (来自 <http://www.shuangsejiu.org/Article.asp?ID=437>):

个时, 网络对函数的逼近效果最好, 20 个时其次, 15 个时最差。

②网络测试。经预处理后的训练样本数加载到所建立的人工神经网络上, 即可用BP算法对网络训练。为了检验训练后神经网络解决问题的能力, 通常需要测试样本对训练后的网络进行测试和检验, 以便对比分析预测结果和实际值之间的误差。通过训练和测试之后的神经网络模型才可用于双色球奖号的预测。

## 4 预测结果分析

以双色球历史开奖数据2008073期~2008133期的数据作为训练样本, 2008130期的数据作为测试样本, 对BP网络进行训练仿真。根据网络多次训练结果比较, 确定隐层节点数为18, 期望误差为0.02。用有效性系数COE和相对误差EQ 来判断经三种改进算法训练后网络的性能, 比较拟合和预测效果, 公式如下:

$$COE = 1 - \frac{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}{\frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2}, \quad EQ = \frac{\sum_{i=1}^m |y_i - \hat{y}_i|}{\sum_{i=1}^m y_i}$$

	红球	蓝球
实际值	02 12 15 18 22 32	08
幸运博士专家工作室	03 07 08 19 23 24 27 28 32 33	01 08
好运来了专家工作室	07 11 12 16 18 19 23 26 31 33	01 16
重庆火锅专家工作室	02 05 14 16 20 22 23 27 29 31	11 08
中国双色球百万工作室	07 09 13 17 19 23 24 28 30 32 33	03 07 08 09
中国双色球后庄工作室	02 04 06 11 15 18 20 23 28 30 32	05 06 08 12

由训练结果和预测结果可知,对传统BP网络学习算法作出不同改进的三种学习算法在实际应用中它们的性能各不相同。`traingdm`函数拟合精度最低,这是因为`traingdm`函数虽然采用了动量梯度下降反向传播算法,能够有效地抑制局部极小,但其收敛速度太慢,训练多次也不能达到期望误差,而且仍然难以找出最优的网络权值和阈值。而`trainlm`和`trainbr`函数均采用Levenberg—Marquardt优化方法,能够快速获得最优的网络权值和阈值,收敛速度较快。虽然`trainlm`函数的有效性系数都要比`trainbr`函数高,相对误差也都比`trainbr`函数低。然而,推广能力是衡量神经网络性能好坏的重要标志。`trainbr`函数采用了Bayesian正则化方法,通过修正神经网络的训练性能函数来提高推广能力,防止出现“过度训练”。虽然有时候`trainbr`函数对训练样本的拟合精度要低于`trainlm`函数,但其有效性系数都高于0.8,甚至绝大多数都大于0.9,相对误差也是绝大多数都小于0.01。经多次数值实验可知,采用`trainbr`函数训练的BP网络稳定性要优于`trainlm`函数和`traingdm`函数,且通常只需几百次的训练就能使网络收敛,找到最优的网络权值和阈值。

## 5 结论与认识

1.本文采用的三种改进算法, `trainbr`函数采用了Levenberg-Marquardt优化方法与Bayesian正则化方法,具有较好的推广能力,预测精度高于`traingdm`函数和`trainlm`函数,且经其训练的网络具有较好的稳定性。

2.作为数据驱动型模型的BP网络,用来建模的学习样本质量至关重要,它直接影响网络的预测精度。`trainbr`函数具有较好的推广能力,但其外推能力仍然较差。若采用引入部分常用公式的结果来增加样本以便让BP网络对样本充分学习,但是如果这种类型的训练样本过多,神经网络模型逼近的函数就会趋近于传统公式,从而失去了采用新模型的意义。所以今后应寻求能增强神经网络外推能力的算法来改进这项有意义的工作。

3.BP神经网络自身还存在着一些缺陷,特别是网络隐含层的层数及隐含层的节点数的选取尚无理论依据,通常是根据实验或经验确定,因此使网络具有很大的冗余性,也会影响拟合和预测效果,而本文选取的隐含层节点数18就是根据多次试验得出。

4.BP神经网络突出的优点是具有很强的非线性映射能力和柔性的网络结构,与传统的人为分析统计建模方法相比,预测精度较高,而且预测方法较为科学。

## 参考文献:

- [1] 高隼.神经网络原理及仿真实例[M].机械工业出版社.2003.8
- [2] 董长虹.Matlab神经网络与应用[M].国防工业出版社.2005.1
- [3] 朱武亭,刘以建.BP网络应用中的问题及其解决.上海海事大学学报.2005, 26(2)
- [4] 陈晓梅,马晓茜.基于改进BP神经网络的锅炉结渣预测模型.华东电力.2005, 33(7)
- [5] 顾军华,宋丽娟,赵文海.基于改进BP神经网络的税收收入预测模型.河北工业大学学报.2003, 32(1)
- [6] 杨琴,谢淑云.BP神经网络在洞庭湖氨氮浓度预测中的应用.水资源与水工程学报.2006, 17(1)

**Abstract:** According to the analysis of (No.2008073~No.2008133) historic recorded data of welfare lottery-Shuangsequiu, a model is developed to predict welfare lottery based on BP neural network. The model is optimized with three different improved BP algorithms and the training results are compared. It's different from the existed analyzing methods of lottery. This method isn't based on times analysis and reducing sample set to predict data, but based on BP neural network to learn the historic recorded data and find the hidden regularity, then automatically produce the next predicting data according to the connecting parameter of neural network joint.

**Key words:** BP neural network improved algorithm welfare lottery predicting