

第四届“互动出版杯”数学中国

数学建模网络挑战赛

承 诺 书

我们仔细阅读了第四届“互动出版杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：

参赛队员（签名）：

队员 1：谢兰军

队员 2：刘宗磊

队员 3：汪海萍

参赛队教练员（签名）： 张建勇

参赛队伍组别：本科组

第四届“互动出版杯”数学中国

数学建模网络挑战赛

编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：
#1225

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

2011 年第四届“互动出版杯”数学中国 数学建模网络挑战赛第二阶段

题 目 保费浮动方案及风险评估研究

关 键 词 BP 神经网络 保费浮动 灰色预测 定量分析

摘 要：

本文使用了无量纲化、数值归一化等方法对收集到的数据进行处理，突出了数据的重要信息，弱化了次要信息，使得对数据的处理更加简单有效，也为模型的建立和评估结果的得出提供了保证。

问题一，影响保费浮动的因素众多，综合有保险期限、险种大类、车辆用途、使用性质、车辆类型、展业方式、保额、出险次数、汽车品牌共 9 种。本文致力于研究融合各影响因素保费浮动方案，建立基于 BP 神经网络的保费浮动体系。结合第一阶段的结论：出险次数对续保率有很大影响，参考《2011 年度北京地区机动车商业保险费率浮动档次升降方案》将保费浮动体系分为四大块，见表 1.1。该方案使得大部分投保者受益，这对降低车辆出险率、鼓励保户续保、发展潜在保户有着很大的拉动作用。

问题二，对保险公司潜在风险的评估，从理赔支出和信誉度两方面考虑，公司的风险指数与理赔支出、信誉度正相关。分析所给数据，将其按时间从 2010/1/1 到 2010/3/31 分为六个阶段，结合灰色预测模型，预测下一阶段该公司的各项指标（见表 2.4）。通过风险评估方法中的定量分析对该公司风险指数计算，得出该公司的资金风险指数及信誉风险指数（见表 2.5）。可知该公司在信誉上存在很大风险，建议该保险公司下阶段内，提高办事效率，减少拖延处理的保单，提高公司信誉度，塑造良好形象。

参赛队号 1225

参赛密码 _____
(由组委会填写)

所选题目 C

ABSTRACT

This article uses the non-dimensional, numerical methods such as normalization of the data collected for processing. With this method, the important information data can be highlighted and what's more, the secondary information could be weaken, making data processing easier and more efficient. And the method provides guarantee for the whole model and evaluating of the results.

Talking about the first problem, we find that there are so many factors which can affect the premium float. It includes the insurance period, insurance category, vehicle use, type of use, vehicle types, and methods to sell the insurance, the total amount of the insurance, the number of accidents, a total of nine kinds of car brands. This integration of the various factors working on the pre BP neural network to establish premiums based on a premium floating system. Combining with the conclusion of the first phase, it says that the number of accidents has a great influence on the renewal rate. Reference to "Commercial Motor Vehicles in Beijing 2011 the Floating Rate of Premium Quality Lifting Program," the premium float system is divided into four blocks (Table 1.1). The program benefits the most of the insured people. At the same time, it decreases the rate of the accidents, encourages more and more people to renew the insurance, and develops potential clients in a large part.

Turning to the second problem, we assess the potential risks of insurance company from its claims expenses and credibility. Company's risk index is related to claims expenses and credibility. After analyzing the data, we divide the time from 2010/1/1 to 2010/3/31 into six stages, predict the next stage of the company's indicators (Table 2.4) based on Grey prediction model, and calculate the company's risk index using quantitative analysis of risk assessment method. It comes to the results about the company's financial risk index and the credibility of the risk index (Table 2.5). The results show that there is a big risk on the credibility of the company. We recommend that the insurance company should improve efficiency and reduce delays in processing to improve the company's reputation, and create a good image.

Keywords: BP neural network Premium floating Greg Model Quantitative Analysis

一、问题重述

问题 1:

汽车保险公司为了降低车辆出险率,鼓励保户续保,发展潜在保户,通常都会对满足一定要求的保户或者投保人给与一定比例的保费浮动优惠,就是通常所说的保费折扣。请根据附件中的参考数据,以及第一阶段中对于影响续保率因素的分析,给出一套较为合理的保费浮动方案。

问题 2:

一些大型的保险公司要在全中国很多地区设立分公司。总公司每年要对分公司的业绩情况进行考核,考核结果直接影响分公司领导班子的去留。传统的考核方法就是计算分公司的保费收入和理赔支出的差额。一些分公司为了提高自己的考核成绩,会使用受理一些风险较大的投保或者故意拖延理赔的处理时间等方法。因此,很多保险公司开始考虑引入风险评估机制来对分公司进行考核,潜在风险较低的分公司会得到较高的考核成绩,请建立合理的模型对参考数据中的汽车保险公司进行潜在风险的评估,并通过对模型的深入分析对该公司今后的风险控制提出建议。

二、问题分析

问题一:

2007 年 7 月 1 日后,与交通违法事故挂钩的交强险费率浮动机制将在全国范围内实行,车主在上一年度交强险保单期满进行续保时将享受浮动费率,安全驾驶者可享受优惠的费率,交通肇事者将负担高额的保费。也就是说,一辆车如果多次出险,来年的保费就会很快上涨,而常年遵守交法不出险的,保费会逐年降低,以此实现“奖优罚劣”。这样既能加强驾驶者的安全意识,防范交通安全事故,又能提升保险公司的利润空间。那么,费率怎样浮动至关重要。制定保费浮动优惠的目的是为了降低车辆出险率,鼓励续保,发展潜在保护。为此,我们所要实现:一,针对每一个不同的保户或投保人,在其提供具体指标后,能明确给出保费折扣;二,对于投保人,在保险公司购保时能清楚自己属于哪一类客户,即汽车保险公司需提供一份保费浮动方案供保护参考。从参考数据中可看到对不同的客户有不同的保费折扣,那么客户的各个不同指标与保费折扣有着紧密的联系,是复杂的非线性关系,利用 BP 神经网络处理大量数据建立客户指标与保费折扣之间的网络关系之后,对具体的客户就能利用神经网络拟合出保费折扣比率。

问题二:

风险评估(Risk Assessment)是指,在风险事件发生之前或之后(但还没有结束),该事件给人们的生活、生命、财产等各个方面造成的影响和损失的可能性进行量化评估的工作。即风险评估就是量化测评某一事件或事物带来的影响或损失的可能程度。

保险公司的风险是由高风险投保人投保造成理赔支出增多和保险公司自身拖延理赔、未及时处理等售后服务导致信誉度下降造成的。本文从资金风险及信誉度两方面考虑,采用定量分析方法对保险公司风险指数定量评估。另外,对该保险公司的发展前景分析,参考前三个月该公司数据,采用灰色理论 $G(1,1)$ 模型预测后阶段风险指数,提出相应建议。

三、模型的假设

- [1]参考数据准确可靠;
- [2]出险次数对保费折扣的影响最大;
- [3]该公司的保费浮动方案短期内不会有较大变化;
- [4]该公司注重公司信誉;
- [5]信誉指标为 0.95;

四、符号说明

| | |
|-------|--------------------------|
| y | 折扣比率; |
| net | 神经网络; |
| x | 投保人的各项信息矩阵; |
| EF | 暴露因子——理赔对保险公司造成损失的程度; |
| SLE | 单一损失期望——特定威胁可能造成的潜在损失总量; |
| ARO | 理赔在半个月内发生的频率; |
| CAR | 信誉风险指数; |
| UNS | 未决率; |
| DCS | 据赔率; |
| FAR | 资金风险指数; |
| ACR | 出险率; |
| INF | 在阶段内的收入; |
| EXP | 在阶段内的理赔支出; |

五、模型的建立与求解

问题一

1 相关理论基础

BP 算法基本原理：

BP (Back Propagation)神经网络，即误差反传误差反向传播算法的学习过程，由信息的正向传播和误差的反向传播两个过程组成。输入层各神经元负责接收来自外界的输入信息，并传递给中间层各神经元；中间层是内部信息处理层，负责信息变换，根据信息变化能力的需求，中间层可以设计为单隐层或者多隐层结构；最后一个隐层传递到输出层各神经元的的信息，经进一步处理后，完成一次学习的正向传播处理过程，由输出层向外界输出信息处理结果。当实际输出与期望输出不符时，进入误差的反向传播阶段。误差通过输出层，按误差梯度下降的方式修正各层权值，向隐层、输入层逐层反传。周而复始的信息正向传播和误差反向传播过程，是各层权值不断调整的过程，也是神经网络学习训练的过程，此过程一直进行到网络输出的误差减少到可以接受的程度，或者预先设定的学习次数为止。

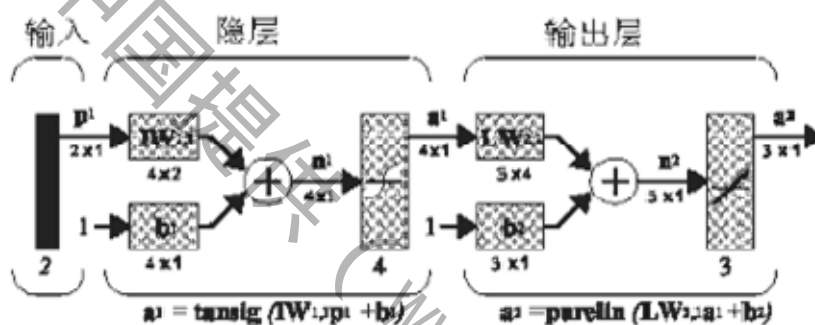


图 1.1 BP 神经网络模型

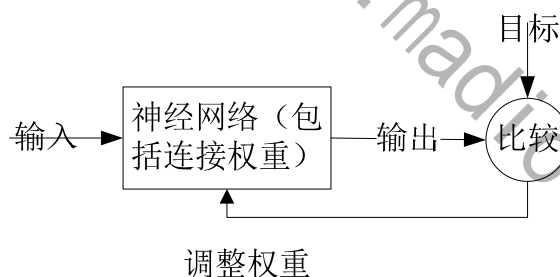


图 1.2 学习和调整权重

MATLAB 工具箱中神经网络参数设置

1.1 Newff

函数功能：构建一个 BP 神经网络。

函数形式：net=newff(P, T, S, TF, BTF, BLF, PF, IPF, OPF, DDF)

P: 输入数据矩阵。

T: 输出数据矩阵。

S: 隐函数节点数。

TF: 节点传递函数，包括硬限制幅传递函数 hardlim，对称硬限制幅传递函数 hardlims，线性传递函数 purelin，正切 S 型传递函数 tansig，对数型传递函数 logsig。

BTF: 训练函数, 包括梯度下降 BP 算法训练函数 **traingd**, 动量反传的梯度下降 BP 算法训练函数 **traingdm**, 动态自适应学习率的梯度下降 BP 算法训练函数 **traingda**, 动量反转和动态自适应学习率的梯度下降 BP 算法训练函数 **traingdx**, Levenberg-Marquardt 的 BP 算法训练函数 **trainlm**……

列表:

| 训练算法缩写 | MATLAB 函数 | 说明 |
|--------|-----------|------------------------|
| GD | traingd | 标准梯度下降算法 |
| GDM | traingdm | 带动态的梯度下降算法 |
| GDA | traingda | 可变学习速率梯度算法 |
| LM | trainlm | Levenberg-Marquardt 算法 |
| BFG | trainbfg | BFGS 拟牛顿法 |
| RP | trainrp | 弹性梯度算法 |
| SCG | trainscg | Scaled 共轭梯度算法 |
| CGB | traincgb | Powell-Beale 共轭梯度算法 |
| CGF | traincgf | Fletcher-Powell 共轭梯度算法 |
| CGP | traincgp | Polak-Ribiere 共轭梯度算法 |
| OSS | trainoss | 一步正割算法 |

BLF: 网络学习函数, 包括 BP 学习规则 **learnngd**, 带动量项的 BP 学习规则 **learnngdm**。

PF: 性能分析函数, 包括均值绝对误差性能分析函数 **mae**, 均方差性能分析函数 **mse**。

IPF: 输入处理函数。

OPF: 输出处理函数。

DDF: 验证数据划分函数。

一般在使用过程中设置前 6 个参数, 后面 4 个参数采用系统默认参数。

1.2 Train: BP 神经网络训练函数

函数功能: 用训练数据训练 BP 神经网络。

函数形式: **[net, tr]=train(NET, X, T, Pi, Ai)**

NET: 待训练网络。

X: 输入数据矩阵。

T: 输入数据矩阵。

Pi: 初始化输入层条件。

Ai: 初始化输出层条件。

net: 训练好的网络。

tr: 训练过程记录。

一般在使用过程中设置前 3 个参数, 后面 2 个参数采用系统默认参数。

1.3 sim: BP 神经网络预测函数

函数功能: 用训练好的 BP 神经网络预测函数的输出。

函数形式: **y=sim(net,x)**

net: 训练好的网络。

x: 输入数据。

y: 网络预测数据。

2 模型的建立

2.1 数据处理与挖掘

据题给附件数据，用 Excel 对其简单处理，发现有部分数据异常。为了建模结果的正确性，我们在运用数据之前做了大量的关于不合理数据剔除的工作，下面通过几个散点图说明。

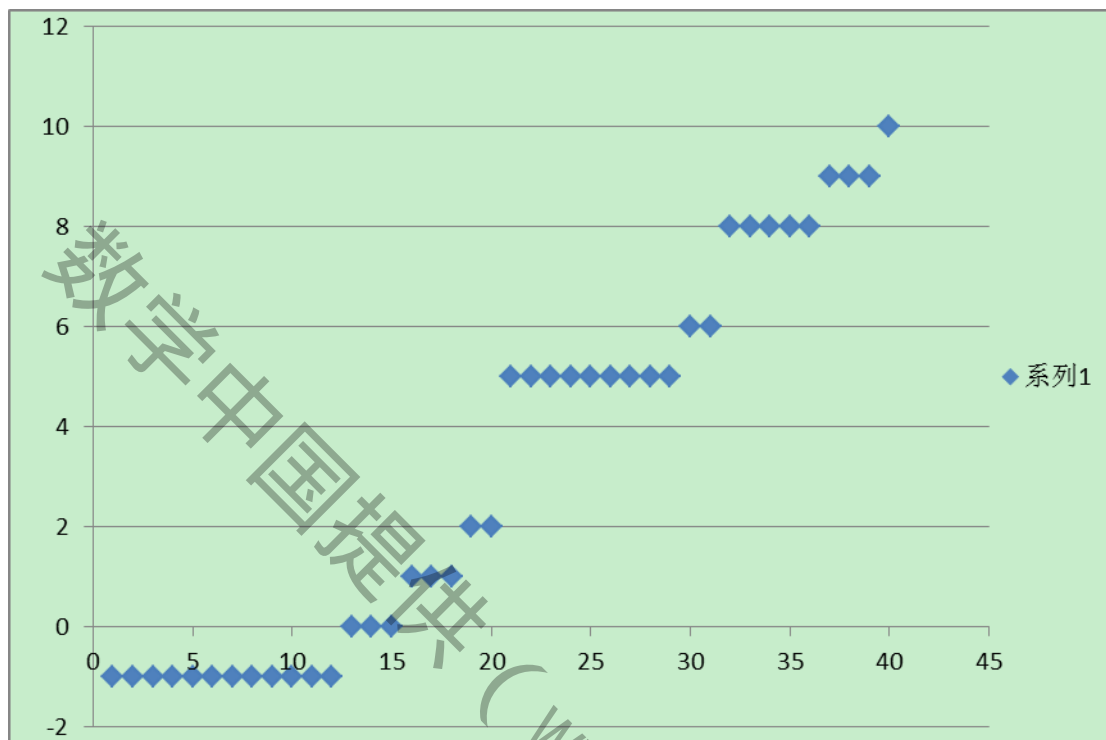


图 1.3

图 1.3 以保险期限为纵坐标，用保险期限在 10 天以下的 40 个数据呈现了具体情况。通过制作的散点图我们发现有的保险期限在 0 天以下，联系实际没有意义，故将其剔除。

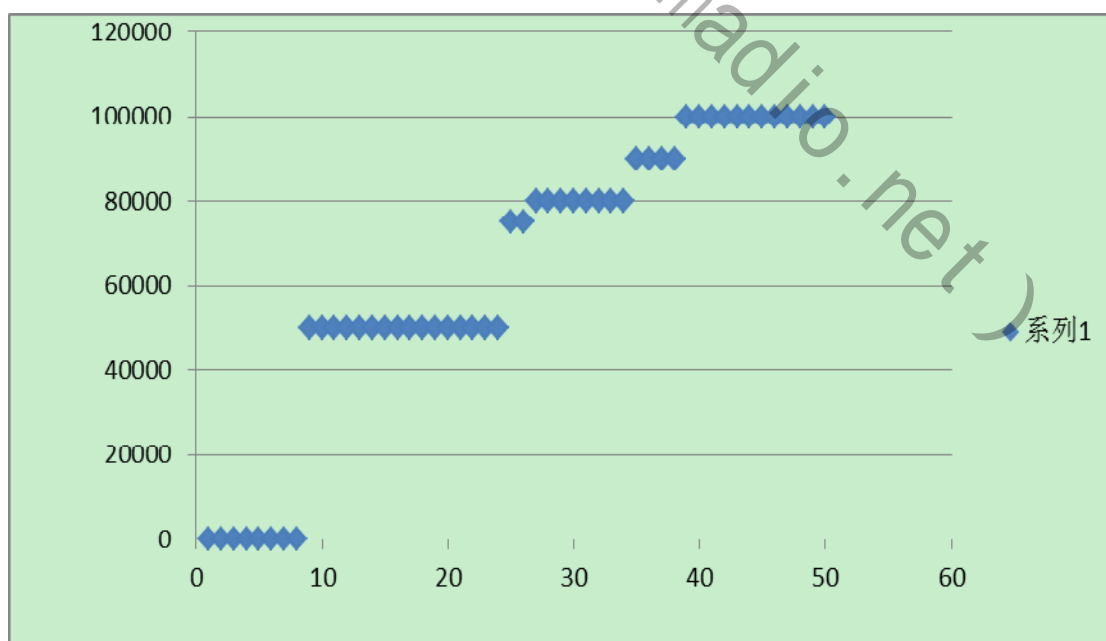


图 1.4

图 1.4 以保额为纵坐标，用保额在 100000 元以下（包括 100000 元）的 50 个数据呈

是保险公司保费浮动方案应用的主体。

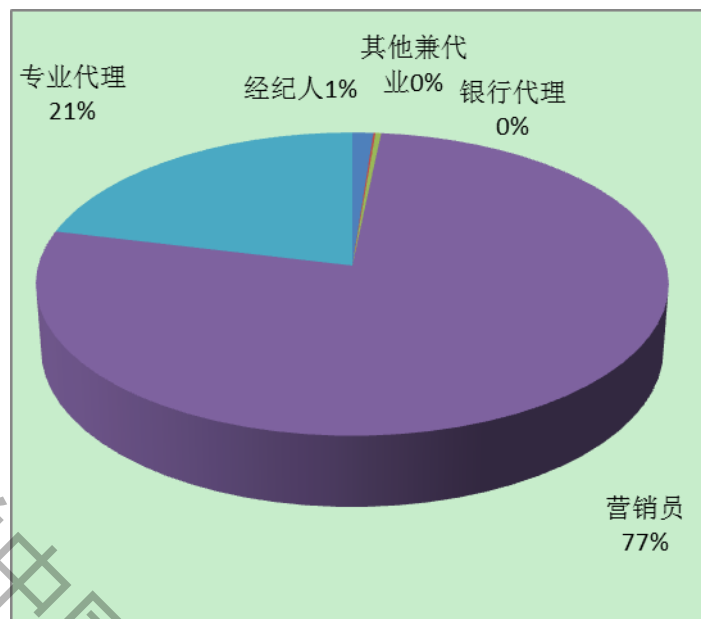


图 1.7

由图 1.7 可以得出各种展业方式比例，营销员所占比例最大，为 77%，传统的投保方式占了大部分。

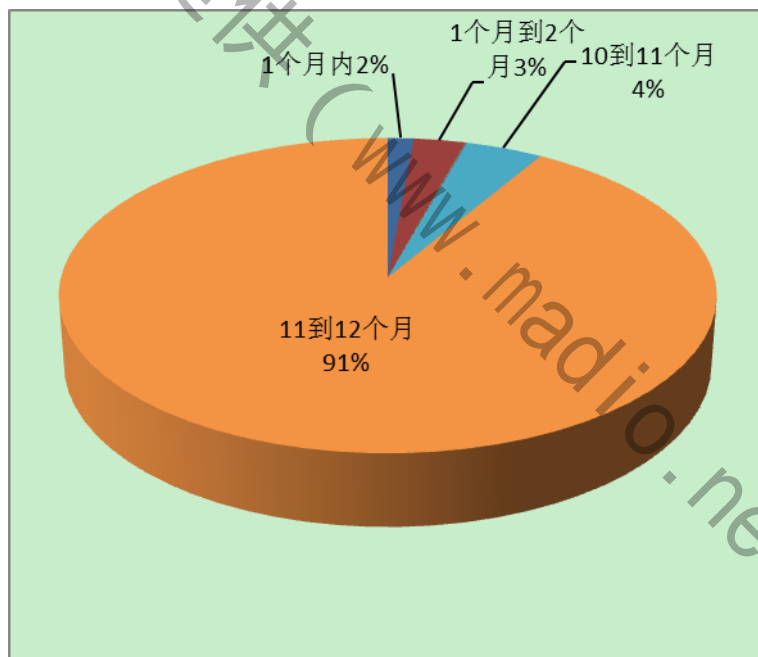


图 1.8

图 1.8 中可以看出长期保险占 91%，长期投保的金额大，优惠多，是保险公司保持客户，持续业务的增长点。

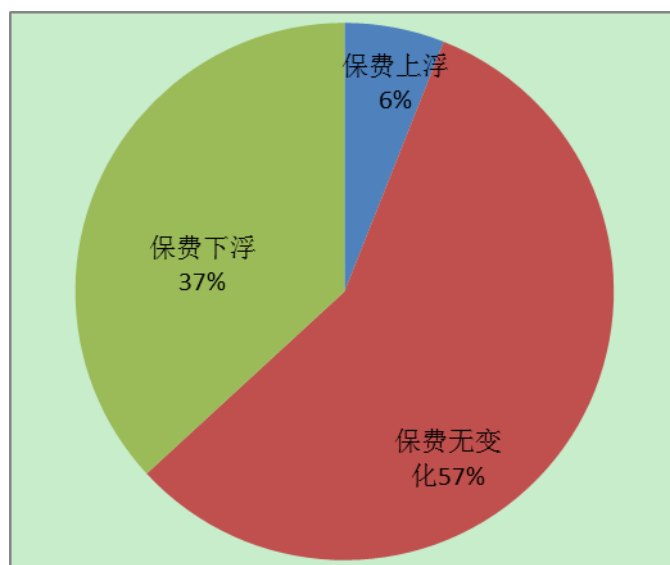


图 1.9

从上图可以看出保费下浮比例 37%，超过 $1/3$ 的投保人受益，并且保费上浮仅占 6%，保险公司既不丢失客户，又一定程度上通过保费浮动控制高风险投保。

2.2 模型的建立

经过数据处理与挖掘，得出有用数据，综合有保险期限、险种大类、车辆用途、使用性质、车辆类型、展业方式、保额、出险次数、汽车品牌共 9 种对保费浮动影响。拥有大量的数据，考虑到程序处理文字的不方便，将文字“数据”转换为对应的数字信息。对大量信息可以运用神经网络处理大规模数据的优越性，人工神经网络具有初步的自适应与自组织能力。在学习或训练过程中改变突触权重值，以适应周围环境的要求。同一网络因学习方式及内容不同可具有不同的功能。人工神经网络是一个具有学习能力的系统，可以发展知识，以致超过设计者原有的知识水平。通常，它的学习训练方式可分为两种，一种是有监督或称有导师的学习，这时利用给定的样本标准进行分类或模仿；另一种是无监督学习或称无导师学习，这时，只规定学习方式或某些规则，则具体的学习内容随系统所处环境（即输入信号情况）而异，系统可以自动发现环境特征和规律性，具有更近似人脑的功能。

基于目前应用最广 BP 神经网络，即误差反传误差反向传播算法的学习过程，由信息的正向传播和误差的反向传播两个过程组成。输入层各神经元负责接收来自外界的输入信息，并传递给中间层各神经元；中间层是内部信息处理层，负责信息变换，根据信息变化能力的需求，中间层可以设计为单隐层或者多隐层结构；最后一个隐层传递到输出层各神经元的的信息，经进一步处理后，完成一次学习的正向传播处理过程，由输出层向外界输出信息处理结果。当实际输出与期望输出不符时，进入误差的反向传播阶段。误差通过输出层，按误差梯度下降的方式修正各层权值，向隐层、输入层逐层反传。周而复始的信息正向传播和误差反向传播过程，是各层权值不断调整的过程，也是神经网络学习训练的过程，此过程一直进行到网络输出的误差减少到可以接受的程度，或者预先设定的学习次数为止。

为便于理解，下图为 BP 神经网络算法流程图。

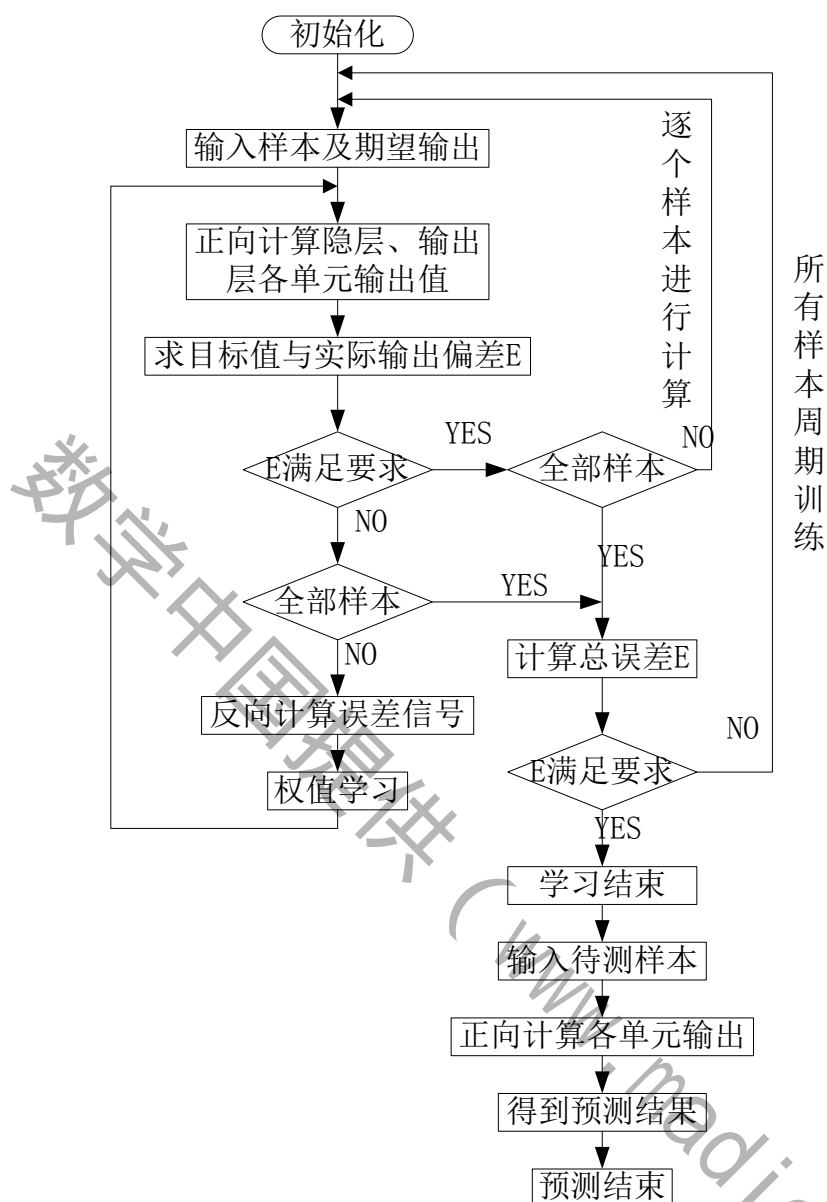


图 1.10 BP 神经网络算法流程图

使用 MATLAB 自带的神经网络工具箱，用 `newff` 建立一个 `net` 的 BP 神经网络，取用处理后数据的 3/4 作为网络训练值。

```
net=newff(P',T',[3,1],{'tansig','purelin'},'traingd')
```

 (1.1)

其中 P 为输入的 9 种影响保费浮动因素数据矩阵， T 为各不同影响因素对应的浮动比率。

然后设置网络参数迭代次数，学习率，目标分别为：

```
net.trainParam.epochs=600;
net.trainParam.lr=0.1;
net.trainParam.goal=0.00004;
```

根据建立的 BP 神经网络，对其进行网络训练。

```
net=train(net,P',T')
```

 (1.2)

训练好的 *net* 神经网络，对其做必要的误差检验，用剩下的 $1/4$ 数据进行检验。

$$y = \text{sim}(\text{net}, x) \quad (1.3)$$

考虑到数据过多，为直观显示结果，以横轴为样本，纵轴为输出函数输出、误差，做出真实值与拟合值对比曲线，误差曲线如下图所示。

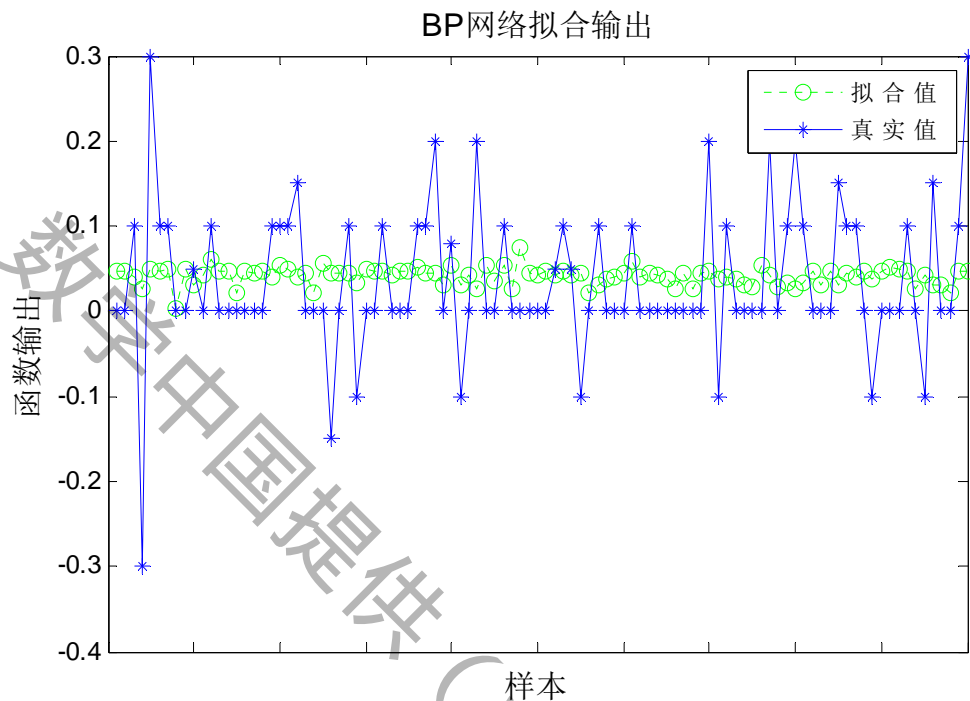


图 1.11

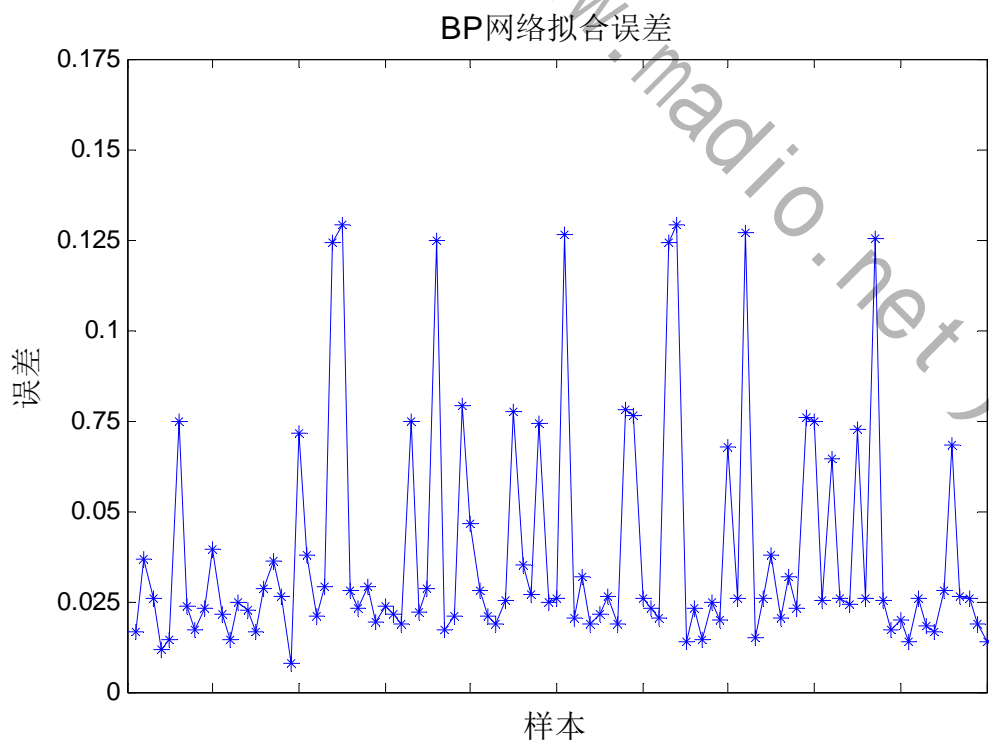


图 1.12

从图 1.12 可以直观的看出所建立的神经网络的误差大部分在 0.05 以下，局部误差能达 0.125，即 5% 的误差，因此在所给数据基础上建立的网络可行。

结合第一阶段结论出现次数对续保率有很大影响，参考《2011 年度北京地区机动车商业保险费率浮动档次升降方案》将保费浮动体系分为四大块，将每一细类保户参保指标代入(1.4)计算，得到保费浮动方案如下表：

表 1.1 保费浮动方案

| 出险次数 | 浮动前保费 (元) | 保额(万 元) | 折扣 | 出险次数 | 浮动前保费 (元) | 保额(万 元) | 折扣 |
|------|--------------|------------|---------|-----------|--------------|------------|---------|
| 0 | 0~1000 | 10~15 | 90.00% | 1 | 0~1000 | 10~15 | 94.97% |
| | | 15 以上 | 100.00% | | | 15 以上 | 118.60% |
| | 1000~2000 | 10~15 | 94.70% | | 1000~2000 | 10~15 | 95.06% |
| | | 15 以上 | 104.70% | | | 15 以上 | 103.82% |
| | 2000~30000 | 10~15 | 94.68% | | 2000~30000 | 10~15 | 95.33% |
| | | 15 以上 | 96.86% | | | 10~15 | 95.33% |
| | 3000 以上 | 10~15 | 92.07% | | 3000 以上 | 10~15 | 99.39% |
| | | 15 以上 | 96.50% | | | 10~15 | 99.39% |
| 出险次数 | 浮动前保费 (元) | 保额(万 元) | 折扣 | 出险次数 | 浮动前保费 (元) | 保额(万 元) | 折扣 |
| 2 | 0~1000 | 10~15 | 97.65% | 3 次以 上 | 0~1000 | 10~15 | 97.80% |
| | | 15 以上 | 134.00% | | | 10~15 | 97.80% |
| | 1000~2000 | 10~15 | 95.26% | | 1000~2000 | 10~15 | 92.00% |
| | | 15 以上 | 100.00% | | | 10~15 | 92.00% |
| | 2000~30000 | 10~15 | 99.05% | | 2000~30000 | 10~15 | 100.00% |
| | | 15 以上 | 99.26% | | | 15 以上 | 98.70% |
| | 3000 以上 | 10~15 | 95.26% | | 3000 以上 | 10~15 | 98.84% |
| | | 15 以上 | 97.85% | | | 15 以上 | 100.60% |

从上表可以看出，大部分投保者受益，这对降低车辆出险率、鼓励保户续保、发展潜在保户有着很大的拉动作用。

问题二

1 相关理论基础

1.1 定量分析的风险评估方法

首先了解一下风险评估步骤，如下图。

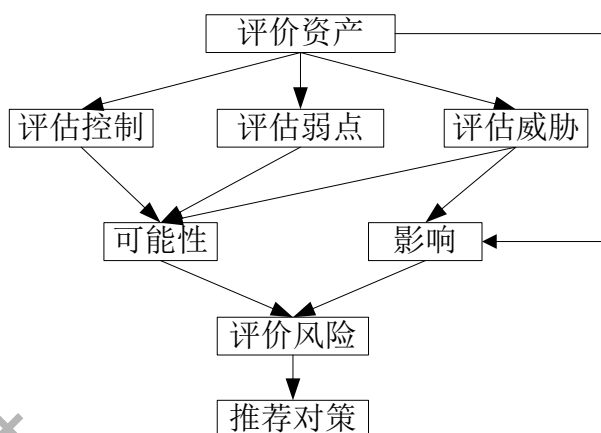


图 2.1

定量分析方法的思想很明确：对构成风险的各个要素和潜在损失的水平赋予数值或货币金额，当度量风险的所有要素（资产价值、威胁频率、弱点利用程度、安全措施的效率 and 成本等）都被赋值，风险评估的整个过程和结果就都可以被量化了。

简单说，定量分析就是试图从数字上对安全风险进行分析评估的一种方法。

定量风险分析中有几个重要的概念：

暴露因子（Exposure Factor, EF）——特定威胁对特定资产造成损失的百分比，或者说损失的程度。

单一损失期望（Single Loss Expectancy, SLE）——或者称作 SOC（Single Occurance Costs），即特定威胁可能造成的潜在损失总量。

年度发生率（Annualized Rate of Occurrence, ARO）——即威胁在一年内估计会发生的频率。

年度损失期望（Annualized Loss Expectancy, ALE）——或者称作 EAC（Estimated Annual Cost），表示特定资产在一年内遭受损失的预期值。

考察定量分析的过程，从中就能看到这几个概念之间的关系：

- （1）首先，识别资产并为资产赋值；
- （2）通过威胁和弱点评估，评价特定威胁作用于特定资产所造成的影响，即 EF （取值在 0%~100% 之间）；
- （3）计算特定威胁发生的频率，即 ARO ；
- （4）计算资产的 SLE ：

$$SLE = \text{Asset Value} \times EF$$

- （5）计算资产的 ALE ：

$$ALE = SLE \times ARO$$

1.2 灰色预测

经过讨论、实验，本文最终采用了灰色系统理论作为建立数学模型的主要方法。有关灰色系统理论的一些概念定义如下，这些定义将在之后的建模过程中被使用。

定义 1 设有序列

$$x = (x(1), x(2), \dots, x(k)) \quad (2.1)$$

则称以下映射为序列 x 到序列 y 的数据变换。

1) 当

$$f(x(k)) = \frac{x(k)}{x(1)} = y(k), x(1) \neq 0 \quad (2.2)$$

称 f 是初值化变换。

2) 当

$$f(x(k)) = \frac{x(k)}{\bar{x}} = y(k), \bar{x} = \frac{1}{n} \sum_{k=1}^n x(k) \quad (2.3)$$

称 f 是均值化变换。

3) 当

$$f(x(k)) = \frac{x(k)}{\max_k x(k)} = y(k) \quad (2.4)$$

称 f 是百分比变换。

4) 当

$$f(x(k)) = \frac{x(k)}{\min_k x(k)} = y(k), \min_k x(k) \neq 0 \quad (2.5)$$

称 f 是倍数变换。

5) 当

$$f(x(k)) = \frac{x(k)}{x_0} = y(k) \quad (2.6)$$

其中 x_0 为大于零的某个值，称 f 是归一化变换。

6) 当

$$f(x(k)) = \frac{x(k) - \min_k x(k)}{\max_k x(k)} = y(k) \quad (2.7)$$

称 f 是极差最大化变换。

7) 当

$$f(x(k)) = \frac{x(k) - \min_k x(k)}{\max_k x(k) - \min_k x(k)} = y(k) \quad (2.8)$$

称 f 是区间值化变换。

定义 2 把数列 x 各时刻数据依次累加的过程叫做累加过程，记作 **AGO**，累加所得的新数列，叫做累加生成数列。具体地，设原始数列为

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

累加生成数列记为

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

且 $x^{(0)}$ 与 $x^{(1)}$ 满足

$$x^{(1)}(k) = \sum_{i=a}^k x^{(0)}(i), k = a, \dots, n \quad (2.9)$$

其中 $a \leq n$ 为正整数。上述累加过程当 $1 < a < k$ 时称为去首累加生成，当 $a=1$ 时成为一般累加生成。

这里，我们只讨论 $a=1$ 时的情形，(2.9) 式中上标 (1) 表示 1 次累加生成，记作 1-AGO。在一次累加数列 $x^{(1)}$ 的基础上再做 1 次累加生成，可得到 2 次累加生成，记作 2-AGO。依次下去，对原始数列 $x^{(0)}$ 我们可以做 r 次累加生成，记作 r -AGO，从而得到 r 次累加生成数列 $x^{(r)}$ 。 $x^{(r)}$ 与 $x^{(r-1)}$ 满足下面的关系：

$$x^{(r)}(k) = \sum_{i=1}^k x^{(r-1)}(i), k = 1, 2, \dots, n \quad (2.10)$$

在实际应用中，最常用的是 1 次累加生成。本文只讨论 1 次累加生成。

2 模型的建立

表 2.1 所示是从参考数据中采集的反应该保险公司 2010-1-1 到 2010-3-31 影响风险的指标。本文将以 2010-1-1 到 2010-3-31 间的数据为基础建立 GM(1,1) 灰色模型来估计 2010 年 4 月的发展状况。通过定量分析模型对该公司进行风险评估。

表 2.1 2010-1-1 到 2010-3-31 该保险公司各项风险影响指标

| 时间 | 保费收入(元) | 财务赔款(元) | 理赔费用(元) | 已决件数 | 未决件数 | 拒赔件数 | 注销件数 | 出险次数 |
|------|---------|----------|----------|------|------|------|------|------|
| 1月上旬 | 3677598 | 1244684 | 59085 | 415 | 18 | 5 | 182 | 620 |
| 1月下旬 | 6886301 | 2651856 | 121138.5 | 836 | 54 | 11 | 305 | 1207 |
| 2月上旬 | 2629976 | 961068.6 | 47225 | 332 | 28 | 0 | 116 | 476 |
| 2月下旬 | 2353398 | 453493.6 | 23650 | 183 | 8 | 0 | 71 | 262 |
| 3月上旬 | 4643439 | 1624417 | 68404.1 | 388 | 30 | 2 | 208 | 630 |
| 3月下旬 | 5412565 | 1219906 | 65603 | 443 | 57 | 1 | 242 | 744 |

由于表 2.1 中的八项数据具有不同的量纲，而且在计算关联系数时要求量纲相同。因此首先对各项数据进行无量纲化。同时，为了易于比较，将各项数据变换到相同的数量级。整个数据变换过程利用式(2.11)均值化处理。

得到表 2.2 所示的整理结果：

表 2.2 处理后无量纲数据

| 时间 | 理赔费用(元) | 财务赔款(元) | 保费收入(元) | 已决件数 | 未决件数 | 拒赔件数 | 注销件数 | 出险次数 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|
| 1月上 | 0.92055 | 0.91572 | 0.86183 | 0.9588 | 0.55385 | 1.57895 | 0.97153 | 0.9444 |
| 1月下 | 1.88736 | 1.95099 | 1.61377 | 1.93146 | 1.66154 | 3.47368 | 1.62811 | 1.83854 |
| 2月上 | 0.73577 | 0.70706 | 0.61632 | 0.76704 | 0.86154 | 0 | 0.61922 | 0.72506 |
| 2月下 | 0.36847 | 0.33364 | 0.55151 | 0.4228 | 0.24615 | 0 | 0.379 | 0.39909 |
| 3月上 | 1.06575 | 1.19509 | 1.08817 | 0.89642 | 0.92308 | 0.63158 | 1.11032 | 0.95963 |
| 3月下 | 1.0221 | 0.89749 | 1.26841 | 1.02349 | 1.75385 | 0.31579 | 1.29181 | 1.13328 |

为了直观的呈现结果，以时间为横轴，做出八项指标对应的曲线，如图 2.2 所示：

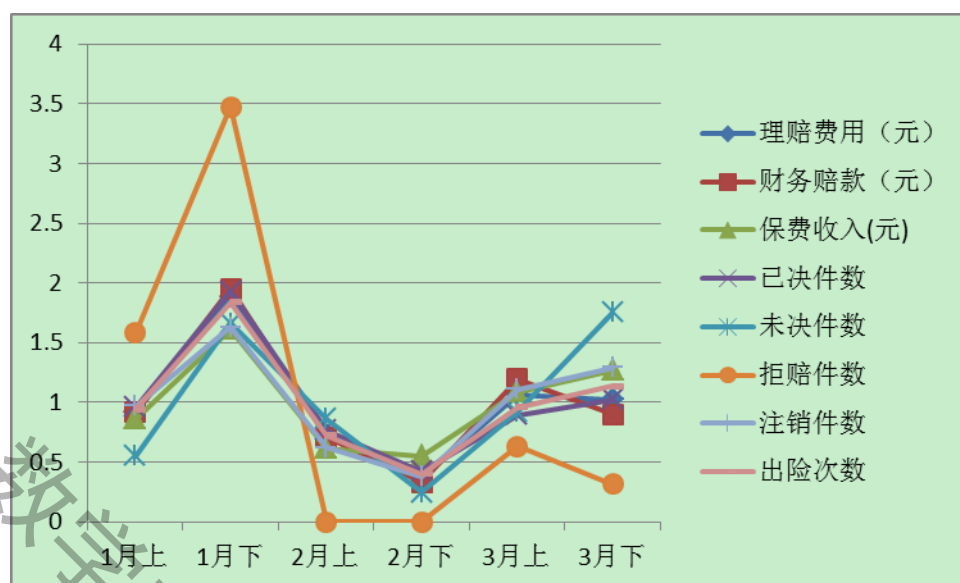


图 2.2

从图 2.2 可以直观的看出风险指标的变化。

(1) 根据初始化后的数据，选取理赔费用为参考数列，首先进行级比检验。建立理赔费用的时间序列如下：

$$x_0 = \{x_0(k)\} = (x_0(1), x_0(2), x_0(3), \dots, x_0(n)) \quad (2.12)$$

其中 k 表示时间，另外 7 项作为比较数列

$$x_i = \{x_i(k)\} = (x_i(1), x_i(2), x_i(3), \dots, x_i(n)) \quad i = 1, 2, 3, 4, 5, 6, 7 \quad (2.13)$$

对于 1 月到 3 月每项指标都有相关数据，经数据处理得到矩阵 $A = (a_{ij})_{6 \times 8}$ ，计算每年的平均值，记为：

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(8)) \quad (2.14)$$

对 $x^{(0)}$ 求级比：

$$\lambda(i) = x^{(0)}(i-1) / x^{(0)}(i) \in (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$$

此题中 $n=6$ ，所以区间范围为：(0.8007, 1.1994)

所得出的级比都在可容覆盖区间范围 (0.8007, 1.1994) 内，故可用 x_0

做 GM (1,1) 灰色预测模型。

(2) 定义：

$$r(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_{0i}(k) + \rho \Delta_{\max}} \quad (2.15)$$

为灰关联系数，其中 $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$ 为绝对差， $\Delta_{\min} = \min_i \min_k \Delta_{0i}(k)$ 为两级最小差，

$\Delta_{\max} = \max_i \max_k \Delta_{0i}(k)$ 为两级最大差。 $\rho \in (0, 1)$ 为分辨系数。

由表 2.2 求绝对差，得数列：

$$\Delta_{01} = (0.00483, 0.06363, 0.02871, 0.03483, 0.12935, 0.12461)$$

$$\Delta_{02} = (0.05873, 0.27358, 0.11945, 0.18304, 0.02242, 0.24630)$$

$$\Delta_{03} = (0.03825, 0.04410, 0.03127, 0.05433, 0.16933, 0.00138)$$

$$\Delta_{04} = (0.36671, 0.22582, 0.12577, 0.12232, 0.14267, 0.73174)$$

$$\Delta_{05} = (0.65839, 1.58633, 0.73577, 0.36847, 0.43417, 0.70631)$$

$$\Delta_{06} = (0.05098, 0.25924, 0.11656, 0.01053, 0.04457, 0.26971)$$

$$\Delta_{07} = (0.02385, 0.04882, 0.01072, 0.03062, 0.10611, 0.11118)$$

比较上述绝对差序列有：

$$\Delta_{\min} = 0.00138, \Delta_{\max} = 1.58633$$

(3) 计算关联系数如下：

$$\text{取 } \rho = 0.5 \quad \zeta_{0j(k)} = \frac{0.00138 + 0.5 \times 1.58633}{\Delta_{0i} + 0.5 \times 1.58633}$$

$$\zeta_{01} = (0.9957, 0.9273, 0.9667, 0.9596, 0.8613, 0.8657)$$

$$\zeta_{02} = (0.9327, 0.7448, 0.8706, 0.8139, 0.9742, 0.7644)$$

$$\zeta_{03} = (0.9557, 0.9490, 0.9637, 0.9375, 0.8255, 1.0000)$$

$$\zeta_{04} = (0.6850, 0.7797, 0.8646, 0.8679, 0.8490, 0.5210)$$

$$\zeta_{05} = (0.5474, 0.3339, 0.5197, 0.6840, 0.6474, 0.5299)$$

$$\zeta_{06} = (0.9412, 0.7550, 0.8734, 0.9886, 0.9484, 0.7475)$$

$$\zeta_{07} = (0.9725, 0.9437, 0.9884, 0.9645, 0.8835, 0.8786)$$

设 ω_k 为指标 k 的权重，满足 $0 \leq \omega_k \leq 1$ ，定义

$$r(x_0, x_i) = \sum_{k=1}^n \omega_k r(x_0(k), x_i(k)) \quad (2.16)$$

为 x_0 对 x_i 的灰色关联度。取 $\omega_1 = \omega_2, \dots, \omega_6 = \frac{1}{6}$ ，算出各列的关联度如下表所示：

表 2.3 各项指标与理赔费用的关联度

| 财务赔款 (元) | 保费收入 (元) | 已决件数 | 未决件数 | 拒赔件数 | 注销件数 | 出险次数 |
|-------------|-------------|---------|--------|---------|---------|---------|
| 0.92938 | 0.8501 | 0.93857 | 0.7612 | 0.54372 | 0.87568 | 0.93853 |

从结果可以看出，直接因素（前 3 个）关联度的排序为 $r_{03} > r_{07} > r_{01}$ 说明在理赔费用的影响中，已决件数是主要因素。在间接因素（后 4 个）中，关联度的排序为 $r_6 > r_2 > r_4 > r_5$ 注销件数是主要的间接影响因素。

构造数据矩阵 B 及数据向量 Y

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ -\frac{1}{2}(x^{(1)}(3) + x^{(1)}(4)) & 1 \\ -\frac{1}{2}(x^{(1)}(4) + x^{(1)}(5)) & 1 \\ -\frac{1}{2}(x^{(1)}(5) + x^{(1)}(6)) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(6) \end{bmatrix}$$

使用 Matlab 计算 \hat{u} ：

$$\hat{u} = (a, b)^T = (B^T B)^{-1} B^T Y = \begin{pmatrix} 0.0248 \\ 1.0148 \end{pmatrix} \quad (2.17)$$

于是得到相应的灰色预测模型为：

$$\frac{dx^{(1)}}{dt} + 0.0248x^{(1)} = 1.0148 \quad (2.18)$$

使用 MATLAB 建立灰色模型，并计算出 4 月上旬和 4 月下旬的相应指标 GM(1,1) 灰色模型预测值。见下表：

表 2.4 灰色模型计算结果

| 时间 | 1 月上旬 | 1 月下旬 | 2 月上旬 | 2 月下旬 | 3 月上旬 | 3 月下旬 | 4 月上旬 | 4 月下旬 |
|---------|----------|----------|----------|----------|----------|----------|----------|----------|
| 保费收入(元) | 3677598 | 6886301 | 2629976 | 2353398 | 4643439 | 5412565 | 6322749 | 6213514 |
| 财务赔款(元) | 1244684 | 2651856 | 961068.6 | 453493.6 | 1624417 | 1219906 | 1858710 | 2236166 |
| 理赔费用(元) | 59085 | 121138.5 | 47225 | 23650 | 68404.1 | 65603 | 76294.6 | 675405.7 |
| 已决件数 | 415 | 836 | 332 | 183 | 388 | 443 | 497 | 593 |
| 未决件数 | 18 | 54 | 28 | 8 | 30 | 57 | 80 | 63 |
| 拒赔件数 | 5 | 11 | 0 | 0 | 2 | 1 | 2 | 4 |
| 注销件数 | 182 | 305 | 116 | 71 | 208 | 242 | 190 | 292 |
| 出险次数 | 620 | 1207 | 476 | 262 | 630 | 744 | 787 | 951 |
| 拒赔率 | 0.011416 | 0.012195 | 0 | 0 | 0.004739 | 0.001992 | 0.00335 | 0.00607 |
| 未决率 | 0.041096 | 0.059867 | 0.077778 | 0.041885 | 0.07109 | 0.113546 | 0.134003 | 0.095599 |

为了直观，将 1 月上旬到四月下旬的拒赔率、未决率在坐标系中画出，如图 2.3 所示：

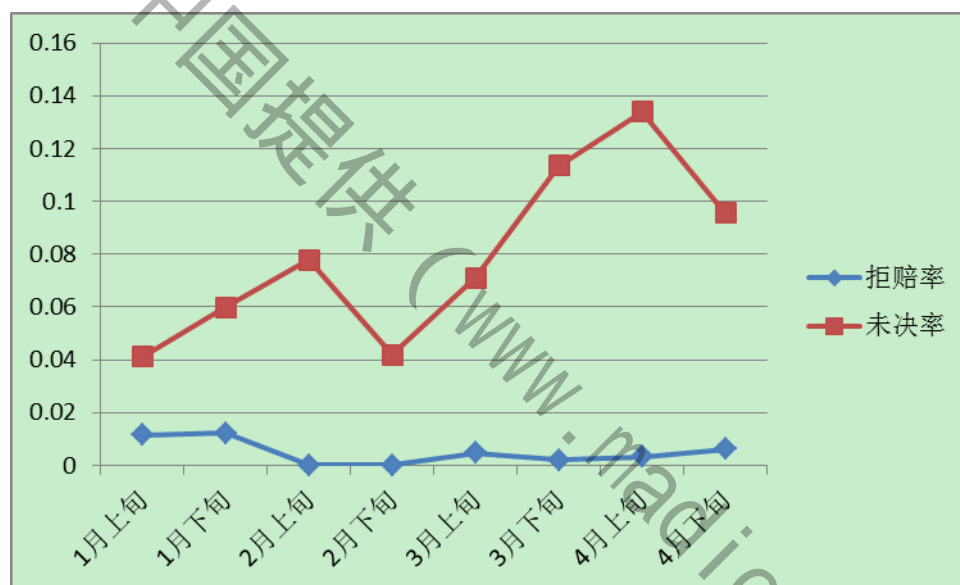


图 2.3 据赔率及未决率

从上图可以看出该的未决率上升趋势，即故意拖延不解决的保单积压过多，很可能造成投保人对该公司的不满，使其信誉度下降。

根据定量分析的风险评估模型对该公司信誉风险指数进行计算：

$$CAR = UNS + DCS \quad (2.19)$$

同时，再对该公司资金风险指数进行计算：

$$FAR = \frac{INF \times ACR - EXP}{INF} \quad (2.20)$$

得出结果如下表所示：

表 2.5 该公司信誉风险指数及资金风险指数

| 时间 | 1 月上旬 | 1 月下旬 | 2 月上旬 | 2 月下旬 | 3 月上旬 | 3 月下旬 | 4 月上旬 | 4 月下旬 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| 信誉风险指数 | 0.053 | 0.072 | 0.078 | 0.042 | 0.076 | 0.116 | 0.137 | 0.102 |
| 资金风险指数 | 0.494 | 0.694 | 0.312 | 0.320 | 0.388 | 0.364 | 0.364 | 0.393 |

为直观分析该公司的风险指数，以时间为横轴，做出其变化曲线如下图：



图 2.4

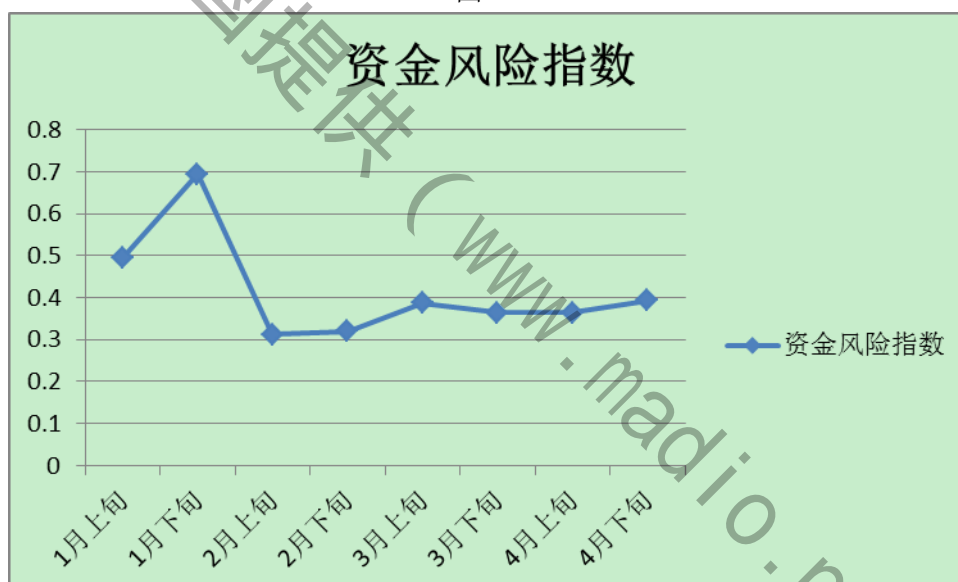


图 2.5

从图中可以看出该公司的信誉风险逐渐升高，并且只有二月下旬小于 0.05，从调查资料得知，国际上大多银行将其信誉指标定在 0.95 及其以上，故取 0.95，那么该公司只有二月下旬达到标准。同时，对该公司资金风险指数来看，在其允许的赔偿范围内，资金上没有多大的风险。

根据上述结论，建议该保险公司下阶段内，提高办事效率，减少拖延处理的保单，提高公司信誉度，塑造良好形象。

六、模型的评价与推广

BP 神经网络实质上实现了一个从输入到输出的映射功能，而数学理论已证明它具有实现任何复杂非线性映射的功能。这使得它特别适合于求解内部机制复杂的问题；另外 BP 神经网络能通过学习带正确答案的实例集自动提取“合理的”求解规则，即具有自学习能力；其次 BP 神经网络具有一定的推广、概括能力使它成为目前应用最多的一种神经网络形式。

在工程应用领域中，应用 BP 网络的好坏最关键的是输入特征选择和训练样本集的准备，若样本集代表性差，矛盾样本多，数据归一化存在问题，那么，使用多复杂的综合算法、多精致的网络结构，建立起来的模型预测效果都不会多好，若想取得实际有价值的效果，还是要从最基础的数据整理工作做起。

灰色系统理论是 1982 年由中国学者邓聚龙提出的，目前许多国家和组织的知名学者都参与到灰色系统理论与应用的研究当中。灰色系统理论由于其适用性，已经广泛应用到工业、农业、社会、经济、能源、交通、气象、水利等研究中，发挥了巨大的作用。

灰色模型是从灰色系统中抽象出来的模型。本文以大量的资料、数据查询为基础，预测了评估的综合和细化指标，在保险公司风险评估中，使得模型评估简单有效。

参考文献

- [1]MATLAB 中文论坛，MATLAB 神经网络 30 个案例分析，北京航空航天大学出版社，2010 年 4 月
- [2]朱凯，王正林，精通 MATLAB 神经网络，北京，电子工业出版社，2010 年 1 月
- [3] 刘思峰，党耀国，方志耕等，灰色系统理论及其应用，北京:科学出版社，2004 年 11 月；
- [4] 傅洋，北京晚报，2010 年 12 月 21 日 星期二
http://bjwb.bjd.com.cn/html/2010-12/21/content_350114.htm

附录

BP 神经网络程序：

```
function BXP(X)
[m,n]=size(X);
k=rand(1,m);
[m1,n1]=sort(k);
%取 3/4 的数据做训练，其余做检验，round 是朝最近的方向取整
n2=round(0.75*m);
input_train=X(n1(1:n2),1:(n-1));
output_train=X(n1(1:n2),n);
input_test=X(n1((n2+1):end),1:(n-1));
output_test=X(n1((n2+1):end),n);
[inputn,inputps]=mapminmax(input_train);
[outputn,outputps]=mapminmax(output_train);
%BP 神经网络构建
net=newff(inputn,outputn,[3,1],{'tansig','purelin'},'traingd');
%网络参数配置（迭代次数，学习率，目标）
net.trainParam.epochs=600;
net.trainParam.lr=0.1;
net.trainParam.goal=0.00004;
%BP 神经网络训练
net=train(net,inputn,outputn);
%拟合
an=sim(net,input_test);
%拟合结果图形
figure(1)
plot(an(1:100),':og');
hold on
h=output_test;
plot(h(1:100),'-*');
legend('拟合值','真实值');
title('BP 网络拟合输出','fontsize',12)
ylabel('函数输出','fontsize',12)
xlabel('样本','fontsize',12)
%网络拟合误差图形
error=abs((output_test)-an);
figure(2)
plot(error(1:100),'-*')
title('BP 网络拟合误差','fontsize',12)
ylabel('误差','fontsize',12)
xlabel('样本','fontsize',12)
```

灰色模型程序：

%通过以前数据预测 2010 年数据

function gray_model()

b=xlsread('预测表格.xlsx');%根据 2010 年各月份数据做分析

b=cumsum(b,2);

[ro,wo]=size(b);

n=wo;

for rw=1:ro

 r(1)=1;

 r(2:n)=b(rw,1:n-1)/b(rw,2:n);

 rm(rw,1:n)=r;

 c=cumsum(b(rw,:));

 for i=1:n-1

 zm(rw,i+1)=0.5*c(i)+0.5*c(i+1);

 B(i,1)=-0.5*c(i)-0.5*c(i+1);

 B(i,2)=1;

 end

 zm1=B(:,1);

 zm(rw,1)=0;

 Y(1:n-1)=b(rw,2:n);

 U=inv(B'*B)*B'*Y';

 for k=0:n+2

 c(k+1)=(b(1)-U(2)/U(1))*exp(-U(1)*k)+U(2)/U(1);

 end

 m(rw,1)=b(rw,1);

 for k=2:n+2

 m(rw,k)=c(k)-c(k-1);

 end

end

B

m

U