

基于主成分分析及BP神经网络分析的地震人员伤亡预测模型研究

田鑫, 朱冉冉

(上海理工大学管理学院, 上海 200093)

摘要:影响地震人员伤亡的因素众多,且彼此之间存在着复杂的交互作用。本文在综合考虑各种因素后选取地震发生时刻、人口密度、地震预报与否、建筑物破坏率、设防水准、震级、烈度等因素作为评价指标,首先运用主成分分析算出其主成分,然后运用神经网络分析方法建立预测模型。实例验证预测效果良好。

关键词:地震灾害;人员伤亡;主成分分析;BP神经网络分析

中图分类号: P315.941

文献标识码: A

文章编号: 1000-0844(2012)04-0365-04

DOI:10.3969/j.issn.1000-0844.2012.04.0365

Study on Earthquake Casualty Forecasting Model Based on Principal Component Analysis and BP Neural Network Analysis

TIAN Xin, ZHU Ran-ran

(School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: There are many factors that influence the casualty in a earthquake, and the factors always relate each other. In this paper, some factors are selected as the evaluation indicators, such as time of earthquake, population density, earthquake forecast, rate of building damage, resistance level, earthquake magnitude, earthquake intensity. First using the method of principal component analysis to calculate the principal components, then using the method of BP neural network analysis to construct the prediction model. Test for true cases prove the model has good prediction result.

Key words: Earthquake disaster; Casualty; Principal component analysis; BP neural network analysis

0 引言

近十年以来,中国发生6.0级以上地震二十余次,造成的人员伤亡及经济损失不计其数,因此建立地震人员伤亡预测模型显得极为重要^[1]。近些年来,国内外学者也提出了一些地震人员伤亡预测模型,这些模型通常分为两类,一类是不考虑易损性方法,另一类以易损性为主要参数的方法^[2-3]。但是,由于地震人员伤亡受诸多因素影响,并且因素彼此之间存在错综复杂的关系,因此得出的预测模型往

往很难达到理想的效果。通过对收集的6.0级以上地震人员伤亡数据的分析,得出其数据符合主成分分析统计规律^[4]。本文通过综合考虑后最终选取地震发生时刻、人口密度、地震预报与否、建筑坍塌率、建筑严重破坏率、建筑一般破坏率、设防水准、震级、地震烈度等因素作为评价指标^[5-9]。首先运用主成分分析对这些因素进行降维,并算出各主成分,然后在此基础上运用神经网络方法建立预测模型,最后通过实例对预测结果进行检验。

1 主成分分析、BP神经网络模型基本原理及模型

1.1 主成分分析基本原理及模型

主成分分析是数学上数据降维的一种方法,其基本思想是设法将原来众多的具有一定相关性的指标(比如 p 项指标)重新组合成一组新的互不相关的综合指标来代替原来指标。最经典的做法就是用 F_1 选取的第一个线性组合,即第一个综合指标的方差来表达,即 $Var(F_1)$ 越大,表示 F_1 包含的信息越多。因此在所有的线性组合中选取的 F_1 应该是方差最大的,故称 F_1 为第一主成分。如果第一主成分不足以代表原来 P 个指标的信息,再考虑选取 F_2 即选第二个线性组合。为了有效地反映原来信息, F_1 已有的信息就不需要再出现在 F_2 中,用数学语言表达就是要求 $Cov(F_1, F_2) = 0$,则称 F_2 为第二主成分。依此类推可以构造出第三、第四、...,第 P 个主成分^[10]。其模型如下。

设有 n 各样本,每个样本有 p 项指标: $X_1, X_2, \dots, X_p$, 得到原始数据矩阵

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (X_1, X_2, \dots, X_p)$$

其中, $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})$, $i = 1, 2, \dots, p$ 。

用数据矩阵 X 的 p 个列向量(即 p 个指标向量) $X_1, X_2, \dots, X_p$ 作线性组合,得综合指标向量

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \vdots \\ \cdots F_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases}$$

也可以简写成: $F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{pi}X_p$, $i = 1, 2, \dots, p$ 。其中 F_i 与 F_j 互不相关($i = 1, 2, \dots, p$), 即 $Cov(F_i, F_j) = 0$, 并有 $Var(F_i) = 1$ 。

1.2 BP神经网络基本原理及模型

BP (Back Propagation)神经网络,即误差反向传播算法的学习过程,由信息的前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信息,经进一步处理后,完成一次学习的前向传播处理过程,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。

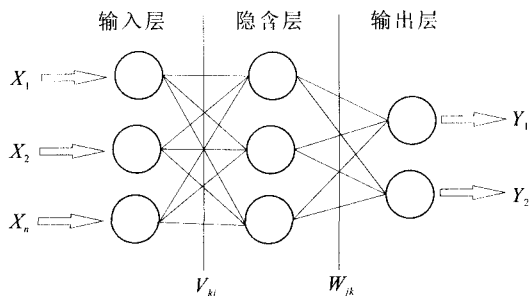


图1 BP神经网络结构示意图

Fig. 1 Sketch of BP neural network.

BP神经网络计算如下:设第 q 层($q = 1, 2, \dots, Q$)的神经元个数为 n_q ,输入到第 q 层的第 i 个神经元的连接权系数为 W_{ij}^q ($i = 1, 2, \dots, n_q; j = 1, 2, \dots, n_{q-1}$)则该多层感知器网络的输入/输出变换关系为

$$s_i^q = \sum_{j=0}^{n_{q-1}} w_{ij}^q x_j^{q-1}$$

其中, $w_{ij}^q = -1; x_0^{q-1} = \theta_i^q; x_i^{q-1} = f(s_i^q) = \frac{1}{1 + e^{-w_i^q}}$;

$i = 1, 2, \dots, n_q; j = 1, 2, \dots, n_{q-1}; q = 1, 2, \dots, Q$ 。

利用该样本集首先对BP网络进行训练,即对网络的连接权系数进行学习和调整,以使该网络实现给定的输入/输出映射关系^[11]。经过训练的BP神经网络,对于实施样本的输入也能给出合适的输出。该性质称为泛化功能。

2 样本收集与分析

2.1 样本收集

为了建模的准确性,本文选取了不同震级和设防水准的地区以及一次地震中不同地段、烈度和震级的地区作为样本,并通过查阅文献综合信息^[5-6, 12-15]最终选取近几十年国内发生的6.0级以上地震作为样本,如表1所示。

2.2 基于SPSS的主成分分析

利用SPSS统计工具,对数据进行因子分析,结果如表2所示。选取特征值大于1的作为主成分,由计算结果得出4个主成分,其累计值为82.491% > 80%,符合要求。然后利用转换关系, $a = b/\sqrt{r}$ 算

出主成分系数,进而利用 F_i 公式可以算出每个评价指标对应的 F_{ij} 值。具体数值如表 2、表 3、表 4 所示。

表 1 地震灾害人员伤亡表

序号	发生地点	发生时刻	人口密度/km ²	地震预报	建筑坍塌/%	建筑严重破坏/%	建筑一般破坏/%	设防水准	震级/M	烈度/I	伤亡人数/个
1	成都	1	890	0	0.026	0.77	21.5	Ⅶ	8	Ⅵ	31 690
2	都江堰	1	498	0	5	46	31	Ⅶ	8	Ⅸ	7 886
3	北川	1	56	0	90	10	0	Ⅶ	8	Ⅺ	18 298
4	江油	1	322	0	2.1	6.4	61.7	Ⅶ	8	Ⅷ	10 454
5	什邡	1	498	0	55	20	15	Ⅶ	8	Ⅷ	39 233
6	锦竹	1	413	0	6.3	36.6	50.8	Ⅶ	8	Ⅺ	48 543
7	玉树	1	6.6	0	100	0	0	Ⅶ	7.1	Ⅷ	15 103
8	唐山	2	401	1	95	3	2	Ⅷ	7.8	Ⅺ	406 000
9	汶川	1	26.9	0	30	50	10	Ⅶ	8	Ⅸ	50 524
10	海城	2	670	1	9.08	42.52	46.28	Ⅶ	7.3	Ⅸ	17 308
11	邢台	1	462	0	35.55	40.31	12.02	Ⅶ	7.2	Ⅹ	46 064
12	通海	1	218	0	69	21	7	Ⅶ	7.8	Ⅸ	42 404
13	松潘	2	2 563	1	5	41	32	Ⅷ	7.2	Ⅸ	797
14	丽江	2	58	1	20	3.9	30.43	Ⅷ	7	Ⅸ	17 221
15	武定	2	55	1	6.13	2.52	0.98	V	6.5	Ⅸ	13 867
16	张北	1	115	0	17.68	25.2	10.16	V	6.2	Ⅷ	11 488

注: (1)设防水准主要参考“我国主要城镇抗震设防烈度”;(2)发生时刻 1 表示白天(7:00—19:00),2 表示夜晚(19:00—7:00);(3)地震预报表示 1,没有预报为表示 0。

表 2 方差分解主成分提取表

成份	初始特征值			提取平方和载入		
	合计	方差的/%	累积/%	合计	方差的/%	累积/%
1	2.382	26.465	26.465	2.382	26.465	26.465
2	2.192	24.354	50.819	2.192	24.354	50.819
3	1.816	20.177	70.996	1.816	20.177	70.996
4	1.035	11.494	82.491	1.035	11.494	82.491
5	0.675	7.503	89.994			
6	0.430	4.781	94.775			
7	0.263	2.925	97.700			
8	0.150	1.669	99.368			
9	0.057	.632	100.000			

表 3 因子载荷矩阵表

评价指标	1	2	3	4
设防水准	0.747	0.128	0.152	-0.523
人口密度	0.668	0.16	-0.424	0.099
地震预报	0.212	0.924	-0.042	-0.081
发生时刻	0.384	0.823	0.027	0.171
震级	0.582	-0.584	0.191	-0.382
坍塌	0.151	-0.072	0.913	-0.037
一般破坏	0.486	-0.262	-0.627	-0.119
烈度	0.551	0.004	0.566	0.436
严重破坏	0.532	-0.452	-0.167	0.604

表 4 指标对应 F 值表

序号	Z1	Z2	Z3	Z4
1	389.96	218.67	71.35	-108.66
2	228.97	148.99	47.36	-29.77
3	-3.46	-10.8	58.62	-2.31
4	157.9	108.22	30.36	-39.5
5	204.02	114.2	70.89	-49.74
6	196.64	134.28	41.31	-26.85
7	144.15	64.38	85.53	-49.32
8	306.85	191.95	59.89	-54.44
9	197.06	118.65	60.49	-28.5
10	76.79	41.33	59.27	-14.18
11	1117.91	621.55	192.21	-281.27
12	30.18	24.63	23.42	-6.96
13	23.91	13.51	17.16	-2.73
14	50.81	37.47	25.36	3.57

指导隐含层的神经元个数的选择,而只是结合实际情况进行试探性选择^[7,11]。在经过综合考虑后,本模型采用隐含层的神经元个数为 4;对于传输函数采用 logsig 及 pureline 函数。在导入数据后,部分训练代码如下:

```
net.trainParam.show=50;
net.trainParam.lr=0.01;
net.trainParam.epochs=10 000;
net.trainParam.time=6 000;
net.trainParam.goal=0.000 1;
SourceDataConvert = importdata('bp_train_
sample_data.dat');
```

2.3 基于 MATLAB 的 BP 神经网络分析

借用 MATLAB 工具箱建立 BP 神经网络模型。其中模型的输入层为表 4 数据,输出层为地震人员伤亡数;模型隐含层神经元个数的选择对网络的学习和计算特性具有非常重要的影响,是该网络构成成败的关键,但到目前为止,还没有完善理论来

SourceData=SourceDataConvert

TargetConvert=importdata('bp_train_target_data.dat');

Target=TargetConvert

net=train(net,SourceData,Target)

训练完成后,利用样本 7、9 作为检验样本来验证模型的准确性,误差率用公式: $p = \frac{|\text{实际伤亡人数} - \text{预测伤亡人数}|}{\text{实际伤亡人数}}$ 来确定,结果如表 5 所示。

表 5 预测误差分析表

样本	实际伤亡人数	预测伤亡人数	误差	误差率
样本 5	15 103	18 298	3 195	0.211 547 4
样本 7	50 524	53 304	2 780	0.055 023 4

由表 5 可以看出,预测结果误差率分别为 21%、5.5%,预测效果比较理想。

3 总结

本文首先对收集的数据进行主成分分析,以此来降低评价指标维度以及减少评价指标之间相互影响,然后利用结果进行神经网络建模分析,最终通过检验取得良好效果。因此,利用该模型可以有效的预测地震人员伤亡数量,为震后救援工作提供了很好的参考数据。

【参考文献】

- [1] 陈玉华,马玉虎,王培玲,等.青海省地震预报回顾与展望[J].高原地震,2008,20(3):1-6.
- [2] 章在墉.地震危险性分析及应用[M].上海:同济大学出版社,1996.
- [3] 诸井孝文,宫村正光.地震人员伤亡预测与研究现状[J].国际地震动态,1993,(3):25-30.
- [4] 王学民.主成分分析和因子分析应用中值得注意的问题[J].统计与决策,2007,6:141-143.
- [5] 张洁,高慧瑛,刘琦.基于汶川地震的地震人员伤亡预测模型研究[J].中国安全科学报,2011,(3):59-64.
- [6] 于山,王海霞,马亚杰.三层 BP 神经网络地震灾害人员伤亡预测模型[J].地震工程与工程振动,2005,25(6):111-117.
- [7] 王海霞,周晓山.预测人员震害损失的神经网络模型[J].世界地震工程,2007,(4):194-198.
- [8] 马占虎,周志宇,高晓明,等.2006 年甘肃文献 5.0 级地震灾害损失评估[J].西北地震学报,2007,29(3):256-263.
- [9] 杨杰英,李永强,刘丽芳,等.地震三要素对地震伤亡人数的影响分析[J].地震研究,2007,30(2):182-187.
- [10] 范金城,梅长林.数据分析[M].北京:科学出版社,2002.
- [11] 周润景,张丽娜.模糊与神经网络设计[M].北京:电子工业出版社,2010.
- [12] 非明伦,朱月芬,周光全,等.汶川 8.0 级地震现场灾害调查统计与分析[J].地震研究,2008,31(S1):535-543.
- [13] 何永年.中国地震年鉴.2004[M]北京:地震出版社,2005.
- [14] 王根龙,刘红帅,张军慧.汶川特大地震之北川县城震害分析[J].西北地震学报,2010,32(2):173-178.
- [15] 李永强,杨杰英,杨东生.1996 年云南丽江 7.0 级地震人员死亡的社会学特征[J].震灾防御技术,2011,9(3):284-290.