

1 网络设计

1.1 网络输出的确定

根据往年大学生的毕业情况，将毕业生的就业情况分为4类(Y)作为网络的输出： y_1 为容易找到待遇好的工作； y_2 为能找到待遇一般的工作； y_3 为不太容易找到工作； y_4 为几乎找不到工作。该网络采用两个输出节点，分别设这四类的输出值为 $y_1(0,0)$ ， $y_2(0,1)$ ， $y_3(1,0)$ ， $y_4(1,1)$ 。

1.2 网络输入的确定

通过咨询多年从事大学生就业工作的老师，根据他们的经验确定了4类影响就业的指标： x_1 为专业热门程度； x_2 为工作经验丰富程度； x_3 为学习成绩； x_4 为其他能力。

要令以上4个指标作为神经网络的输入，应分别对它们进行取值，其取值标准（为计算方便，令它们的取值都在0~1之间）^[2]如下：

x_1 ：根据往年的就业情况和就业指导教师的经验而定（如：辽宁工程技术大学的采矿专业就业率最高，可以对此专业取值为1.00；信息与计算科学的就业率最低，可以对其取值为0.10；其他专业的取值范围在0~1之间）。

x_2 ：按照大学生在校期间是否曾担任职务（如学生会干部）、兼职、家教以及参加的社会活动等，并根据表现情况而定。

x_3 ：根据学生的加权平均成绩并折合到0~1之间而定（如某学生加权平均成绩为84.56，可以取其值为0.85）。

x_4 ：根据大学生在校期间所得奖励和证明证书而定（如英语四六级证书、日语二级证书、计算机等级证书、高级程序员等），在本文中其最小值为0.5，代表没有通过英语四级（日语二级），也没有其他证书的学生。

以上4类取值除 x_3 外，其他都是按实际情况对各自所有影响因素的隶属程度而得。

1.3 数据

按照评定标准，根据往年毕业学生的资料得到如表1所示的数据。

2 网络的训练

2.1 隐层节点的确定

本文中的输入节点数 $n=4$ ，分别为 (x_1, x_2, x_3, x_4) ；输出节点数为2个，取值分别为0或1；根据神经网络确定的原则^[1]可知，隐层节点数 $m=2n+1=9$ ^[3]。

表1 毕业生资料数据
Tab.1 The data of graduates

人员序号	x_1	x_2	x_3	x_4	期望输出(Y)
1	1.00	0.60	0.73	0.50	(0, 0)
2	0.70	0.70	0.81	0.50	(0, 1)
3	0.10	0.90	0.88	0.80	(1, 0)
4	0.50	0.60	0.67	0.70	(1, 0)
5	0.70	0.50	0.70	0.60	(0, 1)
6	0.80	0.40	0.63	0.60	(0, 1)
7	0.20	0.60	0.87	0.70	(1, 0)
8	0.80	0.30	0.67	0.50	(0, 1)
9	0.80	0.20	0.70	0.50	(0, 1)
10	0.30	0.70	0.79	0.70	(1, 0)
11	0.70	0.40	0.70	0.60	(0, 1)
12	1.00	0.10	0.60	0.50	(0, 1)
13	0.60	0.70	0.82	0.50	(0, 1)
14	0.20	0.90	0.80	0.90	(0, 1)
15	0.40	0.10	0.68	0.60	(1, 1)
16	0.40	0.20	0.62	0.70	(1, 1)
17	0.20	0.80	0.80	0.70	(1, 0)
18	0.10	0.30	0.63	0.50	(1, 1)
19	0.10	0.80	0.77	0.80	(1, 0)
20	0.10	0.90	0.92	0.90	(0, 1)
21	0.60	0.50	0.69	0.50	(0, 1)
22	0.60	0.70	0.73	0.70	(0, 1)
23	0.50	0.10	0.61	0.50	(1, 1)
24	0.70	0.20	0.70	0.60	(0, 1)
25	0.10	0.80	0.91	0.60	(1, 0)
26	0.90	0.10	0.70	0.50	(0, 0)
27	0.90	0.40	0.84	0.50	(0, 0)
28	0.50	0.80	0.87	0.70	(0, 1)
29	1.00	0.70	0.80	0.70	(0, 0)
30	0.50	0.50	0.79	0.60	(1, 0)
31	0.30	0.40	0.65	0.50	(1, 1)
32	1.00	0.50	0.87	0.70	(0, 0)
33	0.70	0.90	0.91	0.80	(0, 0)
34	0.10	0.40	0.64	0.50	(1, 1)
35	0.10	0.30	0.61	0.50	(1, 1)
36	0.20	0.40	0.60	0.50	(1, 1)
37	0.60	0.60	0.64	0.50	(1, 0)
38	0.40	0.30	0.61	0.50	(1, 1)
39	0.90	0.60	0.88	0.70	(0, 0)
40	0.60	0.70	0.89	0.80	(0, 1)

注：资料来自辽宁工程技术大学就业指导中心

2.2 训练网络

对以上数据进行处理以及对网络的训练是通过 Matlab^[4]实现的，所用语句为^[5]：“net=newff(minmax(p), [9, 2], {'tansig', 'logsig'}, 'trainlm');”。

在训练了 112 次以后，误差达到 0.05，满足精度要求。

网络的训练结果为：

TRAINLM, Epoch 0/300, MSE 0.429508/
0.05, Gradient 5.13673/1e-010

TRAINLM, Epoch 25/300, MSE
0.0961575/0.05, Gradient 0.123247/1e-010

TRAINLM, Epoch 50/300, MSE
0.0874426/0.05, Gradient 0.0906441/1e-010

TRAINLM, Epoch 75/300, MSE
0.0667494/0.05, Gradient 0.0656093/1e-010

TRAINLM, Epoch 100/300, MSE
0.056572/0.05, Gradient 0.198743/1e-010

TRAINLM, Epoch 112/300, MSE
0.0498696/0.05, Gradient 0.0678862/1e-010

TRAINLM, Performance goal met.

输出的误差图如图 1 所示。

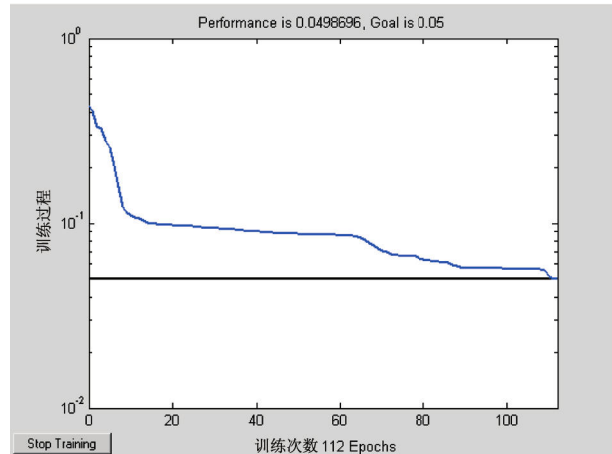


图 1 训练网络误差输出图
Fig. 1 Error output of training network

2.3 网络模型可信度的判别原则

要判断此网络模型是否可信，采用一个 λ 判别原则，即训练好之后的网络的实际输出与期望输出之间会有一定的误差，如果这些误差在一个规定的置信值 (λ) 范围内 [本文的取值在 (0.05~0.10) 之间]，就可以说此网络预测模型是可信的，即可以用来预测。超过了这个误差就可以说误差稍大或较大，也就是网络可信度低或是不可信。表 2 是网络的实际输出与期望输出之间的误差以及在误差允许范围内的实际输出值。

从表 2 可以看出，期望输出与实际输出在误差允许的范围是一致的，可信度可以达到 95% (38/40)，这个可信度很高^[6]。因此，可以利用这个神经网络模型进行预测。

2.4 利用网络进行预测

现有两名待预测的学生 (他们的毕业情况已知)： $s_1(1.00, 1.00, 0.70, 0.70)$ ， $s_2(0.10, 0.10, 0.90, 0.50)$ ；将其作为输入代入上述网络中，输出结果分别为 $y_1(0.000 0, 0.000 7)$ ， $y_2(0.859 6, 1.000 0)$ 。

表 2 期望输出与实际输出的对比

Tab. 2 The expected output compared with the actual output

人员序号	期望输出 (Y)	实际输出	误差允许内的实际输出
1	(0, 0)	(0.000 0, 0.003 2)	(0, 0)
2	(0, 1)	(0.002 1, 0.955 1)	(0, 1)
3	(1, 0)	(0.982 4, 0.003 5)	(1, 0)
4	(1, 0)	(1.000 0, 0.000 0)	(1, 0)
5	(0, 1)	(0.000 0, 1.000 0)	(0, 1)
6	(0, 1)	(0.000 0, 0.999 9)	(0, 1)
7	(1, 0)	(0.999 9, 0.000 4)	(1, 0)
8	(0, 1)	(0.000 0, 1.000 0)	(0, 1)
9	(0, 1)	(0.000 8, 1.000 0)	(0, 1)
10	(1, 0)	(0.757 5, 0.178 2)	(1, 0)
11	(0, 1)	(0.000 0, 1.000 0)	(0, 1)
12	(0, 1)	(0.000 0, 0.950 7)	(0, 1)
13	(0, 1)	(0.000 3, 0.993 6)	(0, 1)
14	(0, 1)	(0.041 7, 0.980 6)	(0, 1)
15	(1, 1)	(1.000 0, 0.996 9)	(1, 1)
16	(1, 1)	(1.000 0, 0.975 7)	(1, 1)
17	(1, 0)	(0.996 2, 0.001 3)	(1, 0)
18	(1, 1)	(0.973 7, 1.000 0)	(1, 1)
19	(1, 0)	(0.999 0, 0.010 5)	(1, 0)
20	(0, 1)	(0.511 6, 0.936 9)	误差较大
21	(0, 1)	(0.000 0, 0.936 7)	(0, 1)
22	(0, 1)	(0.030 6, 0.991 2)	(0, 1)
23	(1, 1)	(0.987 0, 1.000 0)	(1, 1)
24	(0, 1)	(0.014 0, 0.998 4)	(0, 1)
25	(1, 0)	(0.977 1, 0.000 2)	(1, 0)
26	(0, 0)	(0.003 3, 0.000 0)	(0, 0)
27	(0, 0)	(0.004 6, 0.139 2)	(0, 0)
28	(0, 1)	(0.051 8, 0.776 5)	(0, 1)

可见 y_1 在上面给出的置信值 λ 允许的范围内可以取 (0, 0), 即就业情况是容易找到待遇好的工作; y_2 在上面给出的置信值 λ 允许的范围内可取 (1, 1), 即就业情况是几乎找不到工作。

3 结论

根据 BP 神经网络的原理, 对大学生就业这一实际情况建立了神经网络模型, 并通过对神经网络的训练得到了初步模型结果。但这还不能说明模型的可靠。为了保

人员序号	期望输出 (Y)	实际输出	误差允许内的实际输出
29	(0, 0)	(0.000 0, 0.009 3)	(0, 0)
30	(1, 0)	(0.978 0, 0.020 5)	(1, 0)
31	(1, 1)	(0.869 7, 0.992 3)	(1, 1)
32	(0, 0)	(0.000 3, 0.188 0)	(0, 0)
33	(0, 0)	(0.000 0, 0.000 0)	(0, 0)
34	(1, 1)	(0.797 2, 1.000 0)	(1, 1)
35	(1, 1)	(0.996 0, 0.907 2)	(1, 1)
36	(1, 0)	(0.971 7, 0.896 6)	(1, 0)
37	(1, 0)	(0.854 5, 0.073 1)	(1, 0)
38	(1, 1)	(0.980 3, 0.997 6)	(1, 1)
39	(0, 0)	(0.000 0, 0.053 2)	(0, 0)
40	(0, 1)	(0.271 2, 1.000 0)	误差较大

续表

证文章的严密性, 本文又依据可信度判别原则, 对结果进行了可信度分析。结果表明可信度高达 95%, 因此, 此 BP 神经网络模型可以很好的用来预测大学生的就业情况。

本模型可以根据任何一位在校人员的在校表现对其未来进行预测, 实用性广。通过本文可以掌握一种基于神经网络建立模型的方法, 此方法不仅可以用在预测大学生就业情况上, 还可以实现对价格、企业破产、经济等方面的预测。

[参考文献] (References)

- [1] 郭嗣琮, 陈刚. 信息科学中的软计算方法[M]. 沈阳: 东北大学出版社, 2001
GUO S Z, CHEN G. Method of soft computing in information science[M]. Shenyang: NEU Press, 2001. (in Chinese)
- [2] 孙旭东. 大学生学习生活表现分类预警模型[OL]. [2006-9-29].
http://www.paper.edu.cn/paper.php?serial_number=200609-429
SUN X D. A classify & predict model of college student' study and life behavior [OL]. [2006-9-29].
http://www.paper.edu.cn/paper.php?serial_number=200609-429 (in Chinese)
- [3] 余岚, 周清华. 基于模糊聚类的 BP 神经网络模式识别法[OL]. [2006-9-26].
http://www.paper.edu.cn/paper.php?serial_number=200609-378
YU L, ZHOU Q H. BP neural network pattern identification based on fuzzily gathers[OL]. [2006-9-26].
http://www.paper.edu.cn/paper.php?serial_number=200609-378 (in Chinese)
- [4] 飞思科技产品研发中心. 神经网络理论与 Matlab7 实现[M]. 北京: 电子工业出版社, 2005.
Fei-Si Science and Technology R&D Center. Neural network theory and Matlab 7's realization [M]. Beijing: Electronics Industry Press, 2005. (in Chinese)
- [5] 马兴义, 杨立群, 林敏, 等. Matlab 6 应用开发指南[M]. 北京: 机械工业出版社, 2002.
MA X Y, YANG L Q, LIN M, et al. The application development guide of Matlab 6[M]. Beijing: China Machine Press, 2002. (in Chinese)
- [6] 茆诗松, 周纪芾. 概率论与数理统计 (2 版) [M]. 北京: 中国统计出版社, 2000.
MAO S S, ZHOU J X. Probability and statistics (Second Edition) [M]. Beijing: China Statistics Press, 2000. (in Chinese)