

基于人工神经网络的地区高等教育水平评价

孙阳春, 范智勇

(大连理工大学 高等教育研究中心, 辽宁 大连 116023)

摘要: 基于人工神经网络的BP算法, 选取高校数量、中央部委所属普通高校数量和每百万人口普通高校数量等7个指标, 对我国31个省市的高等教育发展水平进行综合评价, 进而获得各地区的高等教育发展水平情况, 了解其特点。

关键词: 高等教育水平; 人工神经网络; BP算法
中图分类号: G521.9 **文献标识码:** A

文章编号: 1008-0627(2012)03-0053-04

现阶段,我国各地区的高等教育发展水平差异很大,地区间高等教育发展水平并不均衡。^[1] 高等教育的发展水平一方面关系到地区经济水平的发展,另外也涉及到高等教育公平性问题,因此如何实现地区高等教育的均衡发展显得尤为重要。目前,有关区域高等教育的研究越来越多,其中有运用熵值法进行定量研究,并得出我国东、中、西三大地区间的高等教育实力发展不平衡,^[2]也有一部分学者在评价指标选取上研究的较为深入。本研究在指标选取上综合了其他学者的研究成果,选取了体现高等教育发展水平的质量指标和规模指标,以及地区在高等教育发展水平上的受益指标;在研究方法选择上,运用人工神经网络方法中的BP算法(反向传播算法)对31个省市的高等教育水平进行综合评价,这一评价方法立足于客观实际情况,而且对于多因素的复杂系统有较好的适应性,研究结果更具有客观性。

一、评价指标的选取

对地区高等教育水平进行评价,其评价质量的高低关键取决于评价方法和评价指标的选择是否合理。因此,在评价指标选取过程中,应尽可能地保证评价指标的全面性和有效性。有学者将各省市高等教育发展水平分为地域型高等教育发展水平和受益型高等教育发展水平,并用每万人口在校大学生数代表地域型高等教育发展

水平,用每万人口大学毕业生数代表受益型高等教育发展水平。^[3]也有学者在比较高等教育区域差异中,选取学校数量、教职工人数等指标进行比较研究。^[4]本研究选取2003年我国各省市的普通高校数量(X_1)、中央部委所属普通高校数量(X_2)、每百万人口普通高校数(X_3)、每十万人口在校大学生数(X_4)、每十万人口大学毕业生数(X_5)、普通高校专任教师人数(X_6)、普通高校教授人数(X_7)7项指标。通过这7项指标建立的BP评价模型,进一步可以得出我国31个省市的高等教育发展水平及其特点。

二、评价方法及评价过程

(一) BP神经网络的原理及其算法

BP神经网络是1986年由Rumelhart和Mcelland为首的科学家小组提出的,它是一种基于误差反向传播算法的多层向前网络,其网络结构一般由2层或2层以上组成,层与层之间的神经元实现权连接,层内的神经元无连接,典型的BP网络是2层的前馈网络,第一层是隐含层,第二层是输出层。输入层神经元负责接受外界的输入信息,并将信息传递给隐含层神经元,最后传递到输出层神经元,经过进一步信息处理之后即完成一次学习过程,当实际输出与期望输出不符合时,则进入误差的反向传播阶段,按照误差梯度下降的方法修正各层间的权值,依次向隐含层、输入层反向传播。随着信息正向传播和误差

收稿日期: 2012-01-10

基金项目: 教育部人文社科青年基金项目(10YJZJ135); 第49批中国博士后基金资助项目(20110491532)

第一作者简介: 孙阳春(1977-),女,辽宁庄河人,副教授/博士,主要研究方向: 高等教育原理。E-mail: sunyangchun2003@163.com

反向传播的不断进行,各层间的权值不断地得到修正,直到网络输出的误差达到可以接受的范围,这时网络学习过程结束。

BP神经网络的具体算法如下:^[5]

1. 初始化网络,设定学习率 μ ,允许误差 ε 和最大学习次数 N 。

2. 随机选取输入样本 k 及其对应的期望输出: $x(k) = (x_1(k), x_2(k), \dots, x_n(k))$

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_o(k))^2$$

3. 计算隐含层和输出层各神经元的输入和输出:

$$hi(k) = \sum_{i=1}^n \omega_{ih} x_i(k) - b_h \quad h=1, 2, \dots, p$$

$$ho_h(k) = f(hi_h(k)) \quad h=1, 2, \dots, p$$

$$yi_o(k) = \sum_{h=1}^p \omega_{ho} ho_h(k) - b_o \quad o=1, 2, \dots, q$$

$$yo_o(k) = f(yi_o(k)) \quad o=1, 2, \dots, q$$

其中, ω_{ih} 为输入层与隐含层的连接权值, ω_{ho} 为隐含层与输出层的连接权值; b_h 为隐含层各神经元的阈值, b_o 为输出层各神经元的阈值;

hi 和 ho 分别为隐含层的输入和输出向量; yi 和 yo 分别为输出层的输入和输出向量; n 、 p 和 q 分别为输入层、隐含层和输出层神经元个数; $f(\cdot)$ 为激活函数。

4. 利用网络期望输出和实际输出,计算误差函数对输出层各神经元的偏导数 $\delta_o(k)$,并结合隐含层到输出层的连接权值,计算误差函数对隐含层各神经元的偏导数 $\delta_h(k)$ 。

5. 利用输出层各神经元的 $\delta_o(k)$ 和隐含层各神经元的输出修正连接权值 $\omega_{ho}(k)$,利用隐含层各神经元的 $\delta_h(k)$ 和输入层各神经元的输入修正连接权值 $\omega_{ih}(k)$ 。

6. 计算全局误差:

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_o(k))^2$$

其中, m 为样本数据个数。

7. 判断网络误差是否满足要求,若误差达到设定误差要求或者学习次数超过最大学习次数,则完成算法。否则选取下一个学习样本及其对应期望输出,返回到第三步,进入下一轮学习。

(二) 地区高等教育水平评价

从《中国教育统计年鉴(2003年)》中可以查出我国31个省市各项指标的具体数值(表1)。

将表1中的各列数据进行归一化处理,这样数据就分布在[-1,1]之间,从数据的分布来看,

表1 31个省市各项指标数据统计(部分)

省市	X_1	X_2	X_3	X_4	X_5	X_6	X_7
北京	73	33	5.01	3121.4	568.87	41904	4619
天津	37	3	3.66	2425.5	397.83	15553	1421
河北	83	5	1.23	850.26	167.59	33617	1918
山西	45	0	1.36	823.68	123.05	20224	1002
内蒙古	27	0	1.13	662.2	104.7	12153	538
辽宁	70	5	1.66	1221.4	234.94	38086	2662
吉林	40	2	1.48	1181.6	194.55	21824	1922
黑龙江	54	3	1.42	1028.2	181	28525	2486
上海	56	9	3.27	2212.3	415.89	24387	2707
江苏	94	10	1.27	1160.8	185.05	49810	3406
浙江	64	2	1.37	1034.5	168.13	29508	1796
安徽	73	2	1.14	639.69	102.47	24744	1250
青海	12	0	2.25	489.21	89.345	2769	89
宁夏	12	1	2.07	605.76	94.155	3415	201
新疆	26	0	1.34	763.32	130.63	10913	371

各列数据多为偏态分布,因此不宜采用等间距线性内插法来构建神经网络的训练数据,故采用自然断点法来选取网络训练数据(表 2)。将评价等级划分成 5 个等级,其中 5 代表这一省市的高等教育水平高,4 代表其高等教育水平较高,3 表示其高等教育水平一般,2 表示其高等教育水平较差,1 表示其高等教育发展水平差。网络的拓扑结构为 $7 \times 15 \times 1$,中间层节点选择数为 15,从训练的结果来看这一节点选择较为恰当,其中输入层得节点数为 7,输出层节点数位 1。训练参数设置如下:基本学习速率为 0.05,最大训练次数为 10000,最大误差为 0.001,采用 Matlab7.5 进行仿真实验,网络训练的误差逼近(图 1)。

图 1 显示,网络训练达到误差要求,即网络训练达标,将表 1 中的数据归一化后,输入已训练好的网络,即可得到 31 个省市的高等教育水平评价值(表 3)。

由表 3 中 31 个省市高等教育水平的评价数据,可以计算出全国高等教育水平的平均值为 2.864,由此可知从全国高等教育发展水平来看,总体水平一般,其中北京 4.9887,上海 4.9335,天津 4.5556,江苏 4.0845,辽宁 4.1535,湖北 4.0761,上述 6 个省市的高等教育水平较高,远超过全国高等教育平均水平。从上述几个省市的

地域分布来看,多处在东部沿海地区,而且经济

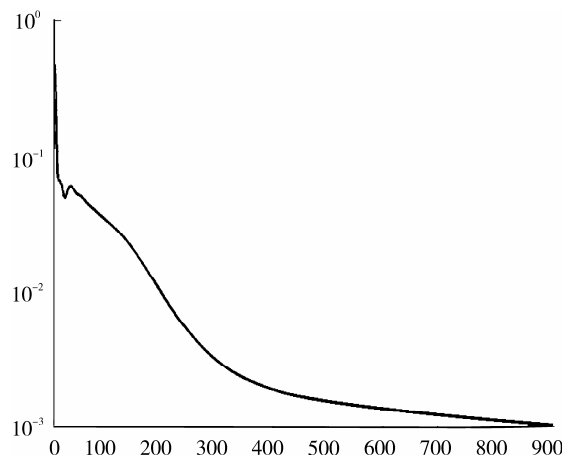


图 1 网络训练误差逼近

较为发达。从表 3 中也可以看出内蒙古、海南、贵州、云南、西藏、甘肃、青海、宁夏、新疆等 9 个省市高等教育发展水平较差,而这 9 个省市多处于西部地区,经济发展相对落后。将上述东部地区和西部地区的高等教育水平评价值进行比较,不难得出高等教育发展水平和地方的经济发展水平有很大的关系。

由表 3 可知,在 31 个省市中高于全国高等教育平均水平的有 14 个,多处于东中部地区,而处于西部地区的省市几乎全部低于全国高等教育平均水平,可见我国高等教育发展水平不

表 2 31 个省市高等教育水平 BP 模型评价标准

网络输入层 Input layer of network							输出层
P_1	P_2	P_3	P_4	P_5	P_6	P_7	Output layer
0.9	0.303	0.6852	0.7456	0.6966	0.9414	0.737	5
0.7444	0.1212	0.1782	0.2372	0.2308	0.6567	0.4153	4
0.5556	0.0606	0.1294	0.1399	0.1309	0.427	0.2697	3
0.3333	0	0.0869	0.0745	0.0586	0.2306	0.1164	2
0.0778	0	0.0051	0.0002	0.0018	0.0354	0.018	1

表 3 31 个省市的高等教育水平 BP 评价值

省市	评价值	省市	评价值	省市	评价值	省市	评价值
北京	4.9887	上海	4.9335	湖北	4.0761	云南	1.7396
天津	4.5556	江苏	4.0845	湖南	3.0994	西藏	1.3043
河北	3.3547	浙江	3.4588	广东	3.0651	陕西	3.8471
山西	2.71	安徽	2.7878	广西	2.0211	甘肃	1.955
内蒙古	1.7843	福建	2.249	海南	1.3299	青海	1.8756
辽宁	4.1535	江西	2.7672	重庆	2.1386	宁夏	1.7665
吉林	3.2635	山东	3.7089	四川	2.199	新疆	1.795
黑龙江	3.4876	河南	2.6033	贵州	1.6859		

均衡。

三、高等教育水平评价启示

1. 加大对西部高等教育的支持力度,从各方面着手提高西部地区的高等教育发展水平。高等教育水平的提高离不开政策和资金的支持,从上述研究结果来看西部地区的高等教育水平远落后于全国平均水平,因此政府应加大对西部高等教育发展的经费支持,另外在高等教育招生中应多给西部高校一些指标,并且提高对西部高校的师资培养,增加西部高校的教授人数等等。

2. 加快西部大开发战略,促进西部地区经济的快速发展。周志刚等研究发现我国各地区高等教育发展的区域差异性的主要动因来自于区域经济发展的不平衡性。^[6]因此要想提高西部地区的高等教育发展水平,同时应提高西部地区的经济发展水平。从西部大开发的角度来看,应加快西部大开发战略,切实有效地提高西部地区的经济水平。

3. 立足地区高等教育发展水平的评价结

果,优化高等教育区域结构布局。从31个省市的高等教育发展水平评价结果来看,一方面每一个地区都能清楚地看到自身所处的位置,这样就有利于各地区认识到自身高等教育的发展水平,另一方面在优化高等教育区域结构布局中这一评价结果也可以作为参考的标准。

参考文献

- [1] 相丽君,徐新.我国高等教育发展水平的区域差异[J].统计与决策,2008(14):94-96.
- [2] 刘丽.区域高等教育发展实力分析[J].教育发展研究,2009(19):41-44.
- [3] 杜育红.我国地区间高等教育发展差异的实证分析[J].高等教育研究,2000(3):44-48.
- [4] 张艳红,邢立新.基于GIS对中国大陆高等教育区域差异分析[J].现代教育科学,2008(2):32-35.
- [5] 杨行峻,郑君里.神经网络[M].北京:高等教育出版社,1992:126-128.
- [6] 周志刚,王凤慧.我国高等教育发展区域差异性的实证研究[J].西南交通大学学报:社会科学版,2008(8):43-47.

An Evaluation on Regional Higher Education Based on Artificial Neural Network

SUN Yang-chun, FAN Zhi-yong

(Center of Higher Education Research, Dalian University of Technology, Dalian 116023, China)

Abstract: The purpose of the study is to survey sampled local universities, ministry-affiliated higher learning institutions and colleges per million population in terms of 7 indexes by the BP algorithm of artificial neural network over 31 provinces and municipalities, in an attempt to assess the regional status of higher education development and traits involved.

Key words: higher education development; artificial neural network; BP algorithm

(责任编辑 夏登武)