

文章编号: 2095-2163(2022)08-0043-06

中图分类号: C913.6

文献标志码: A

基于BP神经网络的养老床位供需分配分析

朱辰超, 刘雅雅

(上海理工大学 管理学院, 上海 200093)

摘要:近年来中国人口老龄化问题日益严重,养老床位的需求日益增加,养老床位的合理化区域分配迫在眉睫。本文选用6个指标:65岁及以上人口数量、城镇单位就业人员数、城镇居民消费水平、城镇人口数据、基本养老保险基金支出、老年人口抚养比,收集历年中国31个省份的数据作为历史样本;运用K均值聚类得到床位需求高、中、低3个类别。依据分类结果分别进行BP神经网络的训练和判别分析的求解,对比发现BP神经网络的精确度更高,更适合用于养老床位需求的分类;运用灰色预测模型和差分整合移动平均自回归模型,得到2022年6个指标的预测值,将其输入BP神经网络,实现对2022年各省养老床位需求的分类;最后,运用熵权法对各类中各省养老床位的需求程度打分,提出一种地区化分配养老床位的方案。研究成果将对养老床位在中国各省的分配提供理论参考,实现养老床位供需均衡。

关键词:灰色预测模型; BP神经网络; 养老床位; K均值聚类; 差分整合移动平均自回归模型; 熵权法

Analysis of supply and demand distribution of pension service beds based on BP neural network

ZHU Chenchao, LIU Yaya

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

[Abstract] Recently, the problem of aging population in China has become increasingly serious, which caused the demand for pension service beds was increasing. Therefore, it is urgent to realize the reasonably regional allocation of pension service beds. The six indexes that number of people aged 65 and above, number of employed persons in urban units, consumption level of urban residents, number of urban population, expenditure of basic endowment insurance fund, and dependency ratio of elderly population are chosen to collect the corresponding data in 31 provinces of China over the years. According to these data, K-means clustering is used to get three classes such as maximum demand, moderate demand and minimum demand. The BP neural network is trained and discriminant analysis is performed at the same time to achieve the accuracy. Based on the predicted results of six indexes in 2022 by Grey Forecast Model and ARIMA (Autoregressive Integrated Moving Average Model), BP neural network has higher accuracy and is more suitable to be selected for predicting the classes of 31 provinces in this period. Finally, the score of every province is calculated by the entropy weight method to evaluate the degree of demand of pension service beds from a regional perspective. The research results provide a plan of regional allocation of nursing beds, and theoretical reference for government to distribute the pension service beds in 31 provinces of China, for realizing the balance between supply and demand of pension service beds.

[Key words] grey forecast model; BP neural network; pension service beds; K-means clustering; autoregressive integrated moving average model; entropy weight method

0 引言

中国老龄化问题日益突出,家庭养老的形式难以满足老年人的养老需求。李林潼^[1]经过调查研究,发现解决现状的重要方法是采用机构养老的形式,但机构养老在现阶段的实施中存在床位供需失衡的问题。关于养老床位需求的研究,余莹等人^[2]运用系统聚类方法分析上海市各个区的养老区域划分。徐佩等人^[3]用时间序列模型对上海市的机构养老床位供给量进行预测。除了针对某一个具体省

份分析外,曹稀哲等人^[4]用多元回归的方法来预测未来六年中国整体的机构床位需求量。王子鑫等人^[5]分析了城市和乡村老年人的养老床位需求。大部分关于养老机构床位的研究都只是分析整体的需求量、或者城乡的需求量,很少细化到各个省份来探讨分析。虽然高巍等人^[6]基于31个省份的数据,运用多目标规划来分析养老服务运营模式,但是最终并没有提出各个省份具体的养老服务运营策略。近年来,很多学者采用聚类分析与判别分析结合的方法研究区域化和老年人能力等级的分类问题。但

基金项目: 国家自然科学基金(62006154)。

作者简介: 朱辰超(2001-),女,本科生,主要研究方向:智能信息处理。

通讯作者: 刘雅雅 Email: yayaliu@my.swjtu.edu.cn

收稿日期: 2022-02-14

是判别分析有很多种方法,选择不同方法,得到的判别函数也不同,有误判的现象发生。

综上所述,现阶段对于机构养老床位的需求分析并没有细化到各个省份,只是停留在城乡的阶段。基于此,本文主要采用BP神经网络和灰色预测模型,地区化分析31个省份养老机构床位的需求量,力图实现省份间的合理分配,解决机构养老床位供需失衡的问题。同时将判别分析与BP神经网络的精确度进行对比分析,以期减少误判现象,实现地区的准确分类。选用65岁及以上人口数、城镇单位就业人员数、城镇居民消费水平、城镇人口数、基本养老保险基金支出、老年人口抚养比这6个指标,考虑到养老床位需求受到历史数据的影响,故收集历年中国31个省份的指标数据,运用K均值聚类的方法将历年省份分为3类,分别为养老床位需求高、中、低的地区,并将分类结果用于BP神经网络的训练。再采用灰色预测模型和差分整合移动平均自回归模型来预测2022年各指标值,对2022年中国各省养老床位的需求进行分类。最后,基于2022年的分类结果,通过熵权法对各类中各省的养老床位需求程度打分,根据分数的高低提出一种地区化分配养老床位的方案,为养老床位供需失衡问题提供解决方案支持。同时,用判别分析对2022年31个省份进行分类,将判别分析的结果与BP神经网络的结果进行对比,实现误判最小化。

1 预备知识

1.1 BP神经网络

BP神经网络(Back Propagation Neural Network)采用误差的反向传播和信号的正向传播来确定权重,误差按照负梯度的方向不断传播,直至误差达到一个极小值时停止,或者在模型的学习次数达到初始设定的最大学习次数时停止,得到最终的完整网络。在BP神经网络中,最常见的是3层反馈网络,分别为输入层、隐含层、输出层。在BP神经网络中,隐含层神经细胞的数量由式(1)来求得:

$$N = \sqrt{N_a + N_b} + a \quad (1)$$

其中, N_a 是输入层中神经元的数量; N_b 是输出层神经元的数量; a 是调节常数,其范围在[1,10]之间。

隐含层神经元的数量需要根据过去的的数据逐步训练BP神经网络,直至得到预测误差最小的网络,从而确定隐含层神经元数量。

在计算过程中,本文拟采用tansig函数作为隐

含层的激活函数,线性传递purelin函数作为输出层的激活函数。设输入层的输入为分类后的2008~2017年各省的65岁及以上人口数 X_1 、城镇单位就业人员数 X_2 、城镇居民消费水平 X_3 、城镇人口数 X_4 、基本养老保险基金支出 X_5 、老年人口抚养比 X_6 ,该层使用tansig函数,对此可表示为:

$$t_j = s\left(\sum_{i=1}^{N_a} W_{ij} X_i + \alpha_j\right) \quad (2)$$

其中, t_j 为隐含层第j个神经元的输出; W_{ij} 为隐含层神经元j与输入层神经元i的权重; α_j 为隐含层神经元的阈值; $s()$ 为隐含层神经元的激活函数tansig。

将高、中、低三个需求类别作为输出层神经元的3个输出值,神经元运用激活函数purelin将激活后的输入值输出,purelin函数的数学形式可表示为:

$$Y = d\left(\sum_{j=1}^{N_b} W_{jb} t_j + \alpha_b\right) \quad (3)$$

其中, W_{jb} 为隐含层神经元j到输出层神经元b的权重; α_b 为输出层神经元b的阈值; $d()$ 为输出层神经元的激活函数purelin。

1.2 判别分析

判别分析要求每类中至少有一个样本,且解释变量必须是可测量的,才能计算每类的均值和方差,用于判别函数的具体表达式与计算。判别分析的计算步骤如下:

- (1) 计算各组类别中各判别变量的均值和协方差阵。
- (2) 计算协方差阵的估计值的逆矩阵。
- (3) 运用逆矩阵和均值确定判别函数的具体表达式。
- (4) 运用判别函数对已有样本进行回判,计算精确度。
- (5) 运用判别函数对待判别样本进行分类。

2 数据预处理

本文从国家统计局中收集2008~2017年中国31个省份的6个指标值作为样本数据。首先,运用数据分析软件SPSS(Statistical Product and Service Solutions)对数据进行归一化处理,去除量纲对于聚类结果的影响;其次,采用K均值聚类的方法将样本分为3类,分别为:床位需求最多、床位需求适中和床位需求最少,最终的聚类结果见表1和表2,最终聚类中心表示各类别中各指标数据的均值。

表 1 最终聚类中心

Tab. 1 Final clustering center

类别	指标					
	65 岁及以上 人口数/人	城镇单位就业 人口/万人	城镇居民消费 水平/元	城镇人口/ 万人	基本养老保险 基金支出/万元	老年人口 抚养比/%
类别 1	0.273 05	0.166 42	0.372 86	0.234 75	0.366 24	0.572 08
类别 2	-0.643 18	-0.635 47	-0.504 99	-0.679 67	-0.684 08	-0.607 27
类别 3	1.670 37	1.984 79	0.798 79	1.939 18	1.531 31	0.560 40

表 2 各类中的样本数

Tab. 2 Number of samples in each category

数量	类别 1	类别 2	类别 3
样本数	110.000	135.000	34.000
总数	279.000	279.000	279.000

由表 1 的数值可知,第一类对应的是床位需求适中的地区,第二类对应的是床位需求最少的地区,第三类对应的是床位需求最多的地区。因每年养老床位的需求都受到过去情况的影响,具有连续性,不是独立于历史存在的离散值,故本文以历史样本数据的分类结果为后续模型预测的基础样本,结果见表 2,所有收集的样本均实现分类,并不存在一个样本属于多个类别的情况。

3 模型应用

3.1 最佳 BP 神经网络训练

依据 K 均值聚类方法的分类结果,将每个类别的 70% 样本作为训练集,30% 样本作为测试集。由于输入变量为 6 个,输出变量为 3 个,根据公式 (1),隐含层个数确定在 [4, 13] 内。在该区间内选取不同的隐含层个数,训练 BP 神经网络,对比各个隐含层个数下 BP 神经网络的预测精度,选择精度最高的网络对应的隐含层个数,最终确定为 10。养老床位需求 BP 神经网络结构如图 1 所示,分类误差直方图如图 2 所示。

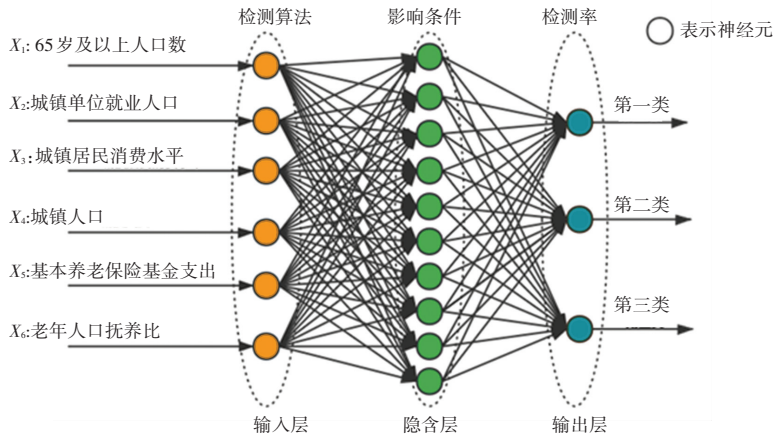


图 1 养老床位需求 BP 神经网络结构

Fig. 1 The BP neural network structure of pension service beds demand

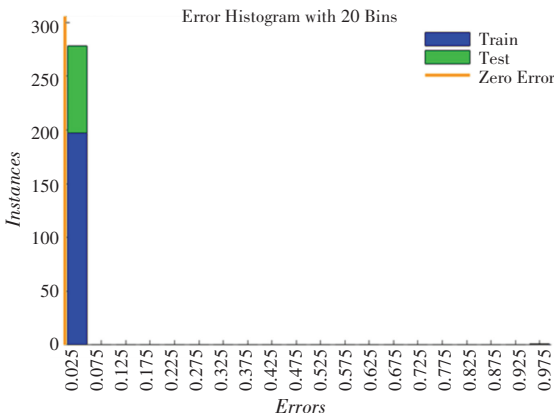


图 2 分类误差直方图

Fig. 2 Histogram of classification error

在神经网络训练的过程中,需要重视过拟合问题。当训练集的精确度极高,而测试集的精确度低时,就会产生过拟合现象。由图 2 可以看出,训练集和测试集的误差均在 0.025 附近,即训练集和测试集的预测精确度基本均接近于 100%, 本文的养老床位需求 BP 神经网络并不存在过拟合现象。

3.2 BP 神经网络与判别分析精确度的对比

判别分析是对已知样本数据进行计算得到判别函数,选择不同方法得到的判别函数也不相同,常用的判别方法有最大似然法、距离判别法、Fisher 判别法、Bayes 判别法。本文将 K 均值聚类方法的分类结果作为基础样本,以 Fisher 判别法为例,通过判别

函数对基础样本进行回判,从而求得判别分析的精确度。与养老床位需求 BP 神经网络的精确度相比,判别分析的精确度较低,故本文选用 BP 神经网络的方法来实现养老床位需求的分类。

3.3 各个指标值预测

本文采用灰色预测模型和差分整合移动平均自回归模型,根据 2008~2017 年 31 个省份的 6 个指标值,预测 2022 年的指标值。灰色预测模型是运用少量数据进行短期预测,在计算残差的方差和指标数据的方差后,将两者的比值称为后验差比值,进行

后验差比检验。当后验差比值大于 0.65 时,则表明灰色预测模型的预测结果对于该指标数据而言不合格,需要选用其他方法进行预测。差分整合移动平均自回归模型是对平稳的时间序列数据进行中短期预测,当数据不平稳时,需要对数据进行差分,使其变成平稳序列。在本文的指标预测的过程中,对每个指标都先运用灰色预测模型进行预测和检验,当检验结果显示不合格,即精确度过低时,将采用差分整合移动平均自回归模型进行预测,最终的预测结果见表 3。

表 3 2022 年 6 个指标 31 个省份的预测值

Tab. 3 Predicted values of six indicators in 31 provinces of China in 2022

2022 年	65 岁及以上 人口数/人	城镇单位 就业人口/万人	城镇居民 消费水平/元	城镇人口 /万人	基本养老保险 基金支出/万元	老年人口 抚养比/%
北京	1 487 982.0	766.002 5	99 419.65	1 842.798 0	19 629 130	12.170 740
天津	1 472 313.0	266.097 7	70 218.07	1 323.585 0	13 319 579	16.869 420
河北	8 567 827.0	641.477 0	25 483.97	4 759.658 0	36 624 835	23.831 060
山西	4 027 498.0	457.537 1	33 692.99	2 376.599 0	14 621 749	20.885 580
内蒙古	1 658 395.0	278.573 3	43 462.27	1 672.693 0	25 150 060	11.557 750
辽宁	5 936 073.0	469.010 3	42 162.02	2 990.439 0	38 662 097	27.005 230
吉林	2 851 461.0	233.075 2	23 470.47	1 613.338 0	26 728 530	20.073 130
黑龙江	4 350 020.0	277.269 2	37 800.62	2 336.298 0	29 997 216	22.423 920
上海	3 440 388.0	855.789 7	82 984.72	2 180.114 0	30 750 031	28.064 280
江苏	8 777 489.0	1 157.917 0	87 473.38	5 972.571 0	43 275 687	24.523 010
浙江	7 324 364.0	891.845 1	51 090.03	4 492.306 0	39 423 649	24.142 510
安徽	7 257 321.0	703.827 0	35 006.64	3 890.583 0	24 864 625	23.222 290
福建	3 107 559.0	611.115 7	43 575.76	2 814.561 0	9 940 904	14.318 110
江西	3 632 046.0	426.908 5	31 419.46	2 930.324 0	15 266 083	15.064 930
山东	12 402 000.0	912.370 4	57 089.71	6 403.501 0	36 394 219	34.451 170
河南	8 022 973.0	736.647 8	38 148.53	5 675.673 0	22 757 992	19.285 340
湖北	6 156 384.0	588.843 0	45 574.06	3 797.931 0	34 513 670	20.343 760
湖南	7 383 908.0	636.284 9	37 773.72	4 303.303 0	13 715 428	23.482 490
广东	8 659 326.0	2 221.010 0	49 068.00	8 908.160 0	88 513 670	13.310 790
广西	3 958 797.0	408.831 3	31 820.66	2 743.833 0	12 539 406	15.878 130
海南	777 331.8	103.549 2	53 665.03	597.429 7	3 730 184	15.950 000
重庆	3 893 540.0	331.450 5	46 465.67	2 274.852 0	13 752 027	25.843 320
四川	9 284 944.0	780.631 8	31 215.64	4 974.710 0	32 100 030	29.467 990
贵州	3 884 745.0	327.070 8	38 551.66	1 986.654 0	6 868 788	23.341 540
云南	4 506 890.0	311.110 4	34 471.28	2 594.114 0	4 754 503	17.669 080
西藏	168 484.4	69.830 64	33 738.76	122.255 8	1 529 201	9.919 364
陕西	3 811 162.0	483.594 20	35 270.17	2 508.313 0	16 245 489	18.366 130
甘肃	2 632 631.0	240.702 60	34 253.81	1 391.851 0	12 264 184	19.392 360
青海	419 210.4	72.255 32	38 894.70	369.686 0	6 548 277	12.837 110
宁夏	580 033.7	67.506 31	42 353.22	449.476 2	3 522 633	17.276 090
新疆	1 754 576.0	288.100 20	34 938.94	1 480.547 0	12 053 859	14.374 270

3.4 中国 31 个省份的分类结果预测

根据差分整合移动平均自回归模型和灰色预测模型预测得到的 2022 年各个省份的指标值,运用训练好的养老床位需求 BP 神经网络,得到最终 2022 年 31 个省份的类别情况,见表 4。

由表 4 可知,在面对中国整体的机构养老床位需求量增加的情况下,关于床位的增加量应按照第三类、第一类、第二类的先后顺序进行分配。优先增加第三类地区的养老床位,其次增加第一类地区的养老床位,最后增加第二类地区的养老床位。

表 4 2022 年 31 个省份预测类别
Tab. 4 China's 31 provinces forecast categories in 2022

类别	地区
第一类: 床位需求适中的地区	北京、天津、山西、内蒙古、吉林、黑龙江、上海、福建、江西、湖南、广西、海南、重庆、贵州、云南、陕西、甘肃、宁夏、新疆
第二类: 床位需求最少的地区	西藏、青海
第三类: 床位需求最多的地区	河北、辽宁、江苏、浙江、安徽、山东、河南、湖北、广东、四川

3.5 最终 31 个省份得分结果

本文根据 2022 年中国 31 个省份对应的 6 个指标预测值, 采用熵权法计算各个指标的权重, 最终得出各类中每个地区的评分, 评分越高则相应地区对于床位的需求越急切。

由于运用 SPSS 软件对该组数据进行标准化时发现存在负值, 故重新采用另一种标准化方法, 保证标准化结果在 $[0, 1]$ 之间, 列出公式见如下:

$$v_{ij} = \frac{r_{ij} - \min\{r_{1j}, r_{2j}, \dots, r_{nj}\}}{\max\{r_{1j}, r_{2j}, \dots, r_{nj}\} - \min\{r_{1j}, r_{2j}, \dots, r_{nj}\}} \quad (4)$$

其中, r_{ij} 表示 2022 年 31 个省份对应的各指标预测值。

熵权法是根据指标的离散程度来分配权重, 具有较强的客观性。熵权法需先对标准化矩阵 v_{ij} 计算其对应的概率矩阵; 其次, 根据概率矩阵求各指标的信息熵和信息效用值; 最后, 将信息效用值归一化, 得到各个指标的熵权。熵权与标准化矩阵的元素进行加权求和, 即为各个样本的评分结果。评分结果是各个地区养老床位需求程度的评分, 评分越高表示该地区对养老床位的需求程度越高, 见表 5。

表 5 2022 年 31 个省份每类的评分 (每类中评分按降序)

Tab. 5 The score of each category of China's 31 provinces in 2022 (descending order)

地区	类	评分 (降序)	地区	类	评分 (降序)
上海	1	6 354 541.243	云南	1	1 624 280.336 0
黑龙江	1	6 350 648.292	海南	1	836 358.953 4
吉林	1	5 492 430.975	宁夏	1	763 617.775 9
内蒙古	1	5 007 543.647	青海	2	1 306 662.070 0
北京	1	3 951 142.797	西藏	2	320 846.249 1
湖南	1	3 773 192.999	广东	3	18 058 712.810 0
陕西	1	3 675 599.337	江苏	3	9 569 469.890 0
江西	1	3 461 902.970	山东	3	8 850 268.439 0
山西	1	3 404 386.218	浙江	3	8 604 681.313 0
重庆	1	3 221 428.765	河北	3	8 272 803.978 0
广西	1	3 001 157.107	辽宁	3	8 236 784.157 0
天津	1	2 755 742.037	四川	3	7 537 277.630 0
甘肃	1	2 736 845.717	湖北	3	7 492 084.914 0
新疆	1	2 556 571.277	安徽	3	5 850 766.242 0
福建	1	2 377 690.964	河南	3	5 577 904.748 0
贵州	1	1 923 063.806			

4 运算结果分析

本文主要采用 BP 神经网络和灰色预测模型来预测分析 2022 年 31 个省份的养老床位需求增加量的情况, 并且针对床位供需失衡的问题提出相应的解决方法:

(1) 对 2022 年的养老床位需求增加量, 应按照第三类 (需求最多)、第一类 (需求适中)、第二类 (需求最少) 地区的顺序来增加养老床位的供给, 先在整体上解决养老床位供需失衡的问题。

(2) 在各组类别中, 各地区按照熵权法的评分结果, 先分配评分高的地区, 再分配评分低的地区, 即各类中按照各个地区对于养老床位需求程度的高低进行分配。

(3) 在 2022 年, 在第三类地区中, 按照广东、江苏、山东、浙江、河北、辽宁、四川、湖北、安徽、河南的顺序依次增加养老床位供给量; 对于第一类地区, 按照上海、黑龙江、吉林、内蒙古、北京、湖南、陕西、江西、山西、重庆、广西、天津、甘肃、新疆、福建、贵州、云南、海南、宁夏的顺序依次增加床位供给量; 针对第二类地区, 则是按照青海、西藏的顺序增加床位供给量。

5 结束语

本文针对中国养老床位供需失衡的问题, 通过分类、预测、评分这 3 个步骤提出一种地区化分配养老床位的方案。首先, 运用 K 均值聚类方法将 2008~2017 年 31 个省份分为 3 类作为样本数据, 分别为养老床位需求最少、适中、最多。运用 BP 神经网络对样本数据进行训练, 同时运用判别分析对样本数据进行回判, 比较 2 个方法的精确度, 确定 BP 神经网络更适用于养老床位需求的分类预测。通过灰色预测模型和差分整合移动平均自回归模型求出 2022 年各个指标的预测值, 输入 BP 神经网络进行分类预测。运用熵权法对各类地区按养老床位的需求程度进行评分, 提出具体的养老床位地区化分配

(下转第 53 页)