

基于DNN神经网络的研究生生源预测方法研究

刁 庶^{1,2}, 蒋川东², 淳少恒¹

¹无锡职业技术学院, 江苏 无锡

²吉林大学仪器科学与电气工程学院, 吉林 长春

收稿日期: 2022年2月24日; 录用日期: 2022年3月22日; 发布日期: 2022年3月30日

摘 要

高校研究生的生源质量是提高研究生培养水平的基础, 获得准确的生源质量信息可以为招生宣传提供科学有效的背景支撑与技术保障。本文以吉林大学某学院近五年招生生源类型为样本, 建立高校研究生生源质量评价指标体系, 基于DNN神经网络拟合高校生源质量变化特点, 在谷歌研发的Tensorflow框架下建立高校研究生生源质量评价模型; 最后, 采用训练完备数据库对下一年研究生生源质量进行预测分析, 指导招生宣传工作的高效执行。本文所提预测方法较好地解决了研究生生源质量难以保障研究生培养水平的问题, 对提高高校研究生整体培养水平以及相应学科建设的发展具有重要意义。

关键词

研究生招生, 培养质量, 评价方法, DNN网络, Tensorflow框架

Research on the Method of Graduate Student Source Prediction Based on DNN Neural Network

Shu Diao^{1,2}, Chuandong Jiang², Shaoheng Chun¹

¹Wuxi Institute of Technology, Wuxi Jiangsu

²College of Instrumentation and Electrical Engineering, Jilin University, Changchun Jilin

Received: Feb. 24th, 2022; accepted: Mar. 22nd, 2022; published: Mar. 30th, 2022

Abstract

The quality of graduate students is the basis and guarantee to improve the quality of graduate training. Obtaining accurate quality information of graduate students can provide scientific and

effective technical support for enrollment promotion. In this paper, the quality evaluation index system of graduate students in colleges and universities is established based on the types of enrollment students in a college of Jilin University in the past five years; then DNN network is introduced to fit the characteristics of the quality change of graduate students in colleges and universities, and the quality evaluation model of graduate students in colleges and universities is established under the Tensorflow framework developed by Google; finally, the data of seven years of training is used to predict the quality of graduate students in the next year. The prediction method proposed in this paper can better solve the problem that the quality of graduate students is difficult to guarantee the level of graduate education, and has certain guiding significance for the follow-up work.

Keywords

Postgraduate Enrollment, Training Quality, Evaluation Method, DNN Network, Tensorflow Framework

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

研究生培养质量是建设一流和高水平研究型大学的关键问题。党的十八大以来，国家全面实施创新驱动发展战略，突出把人才作为创新的第一资源。习近平总书记强调，创新驱动实质上是人才驱动。因此对创新型人才的培养是国家兴旺发达的关键所在。研究生作为创新型人才的主力军，肩负着发展国家科学技术的重任。因而研究生教育质量成为能否为国家输送高层次创新型人才的中心环节[1]。随着研究生招生体制和培养体制的改革，逐步实现从规模化发展向内涵式发展的转变。研究生的生源质量是研究生培养质量的基础，研究生招生工作得到提升，研究生生源质量则提高，直接影响研究生的培养质量[2] [3]。针对目前研究生招生数量大量扩张的情况，研究生生源质量不能得以保障。教育管理的重点：如何适时、适宜地抓好并推进研究生的培养工作，提高研究生的培养质量，已经成为摆在我们面前的首要问题[4] [5]。科学的、合理的、有效的前期准备很大程度上决定了招生单位获取生源的质量，进而影响着整个硕士研究生培养工作的开展。为了更好地提升研究生培养质量，必须加强研究生生源质量评估体系建设。

针对现有生源质量评估办法难以定量分析，为后续招生工作提供数据支撑，本文提出了 DNN (Deep Neural Network)神经网络方法。DNN 网络是数据挖掘技术的重要手段，可以根据大量的数据分析及推理的有效工具，具有很强的实用性，已在数据分析实际中得到广泛应用[5] [6] [7] [8]。研究生生源质量的评价，存在生源质量影响相互制约的关系，不确定分析地推演其中联系，构建合理的不确定性框架，并进行数据训练分析计算，DNN 为生源质量的影响因素的关系建立模型，DNN 神经网络数据分析技术为研究生生源质量的评价与预测提供了良好的技术支撑，因此，本文以吉林大学研究生招生的历史数据为基础，首先从 DNN 网路入手，在谷歌的 Tensorflow 框架上搭建网络并选择合适的优化器来训练网络，进一步通过 2013 年至 2019 年的研究生生源数据来训练优化网络权值与偏置，预测未来，根据招生生源数据预测研究生生源质量，为招生工作提供技术支持，进一步为提升研究生培养质量提供保障。

2. 基于 DNN 网络的预测模型搭建

首先, 通过简要分析数据集的相关影响参数可以初步判断网络的特征输入的个数。进一步地, 根据需要的预测结果确定网络的输出个数。最后, 针对 DNN 网络, 设置其内部隐藏层的层数, 用已知的数据集训练 DNN 网络, 选择合适的梯度下降算法, 优化网络内部的权值、偏置, 使网络训练后得到的输出接近已知数据集。在得到训练后的 DNN 网络模型后, 即预测模型搭建完成, 将得到的特征输入到预测模型中, 得到的网络输出就是预测结果。DNN 网络是在已有基本神经网络结构的基础上, 增加了隐藏层中包含的神经元层数, 在增加了网络权值与偏置个数的条件下, 虽然计算速度有所下降, 但网络模型的预测结果会更加准确。

2.1. DNN 网络基本单元

DNN 网络隐藏层中的每一层都由多个神经元构成, 如下图 1 所示, 以单个神经元模型(也称为感知机模型)为例, 假设它有若干个输入和一个输出。

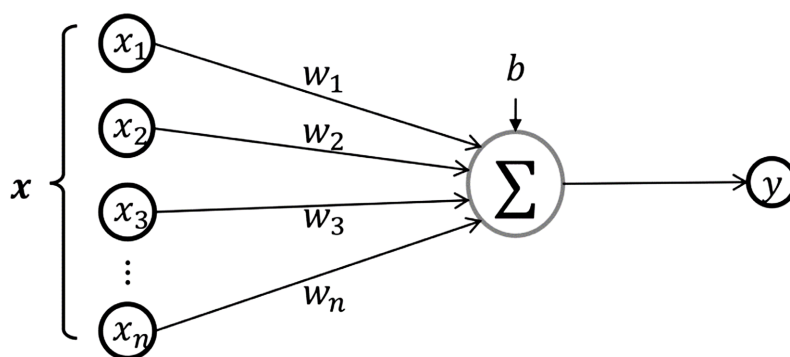


Figure 1. Single neuron linear model

图 1. 单个神经元线性模型

其中间输出结果 z 为:

$$z = \sum_{i=1}^n w_i x_i + b \quad (1)$$

接着, 激活函数得到神经元输出 y 为:

$$y = f_{active}(z) \quad (2)$$

因此对于一个基本神经元的计算, 首先是权重、偏置与输入的线性计算得到中间结果 z , 然后再通过激活函数 f_{active} 求得神经元最终的输出[9] [10] [11]。

2.2. DNN 网络层结构

从 DNN 按不同层的位置划分, DNN 内部的神经网络层可以分为三类, 输入层, 隐藏层和输出层, 如下图 2 示例, 第一层是输入层(Input Layer), 最后一层是输出层(Output Layer), 其余都是隐藏层(Hidden Layer)。因此对于 DNN 网络来说, 其隐藏层中包含很多基本神经元模型可以有很多层, 层与层之间是全连接的, 即第 i 层的任意一个神经元一定与第 $i+1$ 层的任意一个神经元相连[12] [13] [14]。虽然 DNN 看起来很复杂, 但是从小的局部模型来说, 还是和上述的感知机模型一样, 即一个中间结果 z 加上一个激活函数 $f_{active}(z)$ 。

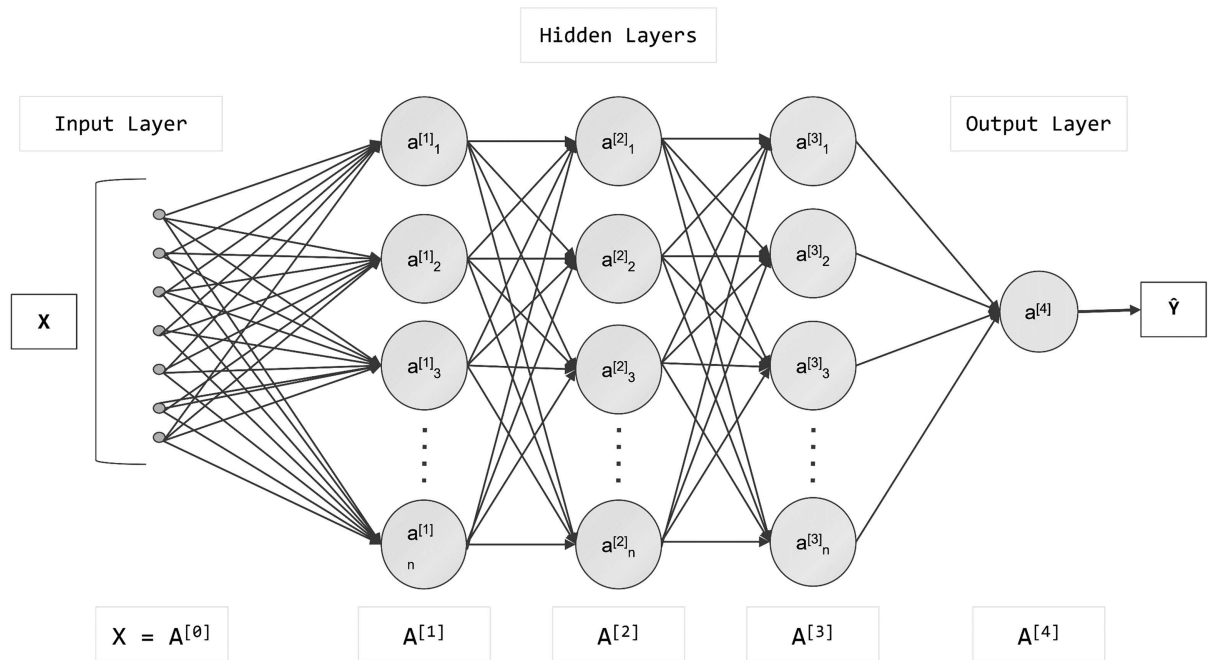


Figure 2. DNN network model
图 2. DNN 网络模型

2.3. DNN 网络常用激活函数

1) Sigmoid 函数

Sigmoid 是常用的非线性的激活函数，它的数学形式如下：

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

Sigmoid 几何图像如下图 3：

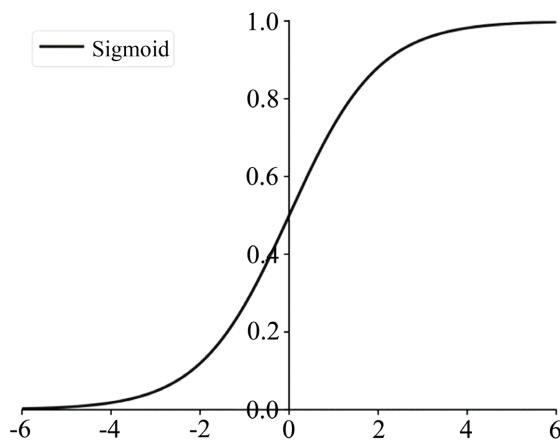


Figure 3. Sigmoid function curve
图 3. Sigmoid 函数曲线

它的一个优良特性就是能够把 $z \in R$ 的输入“压缩”到 $z \in (0,1)$ 区间，这个区间的数值在神经网络中常用来将网络的中间结果转换为概率分布(0, 1)区间的输出，即用来解决二分类问题。

2) ReLu 函数

为了避免使用 Sigmoid 而出现的梯度弥散现象(即网络参数长时间得不到更新, 导致训练不收敛或停滞不动的现象发生), 引入 ReLu 激活函数能很好地避免这一情况发生, ReLu 函数定义为(如图 4):

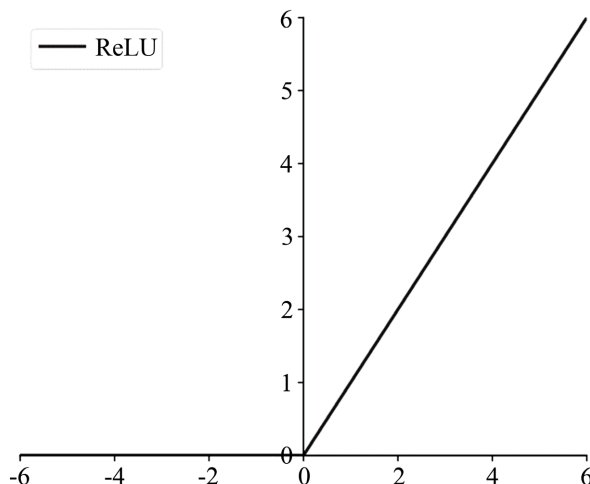


Figure 4. ReLu function curve
图 4. ReLu 函数曲线

对于输出结果为某段或者整个实数空间, 例如房价预测、函数趋势预测等问题, 常用 ReLu 函数作为激活函数[15] [16] [17]。

2.4. 基于 Adam 梯度下降的 DNN 参数优化方法

假设网络预测输出为 V_i , 而实际数据值为 M_i , 则不可避免的预测输出与实际数据会有损失 $loss$, 在回归预测问题中, 损失函数常用均方差如下定义:

$$loss = \frac{1}{N} \sum_{i=1}^N (V_i - M_i)^2 \quad (5)$$

由于网络预测输出为 V_i 与其内部权值、偏置、激活函数相关, 因此损失函数也与网络内部的参数相关, 可以表示为 $loss\{w_i^l, b_i^l, f_i^l(z)\}$, 其中 $w_i^l, b_i^l, f_i^l(z)$ 分别代表第 l 层第 i 个神经元的权值、偏置以及激活函数[18] [19] [20]。在回归问题中的激活函数 $f_i^l(z)$ 默认为 ReLu 函数, 因此, 可以表示损失函数为 $loss\{w_i^l, b_i^l\}$ 。

训练网络的目的是为了使损失函数最小, 通过梯度下降算法更新网络权值, 目前 Adam 梯度下降算法在实际应用中效果良好, 超过了其他的自适应技术。具体的 Adam 迭代更新公式为:

$$v = \beta_1 v + (1 - \beta_1) dw \quad (6)$$

$$s = \beta_2 s + (1 - \beta_2) dw^2 \quad (7)$$

$$w = w - \alpha \frac{v}{\sqrt{s + \varepsilon}} \quad (8)$$

其中规定 $\beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10^{-8}$, w 代表带更新的数值, v, s 是在梯度更新过程中的中间变量。Adam 算法相当于先把原始梯度做一个指数加权平均, 再做一次归一化处理, 然后再更新梯度值。应用此方法可以同时更新网络的权值与偏置。

3. 基于 DNN 神经网络的研究生生源质量预测

3.1. 训练数据

如图 5、表 1 所示，收集了 2013 至 2019 学年的数据作为 DNN 网络训练数据，其中特征为年份与每年的生源总人数，对应特征输出就是通过不同招生方式选拔而来的研究生人数。通过这部分数据集对搭建好的 DNN 预测网络模型，实现其预测未来各项招生计划下的人数。

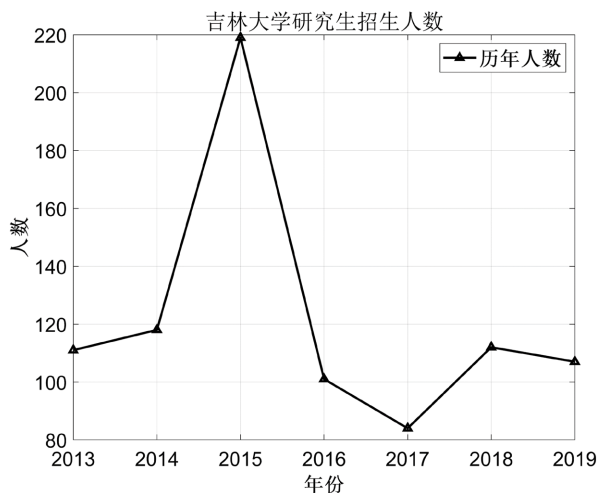


Figure 5. Quality of graduate students in a college of Jilin University in 2013~2019

图 5. 吉林大学某学院 2013~2019 学年研究生生源质量

Table 1. Number of graduate students in various enrollment plans of Jilin University

表 1. 吉林大学研究生各项招生计划人数

年份	全国统考	普通招考	硕博连读	免试推荐	强军计划
2013	68	8	11	22	2
2014	63	23	11	19	2
2015	66	18	16	119	0
2016	57	13	6	24	1
2017	69	0	8	6	1
2018	67	1	21	20	3
2019	55	18	8	25	1

3.2. DNN 网络建立

由数据可知，其特征为年份和每年招生总人数，特征输出为 5 种招生方式的研究生人数。因此，在这里得到 DNN 网络的输入层为 2，输出层为 5。设置中间隐藏层为 2 层。如下图 6 所示。其中，隐藏层的第一层设置为 3 个神经元，第二层设置为 4 个神经元。

3.3. 网络激活函数选择与梯度优化

对于研究生生源质量进行预测的问题，可以归结于回归预测类问题，因此选择激活函数为 ReLu，应

用 Adam 梯度下降算法对网络权值、偏置进行更新。

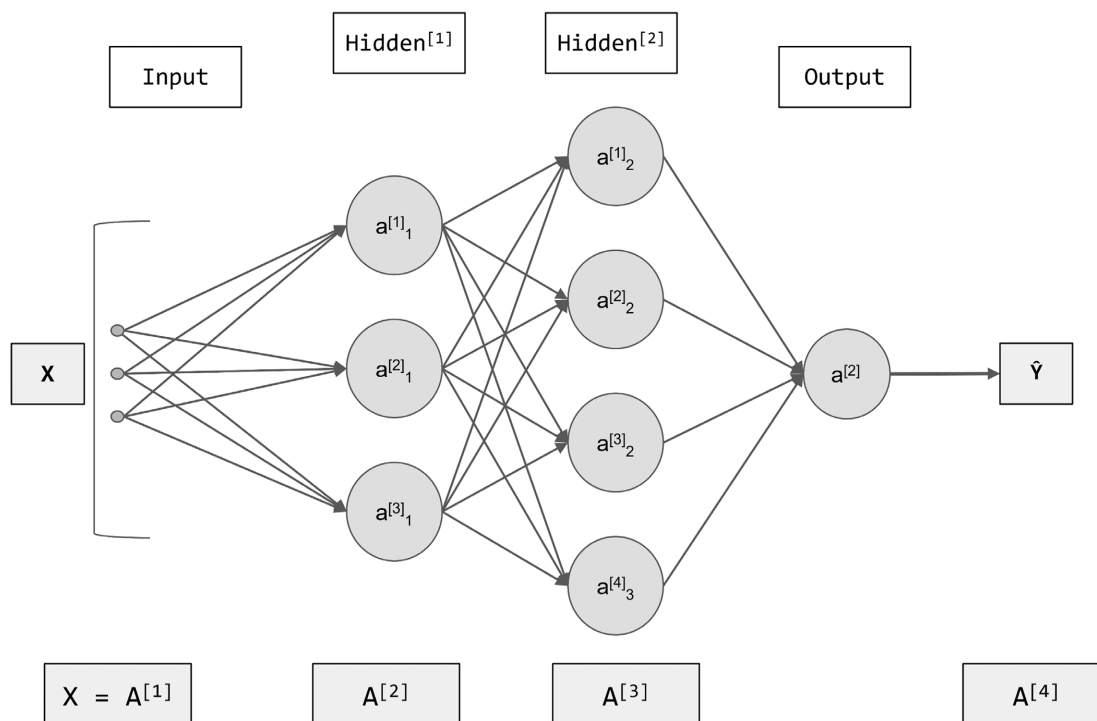


Figure 6. DNN network for graduate student source quality prediction
图 6. 研究生生源质量预测 DNN 网络

3.4. 多损失函数定义

设对于每种招生方式下的研究生损失函数为 $loss_j$ ，则有

$$loss_j = \frac{1}{y_j} \sum_{i=1}^N (y_i'' - y_i^j)^2 \quad (9)$$

其中， $j=1,2,3,4,5$ 分别代表 5 种招生方式； $i=2013, \dots, 2019$ 代表年份； $\overline{y_j}$ 代表第 j 种招生方式下，所有年份招生人数总和的平均值； y_i^j 代表第 j 种招生方式下第 i 年的实际招生人数； y_i'' 代表第 j 种招生方式下第 i 年的网络预测招生人数。

利用 Tensorflow 的多线程模式，开启 5 个会话，利用 Adam 梯度下降方法，设置学习率 $rate=0.005$ ，并采用打乱样本顺序，重复训练的方式训练 DNN 预测网络，直到使得以上 5 个损失函数值都达到 10^{-5} ，完成网络训练。

4. 仿真结果

4.1. DNN 网络验证

在训练已经达到目标损失值后，将训练数据的特征作为验证输入到训练好的网络中，得到验证输出，与实际结果对比，可以看到网络的输出精度很高，网络验证精度达到 99.4%。

4.2. DNN 网络预测

针对本文所搭建的 DNN 预测网络模型，需要得到当年份，即 2020 年份下的报考总人数，根据表 3

每一年份下吉林大学某学院研究生报名总人数, 对这组 2013 至 2019 学年的招生总人数进行回归预测得到 2020 年份预计招生总人数为 110 人(详见表 2)。

Table 2. Verification results
表 2. 验证结果

年份	预测	实际	预测	实际	预测	实际	预测	实际	预测	实际
年份	全国统考		普通招考		硕博连读		免试推荐		强军计划	
2013	67.930	68	8.135	8	11.001	11	22.294	22	1.990	2
2014	62.996	63	23.005	23	10.998	11	18.969	19	2.000	2
2015	65.836	66	18.252	18	15.952	16	118.821	119	-0.006	0
2016	68.957	57	13.381	13	6.142	6	24.222	24	1.559	1
2017	68.995	69	-0.006	0	7.992	8	6.056	6	0.999	1
2018	66.975	67	1.048	1	20.973	21	19.931	20	2.995	3
2019	57.679	55	17.926	18	8.873	8	24.725	25	1.163	1

Table 3. Total number of graduate students enrolled in Jilin University in each year
表 3. 吉林大学每一年份报考研究生总人数

年份	全国统考	普通招考	硕博连读	免试推荐	强军计划	总人数
2013	68	8	11	22	2	111
2014	63	23	11	19	2	118
2015	66	18	16	119	0	219
2016	57	13	6	24	1	101
2017	69	0	8	6	1	84
2018	67	1	21	20	3	112
2019	55	18	8	25	1	107

将 2020 年份预计招生总人数, 110 人作为网络输入, 得到每项计划下的招生预测人数如表 3 所示。

Table 4. DNN network predicts the enrollment results under various plans in 2020
表 4. DNN 网络预测 2020 年份各项计划下的招生结果

年份	全国统考	普通招考	硕博连读	免试推荐	强军计划
2020	69.052	-0.076	19.550	16.284	2.807

结合网络验证、预测结果如图 7 所示分析 2013 至 2020 年各项计划招生以及招生总人数不难发现自 2016 年以后的研究生招生总人数逐渐趋于稳定, 各项计划下的招生人数也趋于稳定, 这不仅标志着国家制度逐渐完善, 也验证了吉林大学研究生招生名额变化幅度的降低, 稳固保持在每年招收 100~115 名研究生。基于这一预测信息, 能够便于吉林大学 2020 年研究生各项招生计划的开展。

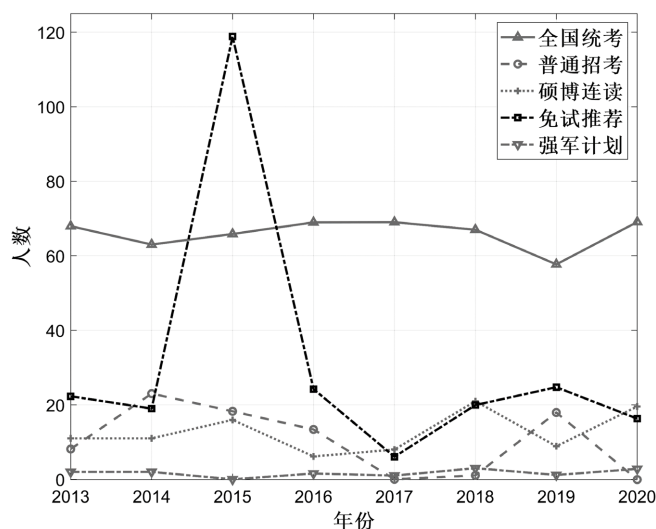


Figure 7. Verification and prediction results of DNN network model
图 7. DNN 网络模型验证及预测结果

5. 结论

通过 DNN 网络的验证输出与实际数据值, 可知网络精度较高, 可以满足预测吉林大学某学院 2020 年研究生生源质量的需要, 而且随着时间推移, 已有数据的增多, 可以将训练样本增多, 重新训练网络权值, 以实现精度更高的研究生生源质量的预测。通过 DNN 神经网络的方法可以对研究生生源质量进行科学有效地预测, 进一步为研究生招生宣传方案提供理论依据, 以提高研究生培养质量。

基金项目

国家自然科学基金青年项目“隧道磁共振旋转探测水害隐患定位成像方法研究”(42104142), 吉林省自然科学基金项目“基于时移核磁共振 4D 成像的地下水动态监测方法与仪器”(20190201111JC); 吉林大学本科教学改革研究项目“金课”视角下“误差理论与数据处理”的教学改革与深化(2019XYB223); 卓越工程师背景下工科学生招生就业创业能力培养机制研究(2019zsyb021)。

参考文献

- [1] 李小华, 肖拥军, 张振华, 郑波. 高校研究生招生和培养模式改革的探讨[J]. 教育教学论坛, 2020(37): 89-90.
- [2] 廖湘阳. 研究生教育发展战略研究[M]. 北京: 清华大学出版社, 2006.
- [3] 龚姚腾, 陈龙泉, 卢致杰. 硕士研究生招生过程预测模型研究[J]. 科技进步与对策, 2006(4): 159-161.
- [4] 刘军. BP 神经网络在教学质量评价体系中的应用[J]. 中国电子商务, 2009(9): 155.
- [5] 黎玺克. 遗传算法优化 BP 神经网络的岩质边坡稳定性预测[J]. 河北工业科技, 2016(1): 1-7.
- [6] 孔一斐, 高明国. 地方高校硕士研究生复试评价体系的改革与探索[J]. 创新创业理论与实践, 2020, 3(24): 75-77.
- [7] Wang, R.Q., Jiang, Y.L. and Lou, J.G. (2020) TDR: Two-Stage Deep Recommendation Model Based on mSDA and DNN. *Expert Systems with Applications*, **145**, Article ID: 113116. <https://doi.org/10.1016/j.eswa.2019.113116>
- [8] Sparsh, M. (2020) A Survey on Modeling and Improving Reliability of DNN Algorithms and Accelerators. *Journal of Systems Architecture*, **104**, Article ID: 101689. <https://doi.org/10.1016/j.sysarc.2019.101689>
- [9] 高迎明. 基于 DNN 的发音偏误趋势检测[C]//第十三届全国人机语音通讯学术会议(NCMMSC2015)论文集. 北京: 中国中文信息学会语音信息专业委员会: 清华信息科学与技术国家实验室(筹), 2015: 366-370.
- [10] 赵跃波. 几种改进的动态神经网络股指预测模型研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2016.

- [11] 宋博文. 基于深度残差网络的股票价格趋势预测模型研究[D]: [硕士学位论文]. 厦门: 厦门大学, 2018.
- [12] 李斌, 朱杰, 马志贤. WebRTC 中一种基于 DNN 的噪声抑制算法的研究[J]. 信息技术, 2019, 43(5): 1-5.
- [13] Halley, S. (2020) Pandemic Opportunities Lead to Largest, Most Diverse Graduate Enrollment Yet. *Recruiting & Retaining Adult Learners*, **23**, 12. <https://doi.org/10.1002/nsr.30679>
- [14] Hope, J. (2019) Engage International Grad Students, Alumni to Support Recruitment Goals. *Enrollment Management Report*, **23**, 6-7. <https://doi.org/10.1002/emt.30592>
- [15] Esq, A.G. (2019) Focus on Branding, Education to Improve Graduate Recruitment Efforts. *Recruiting & Retaining Adult Learners*, **21**, 12. <https://doi.org/10.1002/nsr.30507>
- [16] Krausman, P.R. (2021) An Open Letter to Wildlife Graduate Students: Ideas to Consider When Publishing. *Journal of Wildlife Management*, **85**, 193-194. <https://doi.org/10.1002/jwmg.21992>
- [17] 李桂民. “一流学科”建设背景下研究生学术创新能力的培养[J]. 临沂大学学报, 2021, 43(1): 106-114.
- [18] 姜俊辉. 新时代研究生培养教育的三点思考[J]. 中国研究生, 2020(12): 30-31.
- [19] 冯艺佳, 王兰珍, 路燕. 互联网时代研究生招生宣传的实践与分析[J]. 高教论坛, 2020(10): 71-76.
- [20] 郭巍. 硕士研究生生源质量现状及提升措施研究[J]. 沈阳建筑大学学报(社会科学版), 2020, 22(2): 201-206.