

大学生就业类型预测模型构建和对学生就业工作的启示

谭贞晶^{1,2}, 何咖宙¹, 刘欢婷¹

¹广东医科大学公共卫生学院, 广东 东莞

²华南师范大学教育科学学院, 广东 广州

收稿日期: 2025年10月19日; 录用日期: 2025年11月12日; 发布日期: 2025年11月19日

摘要

挖掘大学生就业数据隐含信息有助于实现大学生就业服务的精准化、信息化。本研究采用标准BP神经网络和主成分分析改进的BP神经网络两种方法, 探究大学生基本信息与就业类型之间关系, 建立大学生就业类型预测模型。大学生就业类型预测模型结果表明: (1) 标准BP神经网络能建立精度良好的大学生就业类型预测模型; (2) 主成分分析改进的BP神经网络模型能提高大学生就业类型预测模型的精确度; (3) 在提取的主成分中, 大学生的学业成绩、户籍和住址、专业类别等因素对就业类型预测模型的贡献率最高。研究结果对大学生就业工作有如下启示: (1) 建立全面就业工作数据系统, 提取就业预测的关键因素和核心指标; (2) 挖掘就业数据信息价值, 建立科学就业信息分析系统; (3) 构建有效的就业预测模型, 构建分层分类的精准化就业指导体系。

关键词

BP神经网络, 主成分分析, 大学生, 就业类型, 预测模型

Construction of a Prediction Model for College Students' Employment Types and Its Implications for Students Employment Work

Zhenjing Tan^{1,2}, Kazhou He¹, Huanting Liu¹

¹School of Public Health, Guangdong Medical University, Dongguan Guangdong

²School of Education, South China Normal University, Guangzhou Guangdong

Received: October 19, 2025; accepted: November 12, 2025; published: November 19, 2025

文章引用: 谭贞晶, 何咖宙, 刘欢婷. 大学生就业类型预测模型构建和对学生就业工作的启示[J]. 服务科学和管理, 2025, 14(6): 838-845. DOI: 10.12677/ssm.2025.146105

Abstract

Mining the latent information embedded in college students' employment data is instrumental in achieving precision and informatization in college students' employment services. This study utilizes two approaches: the standard BP neural network and the BP neural network enhanced through principal component analysis, to delve into the correlation between college students' basic information and their employment types, and to construct a predictive model for college students' employment types. The results of the prediction model for college students' employment types indicate that: (1) the standard BP neural network is capable of developing a predictive model for college students' employment types with commendable accuracy; (2) the BP neural network model refined by principal component analysis can further elevate the precision of the predictive model for college students' employment types; (3) among the principal components extracted, factors including college students' academic achievements, registered residence, address, and major classification exhibit the highest contribution rates to the predictive model for employment types. The research findings offer the following insights for enhancing college students' employment services: (1) establish a comprehensive employment work data system to extract key factors and core indicators for employment prediction; (2) mine the value of employment data information and establish a scientific employment information analysis system; (3) construct an effective employment prediction model and build a hierarchical and classified precision employment guidance system.

Keywords

BP Neural Network, Principal Component Analysis, College Students, Types of Employment, Prediction Model

Copyright © 2025 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

大学生就业问题是重要的社会、经济和民生问题。近十多年来,随着我国高等教育规模的扩大,每年毕业的大学生人数不断增加。2025 年的毕业生规模已经达到了 1222 万[1]。庞大的毕业生群体带来了巨大的新增就业人口压力,使得就业问题成为备受高校、企业和政府等多方关注的问题。高效利用就业数据对大学生的就业管理和服务具有重要的促进作用。为了做好高校毕业生就业工作,国家有关部门先后出台了相关政策和指导意见。2016 年 4 月,教育部发布《关于开展全国普通高校毕业生精准就业服务工作的通知》,要求各地各高校建立健全精准推送就业服务机制,精准推送相应的就业政策、岗位信息、指导服务,实现就业服务个性化、差异化[2]。2017 年 12 月,教育部出台了《关于推动高校形成就业与招生计划人才培养联动机制的指导意见》,指出要充分利用“互联网 + 就业”新模式,准确把握毕业生求职意愿和用人单位岗位需求信息,建立精准推送就业服务机制[3]。2018 年 6 月发布的《教育统计管理规定》要求建立教育统计数据解读、预测预警机制,鼓励深入挖掘数据资源,综合运用多种统计分析方法,提高统计分析和应用能力[4]。

就业服务精准化、个性化和信息化,智能化匹配关键信息,实现“一生一策”动态管理,离不开对大学生就业深度隐含信息数据的挖掘。BP 神经网络具有较强的非线性映射能力、高度自学习和自适应的能力、泛化能力、容错能力,而且结构简单,其预测输出能不断逼近期望输出,被广泛应用于模式识别、数

据挖掘、智能控制等领域。有研究通过灰色系统对大学生就业率变化特点进行拟合，建立大学生就业率预测模型，并分析了该人工神经网络方法的有效性和优越性[5]。而采用灰色模型和神经网络分别对不同大学生就业数量进行建模和预测，能获得比较理想的大学生就业预测结果[6]。有研究运用遗传算法优化的 BP 神经网络技术计量研究高校学生的创业意愿与创业行为间的关系，揭示了大学生创业意愿与创业行为的影响因素[7]。还有学者通过层次分析法对各指标赋权，利用 BP 神经网络数据分析技术，模拟专家的知识与经验对高校毕业生就业质量进行评价，评价结果误差在 5%以内，表明该网络有一定的可靠性[8]。这些研究都表明，神经网络建立的就业预测模型具有较高的预测精度，为做好大学生就业服务工作提供了一种全新的思路和方法。

大学生就业去向类型的识别从本质上属于模式识别或者非线性分类问题。每个大学生作为一个独立个体，其就业信息构成可以视作一个多维数据系统。就业情况都是多维的信息系统，其基本特征是多变量、多层次、强耦合，系统内部各因素存在复杂的非线性相互作用。

本文将运用 BP 神经网络和主成分降维改进的 BP 神经网络探讨大学生基本信息数据与就业类型之间存在的关系或规则，挖掘大学毕业生就业分类的基础规律，建立大学生就业类型预测模型，为毕业生就业指导提供依据和参考，为做好大学生就业服务提出建议，并为政策制定、岗位信息精准对接服务提供支持。

2. 方法与原理

2.1. BP 神经网络

BP (Back Propagation)神经网络是一种基于误差反向传播算法训练的多层前馈神经网络。BP 神经网络能学习和储存大量的输入-输出模式映射关系，在前向传递中，输入信号从输入层经隐含层逐层处理，直到输出层每一层的神经元状态只影响下一层神经元状态[9]。它的学习规则是使用最速下降法，根据预测误差调整网络权值和阈值，从而使 BP 神经网络预测输出不断逼近期望输出。

BP 神经网络前向输入的传导过程如下：

网络隐含层第 i 个节点的输入 α_i ：

$$\alpha_i = \sum_{j=1}^M w_{ij} x_{ij} + b_i$$

网络隐含层第 j 个节点的输出 z_j ：

$$Z_j = h(\alpha_i) = h\left(\sum_{j=1}^M w_{ij} X_{ij} + b_i\right)$$

网络输出层第 j 个节点的输入 a_j ：

$$a_j = \sum_{k=1}^q W_{jk} Z_i + b_j = \sum_{k=1}^q W_{jk} h\left(\sum_{j=1}^M w_{ij} X_{ij} + b_i\right) + b_j$$

网络输出层第 k 个节点的输出 y_j ：

$$y_j = h\left(\sum_{k=1}^q W_{jk} Z_i + b_j\right) = h\left(\sum_{k=1}^q W_{jk} h\left(\sum_{j=1}^M w_{ij} X_{ij} + b_i\right) + b_j\right)$$

BP 神经网络的输入误差的反向传播过程首先从网络的输出层开始逐层计算各层神经元的输出误差，运用误差梯度下降法来调整各层的权值和阈值，使修改后的网络的最终输出能接近期望值。

对应误差纠正函数:

$$E_j = \frac{1}{2M} \sum_{j=1}^q (yn_j - y_j)^2$$

误差纠正就是使基于 yn_j 的目标函数达到最小值, 从而使神经网络中每个输出神经元的实际输出在统计意义上最接近样本应该的输出。

系统对 P 个训练样本的总误差的准则函数为

$$h = \sum_{j=1}^q \frac{1}{1 + e^{-y_j}} \frac{\partial h}{\partial y_j} = \sum_{j=1}^q \frac{e^{-y_j}}{1 + e^{-y_j}}$$

$$h = \sum_{i=1}^M \frac{1}{1 + e^{-z_i}} \frac{\partial h}{\partial z_i} = \sum_{i=1}^M \frac{e^{-z_i}}{1 + e^{-z_i}}$$

根据误差梯度下降法依次调节输出层各权值的修正量 Δw_{ij} , 输出层各阈值的修正量 Δb_i , 隐含层各权值的修正量 Δw_{jk} , 隐含层各阈值的修正量 Δb_j

$$\Delta w_{ij} = \sum_{i=1}^M -\eta E'_j * \frac{\partial h(z_i)}{\partial z_i} * \frac{\partial z_i}{\partial w_{ij}} \quad \Delta w_{jk} = \sum_{k=1}^q -\eta E'_j * \frac{\partial h(y_j)}{\partial y_j} * \frac{\partial y_j}{\partial w_{jk}}$$

$$\Delta b_i = \sum_{i=1}^M -\eta E'_j * \frac{\partial h(z_i)}{\partial z_i} * \frac{\partial z_i}{\partial b_i} \quad \Delta b_j = \sum_{j=1}^q -\eta E'_j * \frac{\partial h(y_j)}{\partial y_j} * \frac{\partial y_j}{\partial b_j}$$

2.2. 主成分分析

主成分分析(Principal Component Analysis, PCA)是一种将多个变量综合成较少的、相互独立且集中反映原始变量较多信息的若干个综合主成分的技术, 它可将高维空间中的多变量问题转化到低维空间中, 形成新的少数的变量(综合变量), 利用这些新变量代替原来变量进行后续处理, 而这种转化需要满足新得到的变量是原变量的一个线性组合[10]。这种做法既能降低多变量数据系统的维度, 又可对系统变量的统计数字特征进行简化, 从而加快了 BP 神经网络收敛速度, 有利于提高预测精度。

由于本研究所采取的预测因子是高维数据, 而高维的数据输入会使得神经网络结构复杂。因为因素间的共线性作用和预测变量贡献率小的因子的影响会降低神经网络预测的精确性。为了剔除和类标签无关的特征, 不能简单地从 n 维特征中去除其余 $n-k$ 维特征, 而需要一种特征降维的方法来减少特征数, 减少噪音和冗余, 减少过度拟合的可能性。

主成分分析的算法步骤是: 计算样本数据集 X 中样本的均值向量 μ , 即 $\mu = \frac{1}{n} \sum_{i=1}^n x_i$; 对每个样本去均值, 即将样本数据中心化, 即 $\tilde{X} = X - \mu$; 构造数据矩阵的协方差矩阵 V , $V = \frac{1}{n} \tilde{X} \tilde{X}^T$; 对矩阵 V 进行特征分解, 求取特征值 λ_i 和对应的特征向量 w_i , 降序排列特征值 λ_i ; 根据贡献率的大小, 取前 d 个特征值 $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_d]$ 和相应的特征向量 $W_d = [w_1, w_2, \dots, w_d]$ 作为子空间的基, 那么所要提取的 d 个主成分为 $F = W_d^T \tilde{X}$; 由所提取的主成分重建原数据 $X = WF + \mu$ 。

3. 研究方法

本研究将大学毕业生专业、性别、民族、户籍、住址、政治面貌、学业成绩、挂科门数、奖惩情况、父母单位、是否留级、是否担任学生干部、是否贫困生等 24 项变量作为 BP 神经网络的输入单元, 把就业去向指标作为输出单元, 构建三层 BP 神经网络预测模型。

3.1. 样本与工具

通过问卷调查方式采集到 467 份大学毕业生个人及就业信息数据样本。在构建 BP 神经网络时，将 300 份样本作为训练集，167 份样本作为测试集。采用 Matlab2016 和 IBM SPSS Statistics 25 软件工具进行数据处理。

3.2. 模型建立

实验分传统 BP 神经网络训练和主成分分析后的 BP 神经网络训练两部分进行。两种训练方法均构建一个三层神经网络，从输入层到隐含层的传递函数为“tansig”，隐含层到输出层的传递函数为“purelin”，选择梯度下降法的学习算法，设定误差目标为学习率 $lr = 0.01$ ，误差目标 $goal = 0.0001$ ，迭代 10 次。

基于 BP 神经网络的大学生就业去向分类算法建模包括 BP 神经网络构建、BP 神经网络训练和 BP 神经网络分类三步。而用主成分分析改进 BP 神经网络训练的步骤是：将主成分分析法与 BP 神经网络算法结合步骤：对原始数据进行主成分分析；将主成分分析后处理好的数据作为 BP 神经网络的输入节点，对数据进行 BP 神经网络预测；对预测结果进行误差分析，误差较大则重新放入 BP 神经网络结构中再进行迭代。

4. 数据结果和分析

4.1. 标准 BP 神经网络模拟仿真预测精度

如表 1 所示，利用 BP 神经网络进行大学毕业生就业去向预测，正确率较高。

Table 1. Accuracy (%) of the employment destination prediction model constructed using a BP neural network
表 1. BP 神经网络构建就业去向预测模型正确率(%)

就业类别	未就业	对口就业	不对口就业	升学	公务员	创业
正确率	70	72.31	87.50	69.23	80	100

4.2. PCA-BP 神经网络模拟仿真预测精度

通过主成分分析，提取输入变量中特征值大于 1 的 8 个主成分。8 大主成分对大学毕业生就业去向预测贡献率如表 2。大学生在校期间的学业成绩、户籍和住址这两项对大学生就业去向预测模型的贡献率最高。

Table 2. Contribution rates (%) of principal components with eigenvalues greater than 1 to the prediction of college graduates' employment destinations
表 2. 特征值大于 1 的主成分对大学毕业生就业去向预测贡献率(%)

序号	主成分	贡献率	累计贡献率
1	学业成绩	18.878	18.878
2	户籍和住址	10.721	29.599
3	专业类别	7.833	37.433
4	补考门数	6.908	44.341
5	父母工作单位	5.435	49.776
6	奖惩情况	4.930	54.706
7	担任学生干部情况	4.759	59.465
8	政治面貌	4.372	63.837

如表 3、表 4 所示，主成分分析减少输入变量后的 BP 神经网络对就业去向预测模型的输入变量减少，隐含层减少，预测准确度有所提升。可见充分利用主成分分析的优势，能有效解决高维数据的“维数灾难”，该方法将原始复杂、冗余且可能存在非线性关系的变量，转化为少数几个彼此正交、线性无关的主成分，而这些主成分对 BP 神经网络预测变量有着更好的预测效果。由此解决了高维数据造成的 BP 神经网络过拟合情况，避免了陷入局部最优点的情况。

Table 3. Accuracy (%) of the employment destination prediction model constructed using a BP neural network after dimensionality reduction via principal component analysis

表 3.主成分分析降维后 BP 神经网络构建就业去向预测模型正确率(%)

就业类别	未就业	对口就业	不对口就业	升学	公务员	创业
正确率	72.73	72.80	69.23	91.67	100	100

Table 4. Comparison of prediction models: standard BP neural network vs. PCA-BP neural network

表 4. 标准 BP 神经网络和 PCA-BP 神经网络预测模型对比

神经网络类型	输入层	隐含层	输出层
标准 BP 神经网络	23	24	6
PCA-BP 神经网络	8	9	6

5. 讨论和启示

我们通过 BP 神经网络模型构建了一个由大学毕业生专业、性别、户籍、学业成绩等作为预测指标的就业预测模型。结果发现，通过主成分分析改进后的 BP 神经网络能够较好地预测大学生就业去向。这一结果对于我们做好大学毕业生就业的工作有所启示。

5.1. 建立全面就业工作数据系统，提取就业预测的关键因素和核心指标

我们建立的 BP 神经网络预测模型发现，通过深度挖掘大学生的数据能够有效地预测大学生就业，这为做好大学生精准就业服务提供了非常有价值的工具。根据 BP 神经网络模型的特性，模型数据样本越大，选取的输入指标越接近本质，则隐含层节点的数目对网络影响越小，神经网络比较稳定，预测效果也比较好。为了建立更加有效的模型，我们可以建立全面的就业工作数据系统，针对大学生及就业相关对象收集更多的数据集，获得更多典型和客观的指标。

在实际工作中，我们可以通过更全面地采集就业供需两方面的数据来建立就业服务工作数据集合。这两方面的数据分别是：(1) 大学生(包括应届毕业生、往届生)信息收集，包括其家庭基本情况、学业水平、求职意向、实习就业、职业心理等等；(2) 就业市场(包括行业职位、用人单位人才需求等)信息采集，真正挖掘就业信息大数据的市场价值。通过基本信息的分类采集管理，对大学生就业涉及的数据参数进行深度分析和数据建模，提取就业预测的关键因素，建立就业预测模型的核心指标体系，可以构建精准就业服务工作数据系统中枢，为大学生就业服务提供有力的数据支持。

5.2. 挖掘就业数据信息价值，建立科学就业信息分析系统

运用人工神经网络建立大学生就业的预测数学模型，不仅仅是一种处理就业数据的方法和技术，更是一种全新的就业服务价值观和方法论。相对于就业信息的数据化和信息化，就业预测模型的建立和应用，旨在进一步丰富智能时代下精准就业服务工作的内涵，深化“以学生为本”的服务理念与管理模式，构建更为高效务实的就业服务流程，提高大学生就业服务精准化水平。

深耕就业数据,根据大学生在校期间的信息较为准确地预测其未来的就业类型,对模型预测的不同就业区域、就业领域、就业类别的大学生进行筛选甄别、分类管理,让就业数据在分析中增值。这既有利于学生工作者提前预判、科学引导,就业问题前置处理,也有利于大学生结合个人条件和市场需求,对未来就业形成更理性的预期,做好相应的就业准备。

5.3. 构建有效的就业预测模型,构建分层分类的精准化就业指导体系

过去传统的就业服务常常是基于经验和直觉来开展的,有时甚至是粗放型的“一刀切”模式。通过有效的就业预测模型,我们的就业服务工作可以从被动响应转为主动干预,从“一刀切”转向分层分类的精准化就业指导。

应用大学生就业去向预测模型,对大学生的就业问题进行前瞻性思考,对就业去向进行预判,对有就业困难的学生形成预警机制,提前介入,精准帮扶;对有考研升学、考公务员和事业单位意向的学生提供相关信息推送和匹配,对于有其他就业需求的学生进行个性化、多元化、定制化的就业信息推送、创新创业指导等,以数据驱动就业服务的创新升级,推动就业工作的“供给侧结构性改革”。这既为学生提供贴合需求的就业信息,也便于保存学生信息,有助于建立长期跟踪和服务的机制,提高学生的就业质量。

模型分析发现,大学生的学业成绩和户籍地是区分和预测大学生就业类型的稳定因素,这一分析结果与就业现实吻合,也为我们的大学生就业服务工作提供了重要方向。在开展就业服务工作时可以根据模型预测结果构建分层分类的精准化就业指导体系。例如,对于学业成绩优秀的学生,就业工作指导应该从基础的求职技能培训转向高级职业发展课程,鼓励他们思考未来3~5年的职业路径,向具有更高发展的行业和岗位迈进。对学业成绩较差的学生,可以利用模型预测结果并结合教务系统的成绩系统,及早识别就业困难风险较高的学生群体,开展针对性的学业辅导和职业技能培训。而对于来自农村或偏远地区的学生,则应加强就业资源的倾斜与支持,提供更多的实习机会,帮助他们克服地域带来的信息不对称和资源匮乏问题,提升就业竞争力。

6. 研究局限与展望

本研究通过探讨利用大学生学习成绩、户籍和就业类型等数据构建就业预测模型,并讨论了结果对于大学生就业服务工作的启示。但研究存在一定的局限性:(1)本研究的样本只有467份,研究样本较少且局限于单一高校,可能带来模型过拟合的风险,同时也会影响结果的代表性和推广性,在进一步研究中,应该从不同高校收集更大样本的数据,使模型的结果更加准确和更具推广价值。(2)本研究只构建了一个就业类型预测模型,由于就业服务工作是一项系统工程,只有进一步开发完善就业服务工作的其他子系统,才能更好地对接就业模型输出的信息。下一步可以基于就业预测模型,通过技术手段对基本信息进行搜索、选取、分析加工和关键信息的智能匹配,生成相关分析报告,包括就业市场分析、毕业生实习就业、职业发展分析等,建立精准就业信息分析系统,使数据隐含信息发挥更大的智能作用,从而提高大学生就业信息的深化整合应用,提高就业管理调控能力。

参考文献

- [1] 中华人民共和国教育部. 2025届全国普通高校毕业生就业创业工作会议召开[EB/OL]. http://www.moe.gov.cn/jyb_zzjg/huodong/202411/t20241114_1162975.html, 2024-11-14.
- [2] 中华人民共和国教育部. 教育部办公厅关于开展全国普通高校毕业生精准就业服务工作的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A15/s3265/201604/t20160401_236231.html, 2016-04-01.
- [3] 中华人民共和国教育部. 教育部关于推动高校形成就业与招生计划人才培养联动机制的指导意见[EB/OL].

-
- http://www.moe.gov.cn/srcsite/A08/s7056/201801/t20180123_325312.html, 2018-01-23.
- [4] 中华人民共和国教育部. 教育统计管理规定[EB/OL].
http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/201807/t20180713_342990.html, 2018-07-13.
- [5] 亓红强, 张福堃, 高大鲲, 等. 基于灰色系统的大学生就业率预测[J]. 现代电子技术, 2019(11): 174-177.
- [6] 李想. 大学生就业的建模与预测研究[J]. 现代电子技术, 2017(21): 117-119+124.
- [7] 朱爱胜, 俞林, 许敏, 等. 大学生创业意愿与创业行为影响因素研究——基于遗传算法优化 BP 神经网络[J]. 技术经济与管理研究, 2015(9): 35-39.
- [8] 杨昱梅, 李继娜. 基于 AHP 和 BP 神经网络的高校毕业生就业质量评价研究[J]. 中国教育学刊, 2015(S1): 148-149.
- [9] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representations by Back-Propagating Errors. *Nature*, **323**, 533-536. <https://doi.org/10.1038/323533a0>
- [10] Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. *WIREs Computational Statistics*, **2**, 433-459.
<https://doi.org/10.1002/wics.101>