基于灰色预测和径向基网络的人口预测研究

徐丽丽 李 洪 李 劲
(昆明理工大学质量发展研究院 昆明 650093)

摘 要 针对经济增长和社会稳定的问题,对人口进行准确预测是极其重要的。因此,文中利用山东省历年的人口总数分别构建了灰色预测模型和径向基网络模型,对 1995－2014 年共 20 年的人口总量进行仿真模拟;并且针对单一模型的局限性问题,还利用标准差法对其预测结果进行了权重的重分配,并在其基础上构建了组合模型。结果表明:相对于灰色模型和径向基网络模型而言,组合预测模型的精度较高,并对 2015－2025 年间的人口总量利用组合模型进行了短期预测。

关键词 灰色预测,径向基网络,组合预测模型
中图法分类号 C924 **文献标识码** A

Research on Population Prediction Based on Grey Prediction and Radial Basis Function Network

XU Li-li LI Hong LI Jin
(Quality Development Institution, Kunming University of Science and Technology, Kunming 650093, China)

Abstract For the problem of economic growth and social stability, it is extremely important to accurately predict the population. Therefore, this paper used the total population of Shandong Province over the years to construct a gray prediction model and a radial basis network model, respectively, to simulate the total population of 20 years from 1995 to 2014. And for the limitation of the single model, this paper also used the standard deviation method to redistribute the weights of its forecast results, and built a combination model on the basis of it. The results show that the accuracy of the combined forecasting model is higher than that of the grey model and the radial basis network model, and a short-term forecast of the total population between 2015 and 2025 is made by using the combined forecasting model.

Keywords Grey prediction, RBF neural network, Combination forecasting methodology

科学有效的人口预测有助于合理地控制人口数量,可以促使经济的发展与资源的合理利用,并对整体规划以及可持续发展都有着重要的意义,其被关注程度也在日益加强。而地处华北地区的山东也是人口重省,截止 2013 年总人口数已达 9700 万人,虽然人口增长率有所下降,但是由于人口基数庞大,其仍然面临着众多的挑战,因此对于人口的预测研究具有重要的意义。

人口预测是以人口发展规律为预测主体来确定参数的,而这些数据的获取以及预测算法的选择在很大程度上影响着预测结果的精确性,不同模型对于人口的预测结果也不同,周志刚等[1]在对湖北省"十二五"人口数量进行分析预测时,将灰色 PGM(1,N)人口预测模型与混沌神经元网络模型的融合,更能反映人口数量及其影响因素间相互制约、相互促进、协调发展的关系,是一种可靠的人口短期预测方法。贾楠等[2]在预测全国总体人口数量时,利用 BP 神经网络建立三层神经网络结构模型,并采用自适应学习速率和附加动量法相结合的方法,对网络进行学习和训练,预测结果显示该方法是可行的。张海峰等[3]在预测西宁市人口规模时,分别构建了一元线性回归模型、马尔萨斯模型、logistic 模型及 GM(1,1)模型,并进行了模型检验,结果表明,GM(1,1)模型的预测精度最高;王宇熹等[4]结合人口精算学递推和灰色动态 GM(1,1)模型,对上海城镇养老保险人口分布进行预测,研究结论对于上海养老保险基金收支缺口测算及新人口政策制定具有重要的理论意义和应用价值。李富荣[5]以灰色预测理论为基础,在分析 GM(1,1)建模思路的基础上结合等维灰数递补方法的优点,应用改进的动态 GM(1,1)模型对中国 2013－2022 年的人口总量进行了预测;Shang 等[6]在研究人口老龄化影响的财政可持续性的定量评估时,将常规人口预测修正嵌入随机人口预测中。由此可见,在对人口数量进行预测时,有数不胜数的模型可以选择,但没有研究表明所选的算法模型是所有模型中最优的,其原因在于模型的选取和求解方法的不同导致其结果的精度不同,但是诸多研究表明单一的模型预测的结果远没有组合预测模型的精度高,如何思兰等[7]在对人口总量进行预测研究时,针对单一预测模型的局限性,采用了标准差法的组合预测模型,结果表明组合模型具有较高的精度。任强等[8]用时间序列的 ARMA 模型对未来的生育率、死亡率进行了估计,并由此构造 Leslie 矩阵,根据自相关函数、偏自相关函数的截尾性或拖尾性,实现对 ARMA 模型的识别,应用证明两者结合的人口随机预测方法是稳健的。蒋若凡等[9]在对我国人口总量进行预测分析时,针对单一指标进行人口总量预测精度不高的问题,建立了多指标灰色 PSO-BP 神经网络人口预测模型,实验表明该模型预测和外

徐丽丽(1992－),女,硕士生,主要研究方向为质量工程管理,E-mail:2436327521@qq.com;李 洪(1990－),男,硕士,主要研究方向为质量统计;李 劲(1971－),男,博士,副教授,主要研究方向为质量管理、人工智能与优化等,E-mail:1286903026@qq.com(通信作者)。

推精度都较高。由此可知,利用组合模型进行相关的分析是可行的,因此本文采用了具有短时预测精度高的灰色预测模型、可利用变量与结果模拟稳定的神经映射的 RBF 神经网络模型以及基于标准差法的组合预测模型,并且对于 RBF 神经网络的预测模型学者们早有研究,如陈毅华等[10]采用 RBF 神经网络方法建立了人口老龄化的定量预测模型,同时用多元线性回归方法进行了比较,结果表明 RBF 预测精度更高。楼旭伟等[11]为了提供交通预测的准确性,提出了基于遗传算法的 RBF 神经网络方法,利用遗传算法优化其权值和阈值,并与 BP 神经网络的预测结果相比较,其结果表明,RBF 神经网络模型对于交通流具有较好的非线性拟合能力,其精度高于 BP 网络。周泰等[12]将 GM(1,1)预测模型和径向基神经网络有效结合,利用灰色系统的贫乏数据建模和神经网络的高度非线性映射能力的优势,构建串联非线性组合预测模型,用来分析铁路投资规模的预测,取得了更加合理精确的预测结果。Wu 等[13]将粒子群自适应优化算法(PSO)引入遗传算法(GA)中来确定径向基函数神经网络的参数,用于降雨预测,提高了预测精度和更好的泛化能力。因此,将 RBF 作为对人口总数的一种预测方法是非常可观的,本文在山东人口总数的预测问题的探讨中,借助相关的研究理论,构建 GM(1,1)、RBF 神经网络以及组合预测模型来进行分析研究。

本文通过 MATLAB 工具,首先对人口数据构建了灰色预测模型和径向基神经网络模型,再对其权重进行分配,最后利用分配的权重构建了基于标准差法的组合预测模型,并对其人口进行了短期的预测。

1 灰色预测模型原理

GM(1,1)模型是时间序列中最常见的一种灰色预测模型,其实质是对原始序列先作累加生成,然后构建一阶线性微分模型,在得到拟合函数后再对系统进行预测。其主要流程如下。

1.1 新序列的累加生成

已知原始序列为 $X^{(0)}=(X^{(0)}(1),X^{(0)}(2),\cdots,X^{(0)}(n))$,其中 $X^{(0)}(\cdot)$表示山东省历年人口总量,对原数据进行一次累加生成,用以弱化原始序列的随机性和波动性,可得到生成序列:

$$X^{(1)}=(X^{(1)}(1),X^{(1)}(2),\cdots,X^{(1)}(n))$$

其中,$X^{(1)}(k)=\sum_{i=1}^{k}X^{(0)}(i),k=1,2,\cdots,n$。

1.2 构建 GM(1,1)模型

利用一阶单变量微分方程进行拟合可得到白化方程 GM(1,1)模型:

$$\frac{\mathrm{d}X^{(1)}}{\mathrm{d}t}+aX^{(1)}=\mu$$

其中,a 表示发展灰数,μ 表示控制灰数。

对于以上两个灰数的求解,可以根据最小二乘法求得,令 $A=(a,\mu)^{\mathrm{T}}$ 可以得到 $A=(B^{\mathrm{T}}B)^{-1}B^{T}Y_n$,其中:

$$B=\begin{bmatrix}-\frac{1}{2}(X^{(1)}(1)+X^{(1)}(2)) & 1\\ -\frac{1}{2}(X^{(1)}(2)+X^{(1)}(3)) & 1\\ \cdots & \cdots\\ -\frac{1}{2}(X^{(1)}(n-1)+X^{(1)}(n)) & 1\end{bmatrix}$$

$$Y_n=\begin{bmatrix}X^{(0)}(2)\\ X^{(0)}(3)\\ \cdots\\ X^{(0)}(n)\end{bmatrix}$$

利用以上的计算结果对其进行还原可得:

$$\hat{X}^{(0)}(k+1)=\hat{X}^{(1)}(k+1)-\hat{X}^{(1)}(k)$$

其中,$\hat{X}^{(1)}(k+1)=(\hat{X}^{(0)}(1)-\frac{\mu}{a})\mathrm{e}^{-ak}+\frac{\mu}{a},k=1,2,\cdots,n$

构建的模型需要检验其适应性,而相对误差和级比偏差检验对于灰色模型的验证是比较可行的,因此其检验的计算公式如下。

(1)相对误差计算

$$\alpha_i^{(0)}=\frac{\varepsilon_i^{(0)}}{X_i^{(0)}}\times 100\%=\frac{X_i^{(0)}-\hat{X}_i^{(0)}}{X_i^{(0)}}\times 100\%$$

一般而言,要求相对误差不超过 5%,则认为其模型的拟合精度较高。

(2)级比偏差检验

将级比偏差值定义为:

$$\beta(k)=\frac{\hat{\sigma}^{(0)}-\sigma^{(0)}(k)}{\hat{\sigma}^{(0)}}\times 100\%$$

$$=(1-\frac{1+0.5a}{1-0.5a}\sigma^{(0)}(k))\times 100\%$$

其中,模型的级比为:

$$\hat{\sigma}^{(0)}=\frac{\hat{X}^{(0)}(k-1)}{\hat{X}^{(0)}(k)}=\frac{1+0.5a}{1-0.5a}$$

而序列的级比为:

$$\sigma^{(0)}(k)=\frac{X^{(0)}(k-1)}{X^{(0)}(k)}$$

由此可计算得到级比偏差。

本文利用灰色预测相比传统预测方法只需少量原数据的优点,对 1995-2014 年的人口数据进行研究,其具体数据以及预测结果如表 1 所列。

表 1 GM(1,1)预测结果

年份	真实值	GM 预测值	残差	GM 相对误差/%	级比偏差
1995	8705	8705.00	0	0.00	
1996	8738	8734.70	3.3	0.04	-0.0026
1997	8785	8790.81	-5.81	0.07	-0.001
1998	8838	8847.29	-9.29	0.11	-0.0004
1999	8883	8904.13	-21.13	0.24	-0.0013
2000	8998	8961.34	36.66	0.41	0.0064
2001	9041	9018.91	22.09	0.24	-0.0016
2002	9082	9076.85	5.15	0.06	-0.0019
2003	9125	9135.17	-10.17	0.11	-0.0017
2004	9180	9193.86	-13.86	0.15	-0.0004
2005	9248.18	9252.93	-4.75	0.05	0.001
2006	9308.9	9312.37	-3.47	0.04	0.0001
2007	9366.97	9372.20	-5.23	0.06	-0.0002
2008	9417.23	9432.41	-15.18	0.16	-0.0011
2009	9470.3	9493.01	-22.71	0.24	-0.0008
2010	9587.87	9554.00	33.87	0.35	0.0059
2011	9637.27	9615.38	21.89	0.23	-0.0013
2012	9684.87	9677.16	7.71	0.08	-0.0015
2013	9733.39	9739.33	-5.94	0.06	-0.0014
2014	9789.43	9801.90	-12.47	0.13	-0.0007

根据模型的构建方法以及利用 MATLAB 工具对数据进行计算可得,GM(1,1)预测模型为:

$$\hat{X}^{(1)}(k+1)=(\hat{X}^{(0)}(1)+\frac{8651.01}{0.0064})e^{0.0064k}-\frac{8651.01}{0.0064}$$

$k=0,1,2,\cdots,n$

根据上述的检验方法对模型结果进行检验，可以得出：GM 的相对误差均未超过 5%，级比偏差趋于 0，并且经后验差检验，可计算方差 $c=S_1/S_0$ 小于 0.05，模型的精度达到 85.85%。由此可知，GM(1,1)对于人口的预测精度是较高的。

2　径向基神经网络模型

由 Moody 和 Darken 提出的径向基(Radial Basis Function，RBF)神经网络是以函数逼近理论为基础而构造的一类前馈网络。径向基神经网络的学习等价于在多维空间中寻找训练数据的最佳拟合平面。网络的每个隐含层神经元传递函数都构成了拟合平面的一个基函数，因此被称为径向基函数神经网络。由于其具有较强的自适应性和泛化能力，因此能够逼近任意非线性函数，可较好地解决难以用数学方法直接解析的但具有一定规律性的难题。

径向基神经网络是由输入层、输出层和隐含层所形成的三层前向型网络，由输入层节点获取数据后传递到隐含层，再通过变换后传出。其隐含层节点则是由径向基函数构成的，其形式多样，而输出层通常为线性函数，其拓扑结构如图 1 所示。

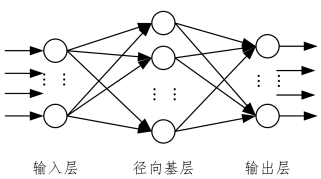

图 1　RBF 神经网络拓扑结构图

本文中的隐含层节点采用的径向基函数是 Gaussian 函数，即：

$$R_i(x)=G(\|X-C_i\|)=\exp(-\frac{1}{2b^2}\|X-C_i\|^2)$$

其中，X 是输入向量 $X=[x_1,x_2,\cdots,x_m]^T$，C_i 是隐含层中的第 i 个神经元中心 $C_i=[c_{1i},x_{2i},\cdots,x_{mi}]^T$。

而其输出可表示为：

$$y_i(x)=\sum_{i=1}^{k}w_{ij}\exp(-\frac{1}{2b_i^2}\|X-C_i\|^2)$$

其中，w_{ij} 是相对应的连接权值，b_i 是第 i 个神经元 Gaussian 函数的宽度。

本文以山东省 20 年间的历史人口总数为基础，构建三层 RBF 神经网络预测模型，其具体流程如图 2 所示。

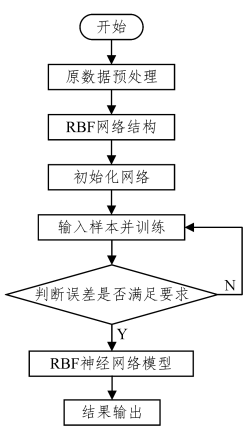

图 2　RBF 网络流程图

2.1　预处理数据

为避免神经元饱和现象，需要对 RBF 网络进行训练的输入原数据进行预处理，常用的方法是进行归一化处理，并且适当的处理可以加速收敛。其归一化方法有很多，而论文中所用的方法为：

$$\bar{x}=\frac{2x-x_{max}-x_{min}}{x_{max}-x_{min}}$$

其中，x 为输入，x_{max} 和 x_{min} 为输入的最大值与最小值，经过此变化后，其数据全在[−1,1]区间之内。

2.2　确定输入、输出和隐含层

对于人口总数的离散型时间序列 $y(t)$ 而言，利用 RBF 神经网络进行预测时，其实质是函数 F 使时间序列 $y(t)$ 满足 $y(t)=F(y(t-1),y(t-2),\cdots,y(t-n))$，以 20 年的数据为基础，通过实验和绝对误差值分析可得，以过去的 6 年数据来进行预测是比较精确的，因此输入层的神经元数为 6。

神经网络的输出层神经元数目是根据问题而确定的，故而，针对山东省每年的人口总数的问题而言，输出层神经元数目为 1。

本文在利用 RBF 网络进行训练时，对于隐含层神经元数量的确定采用的是将其设定为与输入样本数目相等的数目。

2.3　RBF 神经网络参数设置

对于 RBF 神经网络算法来说，需要求解的参数有：基函数中心、方差以及隐含层到输出层的权值。论文选用基于 MATLAB 工具箱的 newrb 函数来生成 RBF 神经网络来以实现对数据的训练和预测，参数的设置如下：输入层神经元数目设为 6，输出层为 1，隐含层最大数为 20，RBF 函数分布密度 Spread 为 0.7。

3　数值模拟

本文采用山东省 20 年间的人口数据进行预测研究，并对未来几年的人口总数进行短期预测，研究利用 MATLAB 工具进行仿真分析，其结果如表 2 所列。

表 2　RBF 的预测结果

年份	真实值	RBF 预测值	残差值	RBF 相对误差/%
1995	8705	8708.33	−3.33	0.038
1996	8738	8741.15	−3.15	0.036
1997	8785	8788.52	−3.52	0.04
1998	8838	8852.81	−14.81	0.167
1999	8883	8886.40	−3.40	0.038
2000	8998	9001.44	−3.44	0.038
2001	9041	9024.27	16.73	0.185
2002	9082	9065.20	16.80	0.185
2003	9125	9128.29	−3.29	0.036
2004	9180	9181.38	−1.38	0.015
2005	9248.18	9247.04	1.14	0.012
2006	9308.9	9308.68	0.22	0.002
2007	9366.97	9372.65	−5.68	0.061
2008	9417.23	9450.30	−33.07	0.35
2009	9470.3	9481.53	−11.23	0.118
2010	9587.87	9577.60	10.08	0.105
2011	9637.27	9628.34	8.93	0.093
2012	9684.87	9688.36	−3.49	0.036
2013	9733.39	9711.28	22.11	0.228
2014	9789.43	9805.88	−16.45	0.168

通过表 1 和表 2 的结果可知，RBF 神经网络预测的平均相对误差小于 GM 预测的平均相对误差，而二者的相对误差

的最大值不超过 0.5%，这表明其都具有较高的预测精度，但灰色预测只具有短期的预测精度，而从平均相对误差而言，可以说明 RBF 对于人口总量的预测精度较高。而 RBF 神经网络也可能具有偶然因素，因此，为了弥补单一模型的缺陷，引入组合预测模型对人口总量进行预测。

3.1 组合预测模型

由 Bates 等[14]于 1969 年提出的组合预测模型是一种将不同的模型结合在一起的预测方法，考虑单一预测方法的局限性，综合利用各种预测方法所提供的信息，扬长避短，使其获得更满意的结果。组合预测的基本思想：先从各种单项预测模型中提取出有效的系统信息，然后找到一种准则或方式，来让这些不同的预测模型进行合理有效的组合，并选择合适的权系数进行加权，最终得到最合理的组合预测模型。目前比较常用的组合预测模型有：平均值法、递归最小二乘法、模糊逻辑系统、BP 网络法、标准差法等。本文选用简单而有效的标准差法，得出最优加权系数作为组合预测模型对山东省的人口总数进行预测。

(1)模型简介

设灰色模型的误差标准差为 σ_1，RBF 神经网络的误差标准差为 σ_2，且 $\sigma=\sigma_1+\sigma_2$，则最优加权系数以及组合预测值分别为：

$$w_i=\frac{\sigma-\sigma_i}{\sigma},X_s=\sum_{i=1}^{m}w_iX_i,i=1,2$$

其中，w_i 是第 i 种预测方法的加权系数，X_i 是第 i 种预测方法的预测值，X_s 是组合预测值。

(2)数值计算

按照上述方法重新计算 GM 预测模型和 RBF 神经网络模型的权重，分别为 0.4417 和 0.5583。则组合预测模型为：

$X_s=0.4417X_1+0.5583X_2$

其组合预测结果如表 3 所列。

表 3 组合模型预测结果

年份	真实值	预测值	残差值	相对误差/%
1995	8705	8706.86	−1.85	0.021
1996	8738	8738.30	−0.30	0.003
1997	8785	8789.53	−4.53	0.052
1998	8838	8850.37	−12.37	0.139
1999	8883	8894.23	−11.23	0.126
2000	8998	8983.73	14.27	0.159
2001	9041	9021.90	19.10	0.212
2002	9082	9070.34	11.66	0.129
2003	9125	9131.33	−6.33	0.069
2004	9180	9186.89	−6.89	0.075
2005	9248.18	9249.64	−1.46	0.016
2006	9308.9	9310.31	−1.41	0.015
2007	9366.97	9372.45	−5.47	0.059
2008	9417.23	9442.40	−25.17	0.267
2009	9470.3	9486.60	−16.30	0.172
2010	9587.87	9567.28	20.59	0.215
2011	9637.27	9622.61	14.66	0.152
2012	9684.87	9683.41	1.46	0.015
2013	9733.39	9723.67	9.72	0.1
2014	9789.43	9804.12	−14.69	0.149

由表 1−表 3 的相关数据可以看出，组合预测模型的最大相对误差为 0.267%，其平均误差为 0.0091%，两者均小于单一预测模型。由此可见，组合预测模型具有较强预测性的优点，所表现的预测精度都要高于单一的预测模型，对于人口的预测问题是非常实用的。

(3)预测实例及分析

根据上述灰色预测模型、RBF 神经网络模型以及组合预测模型分别对山东省 2015−2025 年的人口总量进行预测，其结果如表 4 所列。

表 4 山东省 2015−2025 年的预测结果

年份	GM 预测值	RBF 预测值	组合预测值
2015	9864.88	9852.75	9858.11
2016	9928.26	9948.32	9939.46
2017	9992.04	9995.66	9994.06
2018	10056.24	10046.01	10050.53
2019	10120.84	10097.77	10107.96
2020	10185.87	10155.91	10169.14
2021	10251.31	10213.12	10229.99
2022	10317.17	10264.29	10287.65
2023	10383.45	10324.82	10350.72
2024	10450.16	10380.35	10411.19
2025	10517.30	10439.86	10474.07

利用不同预测模型对从 2015 年到 2025 年短时间段内山东省的人口总量变化进行预测分析，预测结果表明：RBF 预测模型所测算的数值基本都小于 GM 预测模型的数值，而组合模型的数值却位于二者之间，相当于对二者的数值求取均值。但任何一种预测模型的预测结果都不可能是完全精确的，这是因为任何一种模型都是直接或者间接地依据历史数据来进行分析测算，而随着时间的流逝，社会的发展以及各种宏观调控等因素会发生变化，从而影响预测结果，并且根据上一年的预测结果来预测未来人口，这将极大地影响预测的准确性。

未来人口总量的变化规律不是一成不变的，它是一个动态的发展过程。它不仅受诸多外部环境因素的影响，还与国家的政策变化息息相关。具体如下：

(1)经济因素。正如马尔萨斯和马克思人口论所说，经济因素是影响人口增长的决定性因素。其决定了人口的增殖条件和生存条件，通过改变人口的出生率和死亡率来影响人口的自然增长率。

(2)文化因素。随着受教育水平的提高，人们的平均婚龄、生育、育儿等文化理念都发生着变化，已经成为影响人口增减变化的重要因素之一。

(3)医疗卫生因素。医学的进步和医疗卫生事业的发展对人口出生率和死亡率有着直接影响，并且它对控制生育和实行优生优育有着积极的作用。

(4)国家政策因素。自 2016 年 1 月 1 日起，全面二孩政策正式实施。国家鼓励生育，将在一段时间内对人口增减变化产生较大的影响。

结束语 本文从灰色预测、RBF 神经网络入手，以山东省人口总量的短期预测为研究对象，建立了一种基于灰色预测模型和径向基网络模型应用标准差法的组合预测模型。并对提出的组合预测模型方法进行了验证分析，其结果表明组合预测模型具有更高的预测准确性。

参 考 文 献

[1] 周志刚，万立，陈丽红. 灰色混沌神经元网络模型及其短期人口预测[J]. 系统工程，2012，30(10)：118-122.

[2] 贾楠，胡红萍，白艳萍. 基于 BP 神经网络的人口预测[J]. 山东理工大学学报(自然科学版)，2011，25(3)：22-24.

[3] 张海峰，杨萍，李春花，等. 基于多模型的西宁市人口规模预测

[J]. 干旱区地理，201336(5)：955-962.

[4] 王宇熹，汪泓，肖峻. 基于灰色 GM(1,1)模型的上海城镇养老保险人口分布预测[J]. 系统工程理论与实践，2010，30(12)：2244-2253.

[5] 李富荣. 改进的动态 GM(1,1)模型在人口预测中的应用[J]. 统计与决策，2013(19)：72-74.

[6] SHANG H L，SMITH P W F，BIJAK J，et al. A multilevel functional data method for forecasting population，with an application to the United Kingdom [J]. International Journal of Forecasting，2016，30(4)：1098-1109.

[7] 何思兰，孙红兵. 基于灰色预测和 BP 神经网络模型的云南省人口总量预测研究[J]. 计算机与数字工程，2016，44(2)：193-196，236.

[8] 任强，侯大道. 人口预测的随机方法：基于 Leslie 矩阵和 ARMA 模型[J]. 人口研究，2011，35(2)：28-41.

[9] 蒋若凡，姜玉梅，李菲雅. 基于灰色 PSO-BP 人口预测模型的研究与应用[J]. 西北人口，2011，32(3)：23-26.

[10] 陈毅华，李永胜，苏昌贵，孙峰华. 径向基神经网络模型在人口老龄化预测中的应用—以湖南省为例[J]. 经济地理，2012，32(4)：32-37.

[11] 楼旭伟，楼辉波，朱剑锋. 基于遗传算法径向基神经网络的交通流预测[J]. 中国科技论文，2013，8(11)：1141-1144.

[12] 周泰，叶怀珍，王亚玲. 基于灰色径向基函数网络的铁路投资规模组合预测[J]. 北京交通大学学报(社会科学版)，2009，8(4)：33-37.

[13] WU J S，LONG J，LIU M Z. Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm [J]. Neurocomputing，2015，148：136-142.

[14] BATES J M，GRANGER C W. Combination of Forecasts [J]. Operational Res Ouart，1969，20(4)：451-468.

(上接第 430 页)

行效用性分析。对结果进行粗略分析可以看出，普遍不被学生重视的学生思想政治以及心理状态等也是严重影响学生毕业的关键因素，对各个关联规则如“基础数学 0＋数据结构 0→数据库设计及应用 0”的置信度为 0.5，效用度为 0.4 进行分析，可得出若对数据库设计及应用相关课程较差的同学进行基础性、针对性的辅导，则可帮助学生更加有效地提高成绩。

通过对结果的综合分析可得，基础课程的学习情况对学生后续专业性的课程学习影响较大；且掌握一类具有通用性质的专业相关知识也对以后的深入学习有较大的帮助；并且学生的思想政治教育以及历史文化学习在整个学习过程中也有着不小的作用；若学生注重毕业问题以及企业实践等，则可依情况进行管理类或语言类的学习等。若对实验结果进行进一步深入分析，不仅可分析课程、成绩之间的相互关联关系，同时还也可利用教育心理学、社会学对关系成因进行探讨解释。

结束语　本文为挖掘教务数据中学生课业成绩之间的关联关系，提出了基于领域知识的频繁高效有趣项集关联规则挖掘算法，即 FUI_DK 算法，详细解释了 FUI_DK 算法的核心思想，并对算法性能进行了实验证明。以学生教务信息数据为基础，基于学生学科学院、课程性质、课程成绩等跨维度属性进行学生课业关联规则挖掘，得到高效用的学院、课程及成绩等的相关关系。对学生而言得到的规则，可帮助学生在学习过程中根据自身需求进行针对性的学习，也可以避免由于前期学习时方向盲目而造成后期必修学科成绩不佳或必备技能能力不足；对管理者而言，可辅助分析课程安排的合理性，也可根据课程的影响程度酌情改善教学方式或加强教学力度等。

参 考 文 献

[1] FAYYAD U，PIATETSKY-SHAPIRO G，SMYTH P. From data mining to knowledge Discovery：an overview [C]// Advances in Knowledge Discovery and Data Mining. Menlo Park，California：AAAI Press，1996：1-35.

[2] AGRAWL R，SRIKANT R. Fast algorithms for mining association rules[C]// Very Large Data Base. 1994：487-499.

[3] HAN J，PEI J，YIN Y. Mining frequent patterns without candidate generation[C]// Special Interest Group on Management of Data. 2000：1-12.

[4] YAO H，HAMILTON H J，BUTS C J. A foundational approach to mining itemset utilities from databases[C]// Siam International Conference on Data Mining. 2004：482-486.

[5] TSENG V S，WU C W，SHIE B E，et al. UP-Growth：an efficient algorithm for high utility itemset mining [C] // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM，2010：253-262.

[6] WU C W，SHIE B E，TSENG V S，et al. Mining top-K high utility itemsets[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM，2012：78-86.

[7] 李慧，刘贵全，瞿春燕. 频繁和高效用项集挖掘[J]. 计算机科学，2015，42(5)：82-87.

[8] PEI J，HAN J，MORTAZAVIASL B，et al. PrefixSpan：Mining sequential patterns efficiently by prefix-projected pattern growth [C]// Proceedings 17th International Conference on Data Engineering. ICDE，2001：215-224.

[9] 吴倩，王林平，罗相洲，等. 一种快速挖掘 top-k 高效用模式的算法[J]. 计算机应用研究，2017，34(11)：3303-3307.

[10] 王敬华，罗相洲，吴倩. 基于投影的高效用项集挖掘算法[J]. 小型微型计算机系统，2016，37(6)：1212-1216.

[11] 潘海为，韩启龙，印桂生，等. 基于领域知识指导的医学图像关联规则挖掘[J]. 计算机研究与发展，2007，44(z3)：424-428.

[12] 潘海为，谭小雷，韩启龙. 领域知识驱动的医学图像关联模式挖掘算法[J]. 黑龙江大学自然科学学报，2009，26(5)：585-590.

[13] 张晶，张斌，胡学钢. 基于领域知识的冗余关联规则消除算法[J]. 合肥工业大学学报(自然科学版)，2011，34(2)：246-250.

[14] SHEN W，WANG J，HAN J. Sequential Pattern Mining[M]// Frequent Pattern Mining. Springer International Publishing，2014：512-517.