

人才需求组合预测的建模与仿真分析

杨焕海

(山东工商学院 山东 烟台 264005)

摘要: 研究人才需求预测问题,人才需求受到多种因素,不仅具有线性变化规律,而且具有非线性变化规律,单一线性或非线性的都无法获得较高的预测精度。为提高人才需求的预测精度,提出一种 ARIMA-LSSVM 的人才需求组合预测模型。首先采用 ARIMA 对人才需求历史数据线性变化规律进行预测,然后采用 LSSVM 对 ARIMA 的残差进行预测,用以刻画数据中的非线性变化规律,最后将两种预测结果之和作为人才需求最终的预测结果。仿真结果表明,ARIMA-LSSVM 建立的人才需求预测可以获得较高的预测精度,解决了单一模型存在的难题,为人才需求准确预测提供了依据。

关键词: 人才需求;差分自回归滑动平均;最小二乘支持向量机;组合预测

中图分类号: X949 文献标识码: A

Intelligent Modeling and Simulation Analysis for Talent Demand Forecasting

YANG Huan-hai

(Shandong Institute of Business and Technology, Yantai Shandong 264005, China)

ABSTRACT: Research on talent demands. The talent demands affected by many factors has the linear variation and the nonlinear variation, therefore, a single linear or nonlinear cannot obtain higher accuracy. In order to improve the prediction precision of talent demands, this paper proposed a forecast model based on ARIMA-LSSVM for talent demands. Firstly, the ARIMA was used to forecast the linear variation rule for talent demands, and then the LSSVM was used to correct the forecasting residuals of ARIMA prediction to describe nonlinear variation rules. Finally, the two prediction results of talent demands were added as the final prediction results. The simulation results show that the model established based on ARIMA-LSSVM for talent demand forecasting can obtain higher prediction precision and solve the problem in single models, so the prediction accuracy is better than that of the single ARIMA or LSSVM model.

KEYWORDS: Talent demand; Auto-regressive integrating moving average; Least squares support vector machine; Combination forecasting

1 引言

人才需求预测是指对历史资料进行统计和分析,预测未来人才需求变化。人才需求受到企业、城市、经济、政策等诸多因素影响,呈现出错综复杂变化特点,不仅含有线性变化特点,又含有非线性变化规律,获得高精度的人才需求预测结果成为一个亟待解决的难题^[1]。

当前人才需求预测方法主要有:时间序列分析法、回归分析法、灰色预测法、支持向量机和神经网络等^[2-5]。对于

复杂的人才需求预测问题,单一预测模型均存在各自的不足。如回归分析、时间序列分析法均只能预测人才需求线性变化规律,预测精度低;灰色预测法只适合对服从指数规律数据进行预测,但人才需求数据具有趋势性和非平稳性,无法满足人才需求实际预测要求^[6]。神经网络非线性建模预测能力强,但其基于经验风险最小化建模,预测结果易出过拟合,泛化能力差^[7];最小二乘支持向量机泛化能力强,但不具有解释能力^[8]。为了充分利用单一预测模型优点,有人提出了人才需求的组合预测模型,实证表明,相对于一种单一模型,组合预测模型更能描述人才需求变化特点,提高预测精度,是当前人才需求预测主要的研究方向^[9]。

为此,本研究借鉴已有的研究成果,将差分自回归移动

基金项目 “2012 年江西省高校党建研究课题:规划一般项目”

收稿日期: 2013-01-04 修回日期: 2013-01-20

平均法(ARIMA)和最小二乘支持向量机(LSSVM)有机结合,提出了一种ARIMA-LSSVM的人才需求组合预测模型,并通过仿真验证了ARIMA-LSSVM的可行性和有效性。

2 人才需求组合模型

ARIMA假设人才需求是线性变化,但实际上人才需求不是完全线性变化,预测结果出现偏差,就需要采用非线性预测模型对其偏差进行修正,而LSSVM具有优异的非线性处理能力,恰好能弥补ARIMA缺陷。ARIMA和LSSVM具体组合原理为:首先,建立ARIMA模型分析人才需求数据线性部分;然后根据ARIMA模型的人才需求预测残差构建LSSVM模型,最后将预测结果进行有机结合,得到组合模型的人才需求预测值,这样充分发挥ARIMA和LSSVM的各自优势,进行取长补短。ARIMA-LSSVM的工作原理如图1所示。

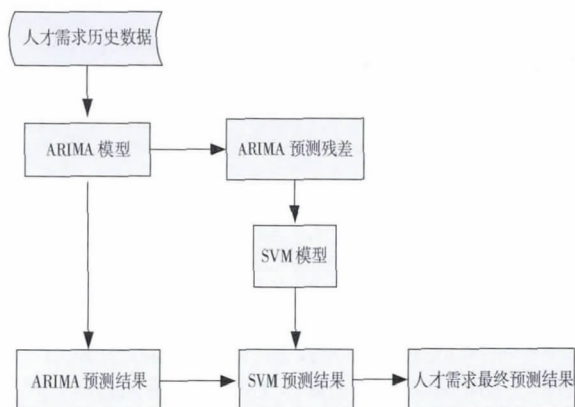


图1 人才需求组合预测模型的工作原理

3 人才需求的线性和非线性预测模型

3.1 人才需求的线性预测模型

ARIMA模型由自回归(AR)、I差分(Differencing)或结合(Integration)和移动平均(MA)三部分组成的^[10]。以 p, d, q 为参数,人才需求的线性预测的ARIMA模型为:

$$y_k = \theta_0 + \varphi_1 y_{k-1} + \varphi_2 y_{k-2} + \cdots + \varphi_p y_{k-p} + \varepsilon_k - \theta_1 \varepsilon_{k-1} - \theta_2 \varepsilon_{k-2} - \cdots - \theta_q \varepsilon_{k-q} \quad (1)$$

式中: y_k 为样本值; $\varphi_i (i=1, 2, \cdots, p)$ 和 $\theta_j (j=1, 2, \cdots, q)$ 为模型参数; ε_k 为随机误差,其均值为0。ARIMA建模包括:样本平稳化处理;模型定阶;模型检验;预测等4个步骤。

3.2 人才需求的非线性预测模型

SVM求解相应的二次规划问题越复杂,计算速度越慢,LSSVM对支持向量机进行了改进,降低了求解的复杂性,因此本研究采用LSSVM作为预测模型。对于人才需求历史数据,LSSVM的回归函数为

$$f(x) = w^T \varphi(x) + b \quad (2)$$

式中 w 为权值向量 b 为偏置量。

引入结构风险函数,式(2)转化为二次优化问题:

$$\min \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 \quad (3)$$

约束条件为:

$$y_i = w^T \varphi(x) + b + \xi_i \quad (4)$$

式中 γ 为正规化参数; ξ_i 为松弛变量。

引入拉格朗日乘子,可得

$$L(w, b, \xi, \alpha) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 + \sum_{i=1}^n \alpha_i (w^T \varphi(x_i) - b + \xi_i - y_i) \quad (5)$$

式中 α_i 为拉格朗日乘子,

根据KKT条件得到

$$\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi_i} = 0, \frac{\partial L}{\partial \alpha_i} = 0 \quad (6)$$

因此,可以得到最后的解为:

$$\begin{bmatrix} 0 & 1^T \\ 1 & x^T x + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (7)$$

引入核函数 $k(x_i, x_j)$ 对式(7)进行转换,得到LSSVM预测模型为

$$f(x) = \sum_{i=1}^n \alpha_i k(x_i, x_j) + b \quad (8)$$

本研究采用RBF函数作为LSSVM核函数,最后得3人才需求的非线性预测模型为

$$f(x) = \sum_{i=1}^N \alpha_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + b \quad (9)$$

式中 σ^2 为RBF核宽度。

4 组合人才需求模型的建立步骤

对于人才需求历史数据的时间序列 (y_t) ,既含有线性自相关规律 (L_t) ,又包含非线性变化规律 (N_t) ,那么人才需求数据可以表示为

$$y_t = L_t + N_t \quad (10)$$

具体建模步骤如下:

第1步:利用ARIMA对 y_t 建立线性预测模型,得到的预测结果为 \hat{L}_t ,那么在 t 时刻,原始人才需求时间序列值与ARIMA预测结果之间的残差为 e_t ,即

$$e_t = y_t - \hat{L}_t \quad (11)$$

第2步:由于ARIMA对人才需求的线性变化部分进行了预测,因此其非线性部分就隐含于 $\{e_t\}$ 中,采用SVM对 $\{e_t\}$ 进行建模预测,就可以挖掘出人才需求中的非线性信息。设时刻 t 的残差值与前 m 个残差相关联,那么残差的计算公式为

$$e_t = f(e_{t-1}, e_{t-2}, \cdots, e_{t-m}) + \varepsilon_t \quad (12)$$

式中 ε 为一随机误差; $f()$ 为一非线性函数,通过SVM对其进行逼近,得到预测结果为 \hat{N}_t 。

第3步:组合ARIMA和SVM预测结果得到人才需求最终预测结果 \hat{y}_t ,即

$$\hat{y}_t = \hat{L}_t + \hat{e}_t \quad (13)$$

从人才需求组合预测建模过程看,通过ARIMA描述其

线性部分,而 SVM 描述非线性部分,充分发挥了两种模型长处,弥补了各自不足,改善了人才需求预测效果。

5 仿真研究

5.1 数据来源

采用某省 1985 ~ 2011 年人才资源数据进行仿真,数据具体如图 2 所示。1985 ~ 2001 年数据作为训练集,建立 ARIMA - LSSVM 人才需求预测模型,2002 ~ 2011 年数据作为测试集检验 ARIMA - LSSVM 模型的预测性能。

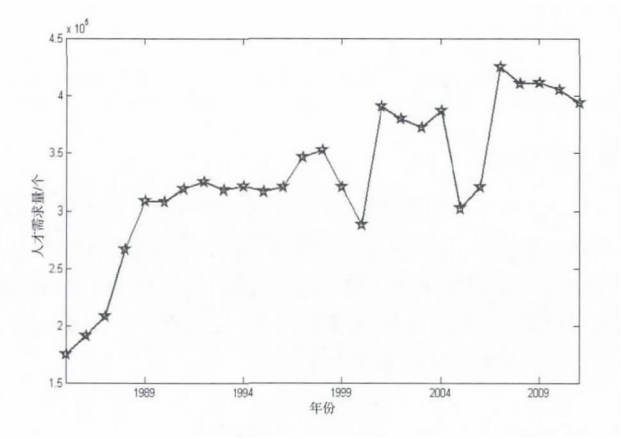


图2 人才需求数据

5.2 对比模型和评价标准

为了衡量 ARIMA - LSSVM 的有效性,选择单一的 ARIMA、LSSVM 进行对比仿真。采用均方误差 (MSE) 和平均绝对误差百分比 (MAPE) 作为评价标准。MSE 和 MAPE 定义如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{14}$$

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100 \tag{15}$$

式中 n 为样本数量; y_i 为人才需求实际值; \hat{y}_i 为模型的预测值。

5.3 ARIMA - LSSVM 人才需求预测模型实现

5.3.1 ARIMA 对人才需求预测

对原始人才需求数据进行 2 阶差分变换后,其时阶序列如图 3 所示。2 阶差分后序列的正态分检验如图 4 所示。从图 3 和图 4 可知,变换后的新人才需求序列具有平稳性,因此可以确定模型参数 $d=2$ 。

人才需求时间序列经过 2 阶差分的样本相关函数和偏自相关函数表 1 所示,从表 1 可知,可以选择 ARIMA(4 2, 3) 对人才需求进行预测。

采用 ARIMA(4 2 3) 对 2002 ~ 2011 年数据进行预测,如图 5 所示结果。

5.3.2 LSSVM 对 ARIMA 的残差预测

采用 2002 ~ 2011 年人才需求实际值与 ARIMA(4 2 3)

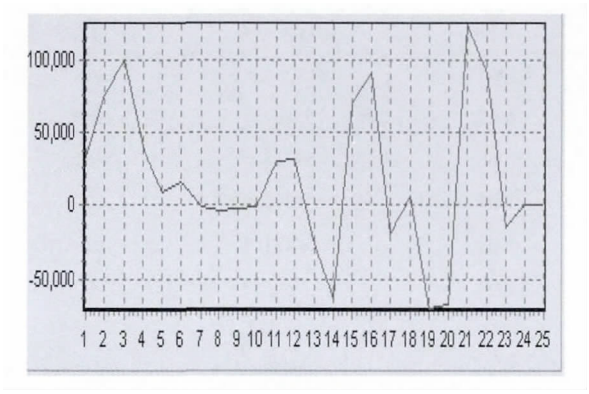


图3 1985 ~ 2011 年的 2 阶人才需求序列

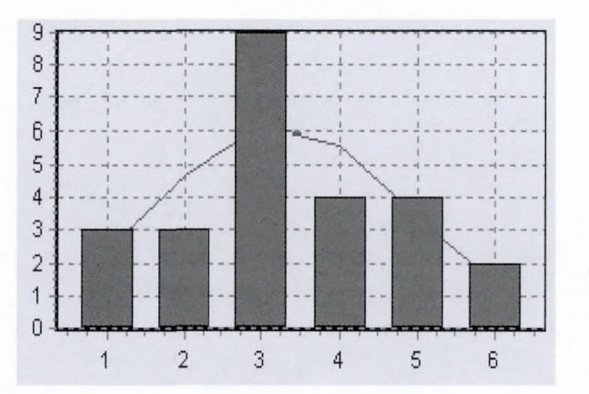


图4 2 阶差分后的正态分检验

预测结果进行相减,得到 ARIMA 的残差序列,采用 LSSVM 对残差序列预测函数 $f()$ 进行逼近,最后到的残有效期预测值如表 2 所示。

表 1 差分后的相关函数和偏自相关函数值

| | 样本自相关函数 | 样本偏自相关函数 |
|----|---------|----------|
| 1 | 0.2262 | 0.2262 |
| 2 | -0.4025 | -0.4781 |
| 3 | -0.1930 | 0.0679 |
| 4 | -0.1422 | -0.3958 |
| 5 | 0.1623 | 0.4211 |
| 6 | 0.3214 | -0.2079 |
| 7 | -0.1111 | 0.0791 |
| 8 | -0.2554 | -0.2091 |
| 9 | 0.0007 | 0.2440 |
| 10 | 0.0534 | -0.2880 |
| 11 | -0.0589 | -0.0157 |

表 2 ARIMA 的残差预测值

| 年份 | ARIMA 残差 | LSSVM 预测值 |
|------|----------|-----------|
| 2002 | -4996.78 | -996.781 |
| 2003 | -37857.9 | -17857.9 |

| 年份 | ARIMA 残差 | LSSVM 预测值 |
|------|----------|-----------|
| 2004 | 84511.25 | 14511.25 |
| 2005 | -35807.4 | -25807.4 |
| 2006 | -84538.5 | 15461.5 |
| 2007 | -10190.5 | -190.469 |
| 2008 | -20181.3 | -10181.3 |
| 2009 | -17625.3 | 2374.656 |
| 2010 | 967.5312 | 1967.531 |
| 2011 | -18055.4 | -8055.41 |

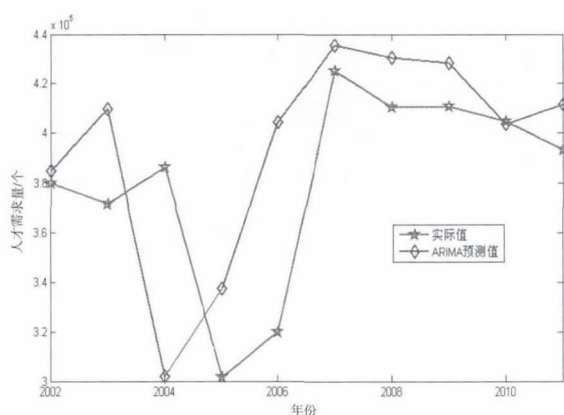


图5 ARIMA的测试集预测结果

5.3.3 两种预测结果的融合

采用 LSSVM 预测结果与 ARIMA(4,2,3) 预测结果相加,对于 ARIMA(4,2,3) 预测的误差进行修正,最终人才需求预测结果如图 6 所示。对图 5 和图 6 进行对比分析可知,ARIMA-LSSVM 的预测精度更高,说明将 ARIMA 和 LSSVM 组合在一起,用于变化复杂的人才需求预测是可行的和有效的。

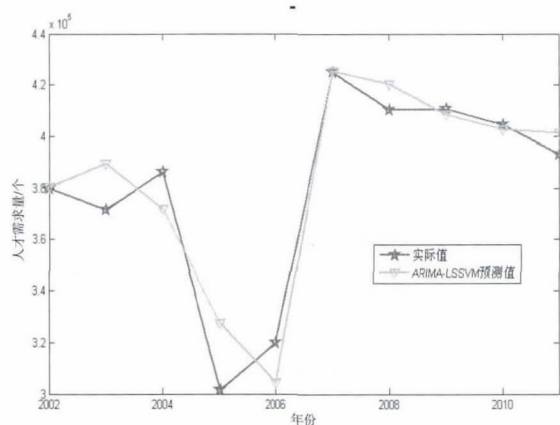


图6 ARIMA-LSSVM对测试集的预测结果

5.4 与单一 ARIMA、LSSVM 性能比较

ARIMA、LSSVM 和 ARIMA-LSSVM 预测结果的误差如表 3 所示。从表可知,相对于单一 ARIMA、LSSVM,ARIMA-LSSVM 的预测误差更小,这说明单一预测模型只能描述人才需求部分变化信息,ARIMA-LSSVM 可以同时从线性和非线性部分对人才需求变化趋势进行挖掘,提高了预测精度,为有效解决人才需求预测问题提供了一种新的预测手段。

表3 各人才需求预测模型的性能对比

| 模型 | RMSE | MAPE |
|-------------|-----------|-------|
| ARIMA | 42635.65 | 8.851 |
| LSSVM | 31546.58 | 6.127 |
| ARIMA-LSSVM | 12702.954 | 2.785 |

6 结束语

人才需求受到多种因素综合影响,具有线性和非线性的复杂变化特点,为提高其预测精度,提出一种基于 ARIMA-LSSVM 人才需求预测模型。仿真结果表明,ARIMA-LSSVM 结构简单,预测精度高,解释能力强,要优于单预测模型,可以准确把握人才需求总体变化趋势。

参考文献:

- [1] 鲍玲. 浅论人力资源预测及规划评估[J]. 经营管理, 2004, 10: 77-78.
- [2] 王鹏涛, 高喜珍. 人才需求及质量预测模型[J]. 系统工程理论与实践, 2000-12: 123-128.
- [3] 孙晋众, 林健. 人才需求预测指标体系及其实证分析[J]. 沈阳航空工业学院学报, 2007, 24(1): 92-94.
- [4] 杨月, 沈进. 多元线性回归分析在人才需求预测中的应用[J]. 商业研究, 2006-11: 33-34.
- [5] 张洪涛, 朱卫东. 安徽省高新技术人才需求的灰预测及灰色关联度研究[J]. 合肥工业大学学报(自然科学版), 2007, 30(20): 222-235.
- [6] 徐荣, 曹安照. 基于灰色系统理论的科技人才需求预测[J]. 安徽工程科技学院学报, 2006, 21(3): 71-74.
- [7] 张建勇, 赵涛. 基于 BP 神经网络的科技人才需求预测[J]. 科技管理研究, 2009-8: 501-502.
- [8] 肖健华. 人才需求预测的支持向量回归模型[J]. 统计与决策, 2009-6: 31-32.
- [9] 曹兴峰. 一种 BP 神经网络气象预测中的应用研究[J]. 科技通报, 2012, 8(28): 56-57.
- [10] 吴虹, 尹华. ARIMA 与 SVM 组合模型的石油价格预测[J]. 计算机仿真, 2010, 27(5): 264-267.



[作者简介]

杨焕海(1978-),男(汉族),山东泰安人,硕士研究生,讲师,主要研究领域为模式识别、机器学习。