

Research on the Application of Econometrics Based on Big Data Analysis in the Analysis of Urban Unemployment Rate

Yingji Cui

Xianyang Normal University, Xianyang, Shaanxi, China

Keywords: Big Data Analysis, Econometrics, Urban Unemployment Rate

Abstract: Influenced by Institutional Factors, China's Unemployment Rate Statistics Still Have Many Imperfections and Cannot Fully and Accurately Reflect the Real Employment Situation in China. If the Reverse Effect of the Large Increase in the Number of Migrant Workers in China on the Registered Unemployment Rate in China's Cities and Towns is Not Considered, the Main Factors That Affect the Increase in the Registered Unemployment Rate in China's Cities and Towns Are the Increase in the Organic Composition of China's Capital and the Increase in the Number of Working-Age Population in China. The Unemployment Certificate Data Lagging 4 Periods, the Labor Market Data Lagging 3 Periods, the Personnel Department Data Lagging 2 Periods, the Unemployment Insurance Data Lagging 1 Period, and the Current Job Search and Unemployment Rate Data Are Used as Explanatory Variables for Predicting the Unemployment Rate. In the Process of Designing an Oa System for an Institute, Combining with the Confidential Nature of the Institute, an Improved Rbac Model Was Designed, Which Inherited the Original Role-Based Authorization of the Rbac Model.

1. Introduction

Unemployment is One of the Most Important Problems Faced by China in the 21st Century. It is Not Only Closely Related to People's Life, But Also Related to Social Stability and Economic Prosperity [1]. However, Due to the Influence of Institutional Factors, There Are Still Many Imperfections in China's Unemployment Rate Statistics, Which Still Cannot Completely and Accurately Reflect the Real Employment Situation in China, Which to a Large Extent Weakens the Monitoring Role of Unemployment Rate in the Process of Macroeconomic Operation [2]. At Present, China Mainly Adopts Urban Registered Unemployment Rate and Survey Unemployment Rate to Reflect the Level of Unemployment [3]. However, in Terms of How to Understand the Influencing Factors of Unemployment in China's Socialist Market Economy, the Views of the Academic Community Are Inconsistent [4]. The Prediction of China's Unemployment Rate by Yu Big Data Technology Can Be an Important Supplement to the Current Incomplete Official Unemployment Rate Statistics [5]. At Present, There Have Been Many Researches on the Statistical Analysis of Unemployment Using Internet Search Data: Askitas and Zimmermannw Successfully Predicted the German Unemployment Rate Data Based on Google Search Keyword Data. The Study Found That the Prediction Based on Certain Keywords Can Be Earlier Than the Official Data to Show Changes in Unemployment Trends [6]. Therefore, How to Objectively and Reasonably Analyze Many Factors Affecting the Urban Unemployment Rate in China, and Establish Effective Prediction, Analysis and Research Models Have Become a Major Research Topic of Many Scholars and Experts [7].

2. Measurement Index of Unemployment Level in Chinese Cities and Towns Becomes Tortuous

2.1 Towns Becomes Tortuous

Under the Condition of Socialist Market Economy, the Existence of a Certain Number of Unemployed People is Conducive to the Formation and Improvement of China's Labor Market [8].

If We Do Not Consider the Reverse Effect of the Large Increase in the Number of Migrant Workers Across the Country on the Registered Unemployment Rate in China's Cities and Towns, the Main Factors Affecting the Increase in the Registered Unemployment Rate in China's Cities and Towns Are the Increase in the Organic Composition of China's Capital and the Increase in the Number of Working-Age Population Nationwide [9].during 2008-2018, Urban Unemployment Continued to Increase [11]. At the End of the 1970s, as the Educated Youth Returned to the City to Generate a Large Number of Urban Unemployed People, in Response to the Peak Unemployment Phenomenon, the Chinese Government Began to Publish the “Unemployed Urban Population” Statistical Regression Model of Nmse and Mse in the “China Statistical Yearbook” the Values Are Shown in Table 1. Table 1 Shows That the Support Vector Machine Test Set Has an Nmse Value of 0.1421 and an Mse Value of 8.0456. Both Values Are Smaller Than Those of Other Regression Methods. It Shows That Compared with Other Prediction Methods, the Support Vector Machine Has a Higher Degree of Fit and Stability, So This Paper Chooses Support Vector Machine as the Prediction Model of China's Quarterly Unemployment Rate.the Model is Linear to Non-Linear, Which Improves the Accuracy and Practicability of the Model, and Reduces the Error to a Certain Extent.

Table 1 Nmse and Mse of Regression Models

Method	Training set		Test set	
	NMS	MSE	NMSE	MSE
Linear regression	0.1421	8.0456	0.0076	5.3426
Regression tree	0.38329	18.1783	0.1587	16.7864
Neural network	0.4865	27.4865	0.7854	7.4586
Support vector regression mechanism	0.4878	4.7878	0.0485	8.4867
Random forest	0.7879	10.7878	0.7894	4.8759

2.2 Model and Algorithm Evaluation

The method of intelligent control principle is to combine fuzzy control with neural network, and to first fuzzify the whole system to obtain the corresponding fuzzy system [10]. For the measurement of unemployment level, there are two indicators commonly used in the world: registered unemployment rate and survey unemployment rate. Different keywords have different meanings and need to be screened by scientific methods. The vast majority of personnel participating in China's labor force survey are on temporary loan, only a small number of statisticians are from the Bureau of Statistics, and most of the rest of the investigators are cadres from various districts. Judging from the living conditions of the unemployed workers, it is very difficult for them to live during unemployment or lay-off.The relief they receive can often only maintain the minimum standard of living, and more people just issue a small monthly cost of living relationship from the original enterprise as the correlation coefficient between the two (see Table 2). Finally, the unemployment certificate data lagging 4 periods, the labor market data lagging 3 periods, the personnel department data lagging 2 periods, the unemployment insurance data lagging 1 period, and the current job search and unemployment rate data are used as explanatory variables for predicting the unemployment rate.

Table 2 Correlation Coefficient Between Search Term of Different Time Lag Order and Unemployment Rate

Key word	t-3	t-2	t-1	t	t+1	T+2	T+3
Unemployment certificate	0.669	0.615	0.589	0.559	0.458	0.245	0.568
Job-hunting	0.547	0.803	0.547	0.148	0.176	0.544	0.147
Unemployment insurance	0.156	0.950	0.259	0.248	0.462	0.368	0.578
Unemployment rate	0.024	0.564	0.987	0.367	0.785	0.248	0.684
labour market	0.315	0.508	0.564	0.156	0.247	0.147	0.247
Ministry of Personnel	0.145	0.826	0.777	0.214	0.298	0.267	0.178

3. Forecasting Model of China's Quarterly Unemployment Rate Based on Support Vector Regression

3.1 Revision of Official Annual Urban Registered Unemployment Rate Data

Support vector machine (SVM) is a set of learning algorithm proposed by vapk et al. In 1990s, when SVM is used in regression estimation, it is called support vector regression machine (SVR). From now to the year that can be predicted, the population and human resource supply in China will increase greatly, and the population peak will reach billion in this century. SVR is similar to neural network in form. The output is a linear combination of intermediate nodes. Each intermediate node corresponds to a support vector, and its weight is the corresponding Lagrange multiplier. Therefore, we have not retained the variable of economic growth rate (GDPZS) in the regression model. After that, with the acceleration of system reform and the reorganization of enterprises and institutions, the number of laid-off workers and unemployed will increase, which will undoubtedly lead to a further rise in the open unemployment rate in cities and towns. In addition, the existing surplus workers' unions in China's enterprises and institutions have been released along with the reform of enterprises and institutions, further aggravating the pressure on urban employment. Apply the annual regression equation and interpolate the annual unemployment rate with the GDP data from the first quarter of 2008 to the first quarter of 2016.

3.2 China's Quarterly Unemployment Rate

Although the Bureau of Statistics has tried to implement the unemployment rate survey system and China's unemployment rate survey is also conducted according to the standards of the International Labour Organization, due to China's economic and political reasons, China's unemployment rate survey indicators have not been released to the public. However, we should also clearly see that up to now, the transfer of rural labor force in our country has been relatively sufficient, and the positive effect of "the number of migrant workers" on reducing the unemployment rate will gradually weaken, which will also pose a challenge to further reduce the unemployment rate in our country. At the same time, various forms of foreign investment and various forms of economy, part-time, temporary, part-time and other forms of employment have sprung up rapidly and have become important channels for expanding employment. In view of the frequency and availability of the data, the GDP, CPI, interest rate, and urbanization rate data, which have a large impact on the unemployment rate, were selected using a variety of independent variables such as univariate, multivariate, and cross-variables. The Cchmv-lin method was used to The annual unemployment rate is interpolated. During the design of the OA system of a research institute, combined with the confidentiality of the research institute, an improved RBAC model was designed, which inherited the original role-based authorization of the RBAC model. The coefficient matrix A in the model is obtained by using the data from September, 2008 to April, 2009, and then the data of the main influencing factors) are brought into the model to obtain the unemployment rate for the whole time period (18 months). Finally, the unemployment rate obtained by the model is compared with the actual unemployment rate to test the error of the model. As shown in figure 1.

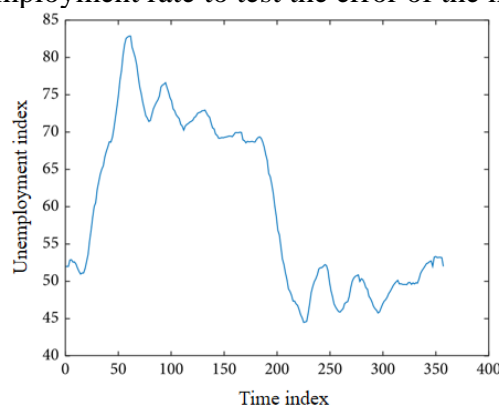


Fig.1 Conclusion of Multiple Regression Model

4. Conclusion

Although there are many factors affecting China's unemployment rate, the improvement of the organic composition of capital is still the primary factor affecting China's unemployment rate. Starting from the reality that our country is still in the middle and late stage of industrialization, compared with other statistical methods, support vector machine has a better effect in unemployment rate prediction. Support vector machine algorithm has its own advantages in solving nonlinear, small sample and high dimensional machine learning problems. In accordance with the different levels of confidentiality of the staff members and the secret content of different departments, access rules can block specific IP. The current unemployment measurement system still has problems such as poor timeliness of statistical data, incomplete statistical objects, insufficient depth of statistical indicators, multiple distortions of statistical data, and lack of international comparability of the statistical system. Using the advantages of big data sources, fully integrate the data from different sources, such as administrative records, e-commerce enterprise data, Internet search data, and sort out the basic information database of labor force that can be updated regularly. The calculation model of unemployment rate is also lack of validity for the calculation of long-term unemployment rate.

References

- [1] Wilson, J. (2019). Rose J . A predator in the park: mixed methods analysis of user preference for coyotes in urban parks. *Leisure Studies*, vol. 38, no. 1, pp. 1-17.
- [2] Erfu, DAI., Zhuo, W.U. Xiaodian, D.U. (2018). A gradient analysis on urban sprawl and urban landscape pattern between 1985 and 2000 in the Pearl River Delta, China. *Frontiers of Earth Science*, vol. 12, no. 4, pp. 791-807.
- [3] Zhou, C. (2019). Dynamic Behavior Analysis of Urban Light Rail Low Floor Train with Different Wheelset Passing Turnout Area. *Journal of Mechanical Engineering*, vol. 55, no. 2, pp. 98.
- [4] Lin, Z.S, Dai, Q.F. (2018). Zhang L, et al. Visualized simulation analysis of open urban community model. *Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)*, vol. 48, no. 1, pp. 336-343.
- [5] He, C. Zhu, L. Tian, W. (2018). Sensitivity analysis of urban building layout on energy consumption. *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology*, vol. 50, no. 4, pp. 174-180.
- [6] Gebhard, Kirchgässner. (1983). *Ökonometrie: Datenanalyse oder Theorienüberprüfung? / Econometrics: Data Analysis or Statistical Testing of Economic Theories?. Jahrbücher für Nationalökonomie und Statistik*, vol. 198, no. 6, pp. 511-538.
- [7] Wang, H. Yao, H. Kifer, D. et al. (2017). Non-Stationary Model for Crime Rate Inference Using Modern Urban Data. *IEEE Transactions on Big Data*, no. 99, pp. 1-1.
- [8] He, X, Zhang, H. Luo, T. et al. (2017). Network capacity analysis for cellular based cognitive radio VANET in urban grid scenario. *Journal of Communications & Information Networks*, vol. 2, no. 5, pp. 1-11.
- [9] Cui, W.P, Li, Y.B, Li, R.K, et al. (2017). Landscape ecological effect of urban expansion in the central area of Chongqing city based on gradient analysis. *Chinese Journal of Ecology*, vol. 36, no. 1, pp. 205-215.
- [10] Zhang, J.B. Song, G.H, Yu, L, et al. (2018). Identification and characteristics analysis of bottlenecks on urban expressways based on floating car data. *Journal of Central South University*, vol. 25, no. 8, pp. 2014-2024.
- [11] Liu, J. He, Z., Wang, Y, et al. (2017). [Monitoring and analysis on variation of quality for floodwater in Wuhan urban areas in 2016]. *Wei sheng yan jiu = Journal of hygiene research*, vol. 46, no. 6, pp. 956.