Prediction Model of English Major Enrollment in Jiangsu University of Science and Technology

Xia Zhang*, Jing Zhang

Jiangsu University of Science and Technology, Zhenjiang, Jiangsu Province, China

*Corresponding author

Keywords: English major, enrollment, data modeling

Abstract: Enrollment is an important way to improve the quality of college students. COVID-19 has brought great challenges to college enrollment. This article uses the enrollment data of the School of Foreign Languages of Jiangsu University of Science and Technology as reference data. Support vector regression (SVR) is used to comprehensively analyze the previous enrollment data, and the computer data modeling method is used to predict the future enrollment situation. Since the parameter setting of support vector regression (SVR) directly affects the modeling effect, particle swarm optimization algorithm (PSO) is used to optimize the parameters c (penalty factor) and g (radial basis kernel parameter) of SVR, and the optimized model is used to predict the enrollment of students and compared with the SVR model. The experimental results show that both SVR and PSO-SVR are suitable for enrollment prediction, and PSO-SVR has a higher prediction accuracy, which provides an effective basis for admissions office to make future enrollment plan.

1. Introduction

Enrollment is an important work of college enrollment office, the quality of new students plays a role in school running. Under the background of COVID-19 epidemic, if the school of Foreign Languages wants to recruit students from the college entrance examination to meet the enrollment plan, it must do a good job in the early stage of enrollment data analysis and prediction. The school needs to master the admission status of the major and know the distribution of students in each region timely, so as to facilitate the development of future enrollment plan and the arrangement of enrollment publicity

Common prediction algorithms include multiple linear regression [1], BP neural network [2], adaptive fuzzy neural network reasoning system [3], etc. Among them, support vector regression (SVR) is the most widely used prediction algorithms at present, which have been applied in many fields.

2. Analysis of Enrollment Data for English Majors

There are many factors that affect the change in the number of enrollments each year, mainly including national education regulations, the ranking of colleges and universities in the university rankings, the ranking and setting of majors, the promotion of enrollment publicity work, etc. Specific as follows: (1) The country adopts the method that determining the distribution of the enrolment figure according to different situation of each province and admitting according to enrolment cut-off point. The enrolment figure in different regions is not evenly distributed according to the number of candidates, so the number of admissions in different regions will vary [4]. (2) There are several important factors that can affect the choice of candidates, including blindly following the trend to pursuit popular universities and majors and a comprehensive consideration of the strength and location of the university and future employment orientation. (3) In the case of COVID-19, the advantages of attending university in the province are obvious. (4)The promotion of enrollment publicity work and the policy aiming at the attraction of students of universities are different. Also, the opacity of enrollment information will lead to changes in enrollment data [5].

Some intermediaries often take advantage of the asymmetry of information to infringe on the legitimate rights and interests of examinees and breed anomalies in admissions [6].

The data used in this study (the enrollment data of English major in Jiangsu University of Science and Technology from 2014 to 2019) are all from the enrollment system of Jiangsu University of Science and Technology, covering students' names, test numbers and the source places of students. This article mainly analyzes the number of students enrolled in English majors and the source places of students each year. Figure 1 shows the changing trend of the number of English majors in and outside the province from 2014 to 2019.The comparative analysis shows that the English majors enrolled in Jiangsu University of Science and Technology are mainly from Jiangsu province, and the rest of the students are scattered from different regions outside the province. The number of students' source places increased from 18 different provinces (autonomous regions and municipalities) in 2014 to 25 in 2019. As can be seen from Figure 1, the number of students enrolled in Jiangsu province showed a downward trend in the past two years, while the number of students enrolled outside the province tended to be stable at about 60.This indicates that the source places of students of English majors in Jiangsu University of Science and Technology is more diversified, which may be related to the increase in open degree of running a school of Jiangsu University of Science and Technology in recent years and its stronger attraction of students.

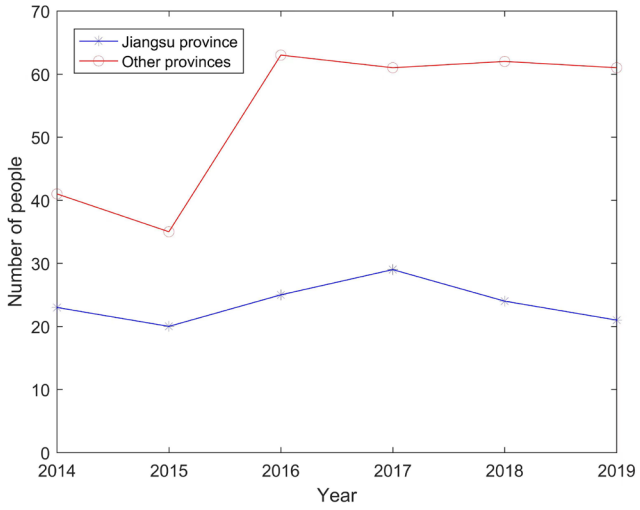

Figure 1 Enrollment of English majors in Jiangsu University of Science and Technology in and out of the province from 2014 to 2019.

3. Support Vector Regression Model

Support vector regression model [7] is based on the improvement and extension of standard support vector machine, which is a form of support vector machine under the quadratic loss function. By constructing the loss function, the quadratic optimization problem of the original support vector machine algorithm is transformed into solving linear equations. Because of its fast solution speed, high-precision function approximation ability, and great advantages in data analysis and pattern recognition, it can be used for classification and regression analysis, so it has been widely applied and developed in various fields.

The idea of SVR modeling is to map the input data from the original space to the high-dimensional feature space through nonlinear mapping, and construct the optimal linear regression function in the high-dimensional feature space to realize the nonlinear regression problem in the original space. This algorithm is based on structural risk minimization, and it can perform well in learning even with a small number of statistical samples.

The factors that determine the performance of SVR algorithm are the selection of penalty factor c and kernel function parameters. In this article, radial basis function(RBF) is selected as the kernel function of SVR algorithm, so the problem is simplified to find the appropriate penalty factor c and

radial basis kernel parameter g, so that SVR has the best learning ability and detection ability. Therefore, this article uses cross-validation to obtain the best c and g, calculates the mean square error (MSE) of cross-validation, and selects the parameter combination with the smallest MSE as the best parameter. The best parameters are used to train the training sample data, and the trained model is used to test the data.

4. Improved Support Vector Regression Algorithm

Particle swarm optimization (PSO) is an intelligent optimization algorithm, which originates from the research on the predatory behavior of birds [8]. The basic idea of particle swarm optimization algorithm is to find the optimal solution through the cooperation and information sharing among individuals in the group. Its advantage lies in its fast convergence speed and no adjustment of many parameters. Support vector regression (SVR) can ideally solve small sample classification, high dimension, nonlinear and regression problems. Support vector regression (SVR) uses kernel functions to map the data in the input space nonlinearly to high-dimensional space to effectively solve nonlinear problems. The prediction accuracy of support vector regression depends on the selection of parameters (c, g) [9].Therefore, this study uses PSO to optimize SVR, and the specific steps are seen in the literature [10].

5. Analysis of Modeling Result

This article takes the enrollment data of English majors in Jiangsu University of Science and Technology from 2016 to 2020 as the input sample set, and the enrollment data of different regions in 2021 as the output sample set. The 34 provinces (autonomous regions and municipalities) were sorted alphabetically, with the first 24 as the training sample set and the remaining 10 as the prediction sample set. MATLAB 2017B was used to edit the program, and the data were calculated and analyzed.

Table 1 Comparison between predicted and actual numbers of enrollment in different regions

region	actual numbers of enrollment	predicted numbers of enrollment (SVR)	predicted numbers of enrollment (PSO-SVR)	region	actual numbers of enrollment	predicted numbers of enrollment (SVR)	predicted numbers of enrollment (PSO-SVR)
Anhui	2	2.03	1.98	Jiangxi	2	1.94	1.91
Macau	0	0.04	0.02	Liaoning	2	2.88	2.05
Beijing		2.11	1.24	Inner Mongolia	0	0.04	0.05
Chongqing	2	1.85	1.96	Ningxia	2	1.86	1.97
Fujian	2	1.97	2.03	Qinghai	0	0.04	0.05
Gansu	2	1.88	2.10	Shandong	3	2.84	2.79
Guangdong	2	1.86	1.96	Shanxi	3	2.85	2.97
Guangxi	2	1.94	1.98	Shaanxi	3	2.76	3.12
Guizhou	3	2.11	2.95	Shanghai	0	0.03	0.02
Hainan	0	0.04	0.05	Sichuan	2	2.18	1.96
Hebei	3	2.89	3.02	Taiwan	0	0.04	0.03
Henan	4	3.72	3.87	Tianjin	2	1.73	2.08
Heilongjiang	2	2.04	1.98	Tibet	0	0.04	0.02
Hubei	4	4.14	4.08	Hongkong	0	0.04	0.04
Hunan	3	2.89	3.06	Xinjiang	0	0.05	0.03
Jilin	2	2.01	2.05	Yunnan	2	1.84	2.08
Jiangsu	21	20.98	21.06	Zhejiang	4	3.63	3.79

Select the radial basis function, and use the cross-validation method to obtain the optimal parameter combination c and g, whose values are 1118.58 and 17.97, respectively. The SVR model

is trained and predicted with the training sample set and the test sample set respectively. At the same time, the parameters of PSO-SVR model are set as follows: the population size is 50, the maximum number of iterations is 100, the learning factors c1 and c2 are 2, the optimization range of c is [0.1, 1000], and g is [0.001, 100].The final prediction results of the enrollment of English majors in 34 different regions of Jiangsu University of Science and Technology in 2019 are shown in Table 1.

It can be seen from Table 1 that the SVR model based prediction model of numbers of enrollment in different regions has good prediction accuracy. The relative error between the predicted numbers of enrollment of English major in Jiangsu University of Science and Technology in 2019 and the actual numbers of enrollment is small, and its accuracy is high. Although SVR model has achieved good results in the prediction of enrollment, the comparison in Table 1 shows that PSO-SVR model has better prediction effect and its accuracy has been improved compared with SVR model. Therefore, the PSO-SVR model can effectively predict the enrollment and can be used as a prediction model for the enrollment of English majors in Jiangsu University of Science and Technology. Using this model can help the School of Foreign Languages to understand the distribution of students in various regions and facilitate the arrangement of appropriate enrollment and publicity work.

6. Conclusion

In this article, SVR and PSO-SVR models are used to predict the enrollment situation of School of Foreign Languages, Jiangsu University of Science and Technology. Both models have achieved good prediction accuracy. The results show that the SVR model optimized by PSO has smaller relative error and higher prediction accuracy, which is suitable for the prediction of enrollment in the case of COVID-19.

References

[1] Han Qianqian, Yang Keming, Li yanru, Gao Wei, & Zhang Jianhong (2021). Svd-anfis hyperspectral prediction model of heavy metal lead content in maize leaves. Spectroscopy and spectral analysis, 41, 6, 1930-1935.

[2] CHEN Nan, FENG Hui-lin, YANG Yan-dong, et al. (2021) Establishment of hyperspectral prediction model for cadmium content in flue-cured tobacco leaves. Journal of Agricultural Resources and Environment, 38, 4, 570-575.

[3] Jia Xueqin, Feng Meichen, Yang Wude, Wang Chao, sun Hui, Wu gaihong, & Zhang song (2018). Study on spectral prediction model of soil water content based on spa-mlr method .Agricultural research in arid areas, 36, 3, 266-269+291.

[4] Zhang Meng, Jiao Hejun (2008). Research on the prediction of college enrollment based on least squares support vector machine. Journal of Zhengzhou Institute of Aeronautical Industry Management, 26, 1, 142-144.

[5] Shao Fengxia. (2019) Analysis of Factors Influencing the Improvement of Student Quality in Colleges and Universities and Suggestions on Countermeasures. Beijing Education (Higher Education), 12, 65-67.

[6] Feng Jiemei. (2021) Countermeasures for College Admissions Publicity under the Impact of covid-19. Beijing Education (Higher Education), 6, 45-47.

[7] Han Xiaohui, Du Songhuai, Su Juan, et al. (2014) Determination method of electric shock current based on parameter-optimized least squares support vector machine. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 30, 23, 238—245.

[8] Sun Ning, Chen Tian, & Xu Gui' an (2020). Particle swarm optimization support vector regression algorithm for air pollution prediction Journal of Shanghai Institute of Electrical Engineering, 25, 1, 7-11+16.

[9] Yan Jinlong, Li Xianli, & Hou Zhoubo (2021). Research on the academic early warning model of higher vocational students Based on PSO-SVR algorithm Jiangsu Science and Technology Information, 38, 15, 57-59.

[10] Xuan Nam Bui, Chang Woolee, Hoang Nguyen, Hoang Bacbui , Nguyen Quoc Long, Qui Thao Le, Van Duc Nguyen, Ngoc Bich Nguyen & Hossein Moayedi. (2019). Estimating PM10 Concentration from Drilling Operations in Open-Pit Mines Using an Assembly of SVR and PSO. Applied Sciences, 9, 14, 2806.