Research on City Water Consumption Forecast Based on SVR Model

Sifu Li, Jie Li*, Yue Zhang

School of Economics and Management, China University of Geosciences, Wuhan, Hubei, 430074, China

*corresponding author

Keywords: Support vector regression, SVR, Urban water use, Simulation analysis

Abstract: Due to the continuous growth of population, rapid economic development and continuous improvement of living standards, the demand for industrial and domestic water in the city has increased substantially, which makes the contradiction between the supply and demand of urban water resources intensified, and the problem of urban water shortage is becoming increasingly serious. The reasonable prediction of urban water consumption is of great significance to the optimal allocation and development of urban water resources. Based on the support vector regression (SVR) method, this paper constructs a prediction model of urban water consumption with MATLAB software. The results show that the model has certain accuracy and applicability, and the prediction results can be used for urban water supply optimization.

1. Introduction

Water resources are irreplaceable resource elements and ecological conditions for human survival, production and development, and it is also a guarantee of safety. my country's per capita water resources are only 28% of the world average. In normal years, the country's water shortage reaches more than 50 billion m^3, and nearly two-thirds of cities suffer from water shortages to varying degrees. With the rapid advancement of industrialization and urbanization and the intensification of the impact of global climate change, my country's water resources problems will become more complex in the future.

The national "Thirteenth Five-Year Plan" outline clearly puts forward the requirements of "implementing the most stringent water resources management system, using water to determine production, using water to determine the city, and building a water-saving society", a major and urgent strategic task. When planning urban water resources, urban water consumption prediction is one of its important basic contents. Changsha is the political, economic, cultural, transportation, and science and education center of Hunan Province. It analyzes the water consumption of Changsha City and has a substantial guiding role in further research on the balance of water supply and demand in the region and the optimal allocation of water resources.

2. Literature Review

The prediction results of urban water consumption not only directly affect the reliability and practicability of water system scheduling decisions, but also directly affect the sustainable use of urban water resources and the sustainable development of social economy.

At present, foreign researchers mostly optimize algorithms for water resource trend forecasting and analysis. G-CHEN et al. proposed a multi-random forest model that integrates wavelet transform and random forest regression (W-RFR) for the prediction of urban water consumption. LCP Velasco et al. used artificial neural networks (ANN) to predict the demand in 8 areas. Domestic researchers have studied the impact of related factors on water consumption in terms of forecasting. Zou Qingrong and Liu Xiuli analyzed the influencing factors of my country's industrial water demand, and constructed a multiple regression model to analyze the relationship between industrial added value, the square of industrial added value and industrial water reuse rate and industrial water consumption. Guo Lei et al. analyzed the negative correlation between total water consumption and regional GDP, constructed a trend model based on the correlation analysis results,

  DOI: 10.25236/iccem.2021.037

and based on economy and population as independent variables. Fu Chuanjun et al. constructed the GM (1,1) gray theory model to predict Haikou's industrial, agricultural, tertiary industry development and population growth trends, and based on this, used the improved BP neural network model to analyze Haikou's The forecast of water resource demand realizes the organic combination of grey theory and neural network.

2.1 Summary

Through combing the literature, the author divides the research methods of scholars at home and abroad into two categories: one is the classical model based on statistics, such as multi-factor regression analysis method, exponential smoothing method, trigonometric function prediction Method ; The other is a prediction model based on machine learning, such as nearest neighbor algorithm, BP neural network, etc. Among them, the first type of research methods are based on mathematical equations, with simple structure and small amount of calculation, which is more convenient in practical applications, but there are problems such as high requirements for data accuracy and poor model generalization ability; the second type of research methods are based on data Mining, the model has high stability and reliability, but requires high data volume, and the model structure is single, which is prone to error gradient accumulation problems.

According to the above-mentioned literature, the author found that the support vector regression method (SVR) and the neural network method have a similar functional structure, but it successfully overcomes the shortcomings of the neural network method in the application of water consumption regression prediction, so it has a wide range of application prospects. Therefore, this paper uses the support vector machine model to predict urban water consumption. The research methods and prediction results can provide reference and data support for the optimal allocation of urban water resources.

3. Research Methods

3.1 Indicator Selection and Data Sources

Water consumption is affected in many ways, and the analysis of factors affecting water consumption in a certain area is very complicated. The water consumption of an area is mainly composed of four parts: urban domestic water, agricultural water, industrial water and ecological water, and each part of the water will be affected by different factors. Due to time constraints and the author's limited ability, this article only starts from the aspect of urban domestic water consumption and analyzes the factors that affect urban water consumption.

The number of people is the factor that most directly reflects the number of people in an area. Obviously, the more populous an area, the higher the domestic water consumption. Residents' disposable income is usually used as an important indicator to measure the level of national economic development, which will have a certain impact on domestic water consumption. In addition, domestic water consumption is directly related to water prices. Therefore, this paper takes Changsha City as the research object, uses support vector regression machine to establish the urban water consumption prediction model, selects the data of water consumption population, residents' disposable income, and domestic water price as sample data to model and analyze the annual water consumption of Changsha City. And evaluate the performance of the prediction model. The indicator data comes from the 2008-2018 Hunan Provincial Statistical Yearbook, Hunan Provincial Water Resources Bulletin, and Changsha City National Economic and Social Development Statistical Bulletin.

3.2 Model Introduction

3.2.1 Basic Principles of Support Vector Regression Machine

SVR (Support Vector Machine Regression) prediction algorithm is the application of the support vector machine principle proposed by Vapnik et al. in data regression prediction. The core idea of SVR is the principle of risk minimization. Its principle is to map the original data set to a new feature space of high order or even infinite dimensions by constructing a nonlinear mapping. A linear function can be found in this space. This linear function can construct a mathematical relationship between the input value and the output value, and the predicted value can be obtained through this function. The basic form of the SVR prediction model is as follows:

$$f(x) = \mu + \sum_{i=1}^{n} \omega_i \varphi(x, x_i) \tag{1}$$

In the formula: μ is a constant; ω_{i} is a coefficient; φ is a basic function.

In the linear regression problem, the insensitive loss function ε is introduced, and the construction of the proxy model becomes the following constrained convex quadratic optimization problem:

$$\begin{cases} \min J = \dfrac{1}{2}|\omega|^2 \\ \text{sub } y_i - \omega x_i - \mu \le \varepsilon \\ \omega y_i + \mu - y_i \le \varepsilon \end{cases} \tag{2}$$

3.2.2 Results Evaluation Method

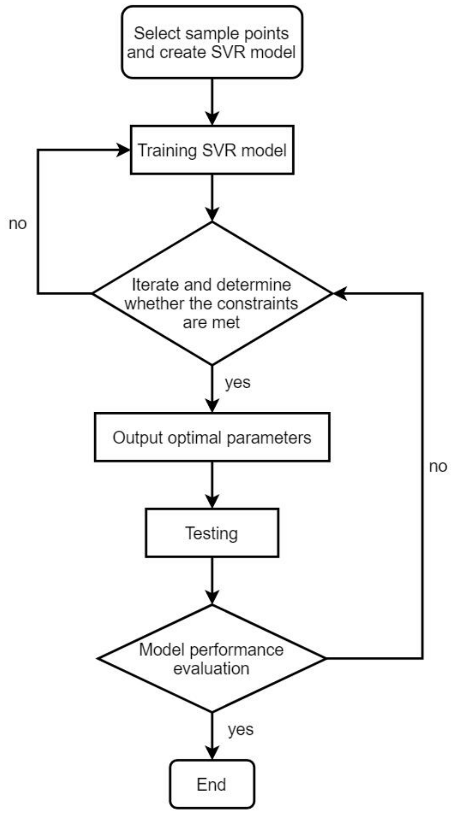
For the evaluation of prediction results, mean square error (MSE) and correlation coefficient R^2 are selected as the evaluation parameters of the pros and cons of the SVR prediction model. The smaller the relative error, the higher the accuracy, and the closer the coefficient is to 1, the better the fitting effect. Calculated as follows:

$$E_{\mathrm{i}} = \frac{\left|\hat{y}_i - y_i\right|}{y_i} (i = 1, 2, \cdots, n) \tag{3}$$

$$R^2 = \frac{(n\sum_{i=1}^{n} \hat{y}_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \hat{y}_i)^2}{\left[n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} \hat{y}_i)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2\right]} \tag{4}$$

3.3 Basic Steps of Modeling

The algorithm flow of the urban annual water consumption prediction model is shown in the following figure:



4. Numerical Experimental Results

This article takes Changsha City as the research object, selects 8 sample data from 2008-2015 as the training set for model training and parameter selection, and selects 4 sample data from 2015-2018 as the test set for comparison of prediction results analyze.

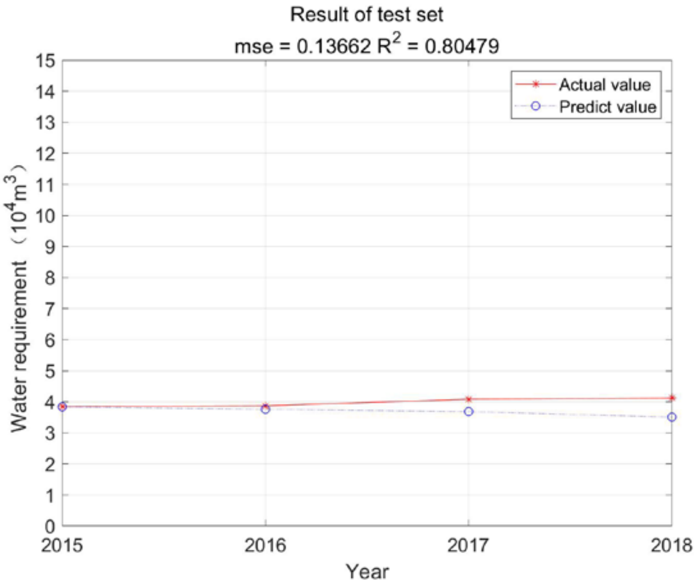


Fig.2 Comparison of Prediction Results of Training Set

Use Matlab simulation software to train the sample data, and see Figure 2 for the simulation results of the training set. It can be seen from the figure that the mean square error of the training set test of the prediction model is 0.0538, and the correlation coefficient is 0.910. The water consumption has a strong correlation with the water use population, the disposable income of residents, and the domestic water price. The comparison between the sample training value and the real water consumption is shown in Table 1.

Table 1 Comparison between Sample Training Values and Real Values

Year	Water use population (10,000 people)	Resident disposable income (RMB)	Domestic water price (RMB / ton)	Real water consumption (M³)	Training water consumption (M³)	Deviation(%)
2008	403.37	17890	1.88	3.69	3.693	0.0007
2009	416	20238	1.88	3.81	3.808	0.0007
2010	476.58	22814	1.88	4.2	4.202	0.0006
2011	485.64	26451	2.58	3.99	3.993	0.0007
2012	495.84	30288	2.58	4.13	3.963	0.0405
2013	509.86	36826	2.58	3.77	3.771	0.0003
2014	528.88	39961	2.58	3.7	3.702	0.0007
2015	552.78	39961	2.68	3.84	3.838	0.0006

Through calculation, it can be seen that the training error of the selected prediction model of the 8 sets of sample values is 0.0405% at the maximum. The error result shows that the accuracy of the prediction model established by SVR is very high. In order to evaluate the performance of the model, 4 sample data of the test set are tested, and the simulation results of the test set are shown in Figure 3.

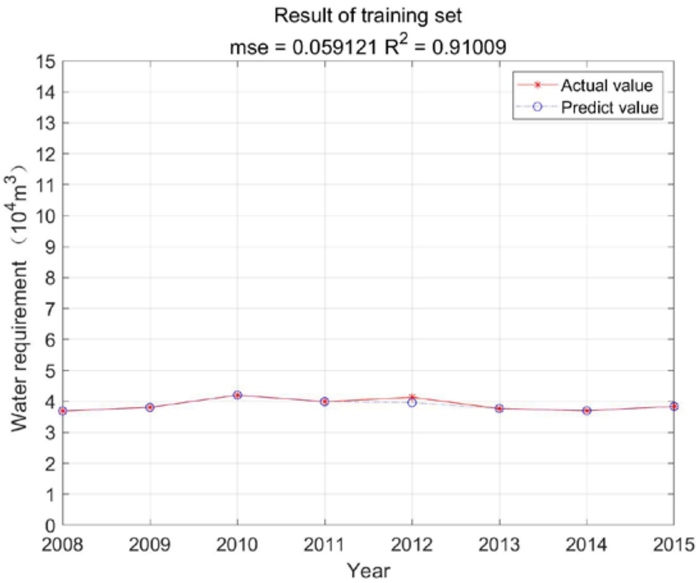


Fig.3 Test Set Simulation Results

The results show that the correlation coefficient of the test set is 0.80479, and the error coefficient is 0.13662. The trend of the predicted value of water consumption is basically consistent with the actual value of the curve. The comparison between the test values and the true values of the four groups of samples is shown in Table 2.

Table 2 Comparison of Test Sample and Real Value and Training Value

Year	Water use population (10,000 people)	Resident disposable income (RMB)	Domestic water price (RMB / ton)	Real water consumption (M³)	Training water consumption (M³)	Deviation(%)
2015	552.78	39961	2.68	3.84	3.838	0.0006
2016	580.97	43294	2.68	3.87	3.758	0.0289
2017	614.38	46948	2.88	4.08	3.684	0.0971
2018	645.23	50792	2.88	4.12	3.506	0.149

It can be seen from Table 2 that the maximum test error of the selected four groups of sample values prediction model is 0.149%. Therefore, it is feasible to use this method to predict the city's annual water consumption.

5. Conclusion

This paper uses the support vector regression model to carry out preliminary research and discussion on the water consumption forecast in Changsha City. The research results show that the forecast model is reasonable, effective and has high forecast accuracy. The optimized scheduling has certain theoretical significance and practical value. It should be pointed out that there are many factors that affect urban water consumption. This article only selects several indicators of water consumption, residents' disposable income, and living water price for analysis, and data collection is limited, which may affect the accurate prediction of urban water consumption to a certain extent. For future work, I hope that more factors related to water consumption can be added to the model to make the model more complete.

References

[1] CHEN G, LONG T, XIONG J, e tal. Multiple random forests modelling for urban water consumption forecasting. Water Resources Management,

[2] JR TMF, ASEFA T, WALKE AG. Short-run water demand forecast accuracy for the Tampa Bay Area (in press). Journal-American Water Works Association, 2016, 108 (3) :126-136.

[3] Zou Qingrong, Liu Xiuli. Research and application of my country's industrial water demand prediction model. Mathematics in Practice and Knowledge, 2015, 45(13): 96-103.

[4] Guo Lei, Huang Bensheng, Qiu Jing, Huang Fenghua. Water demand prediction of the Pearl River Delta urban agglomeration based on trend and regression analysis. Water Resources and Hydropower Technology, 2017, 48(01): 23-28. DOI: 10.13928/j. cnki.wrahe.2017.01.005

[5] Fu Chuanjun, Ma Chao, Xu Dong. Forecast of water resources demand in Haikou City based on grey theory and neural network. Journal of Water Resources and Water Engineering, 2010, 21(05): 96-98+101.

[6] Man Guangsheng, Xu Deqian, Chen Guowei. Research on the prediction method of urban industrial water consumption. Journal of Hefei University of Technology (Natural Science Edition), 2002, (02): 204-207.

[7] He Wenjie, Wang Jizhen, Zhao Hongbin, Han Hongda. Research on the simulation method of urban water consumption in Tianjin. Water Supply and Drainage,2001,(10):43-44.DOI:10.13789/j.cnki.wwe1964.2001.10.011

[8] LiuHongbo, ZhangHongwei.Comparison of the city water consumption short-term forecasting methods.Transactions of Tianjin University, 2002, 8 (3) :211-215.

[9] Yang Fang, Zhang Hongwei, Liu Hongbo. Short-term prediction method of urban water supply load. Journal of Tianjin University, 2002, (02): 167-170.

[10] Yan Xu, Li Siyuan, Zhang Zheng. Application of BP neural network based on genetic algorithm in urban water consumption prediction. Computer Science, 2016, 43(S2):547-550.

[11] Cui Huimin, Xue Huifeng, Wang Lei, Zhao Chenxiao. Water consumption forecast based on BP neural network screened by PCCs-DEMATEL index. Water Saving Irrigation, 2019, (05): 87-91+98.

[12] Tang Fa. Research on support vector machine algorithm based on statistical learning theory. Huazhong University of Science and Technology, 2005.