

Research on Model of Predicting Irrigation Water Requirement Based on Kernel Method

ZHONG Bingxiang

College of Electrical and Information Engineering

Chongqing University of science and technology ,Chongqing, China

zbxcq@126.com

Keywords: Irrigation water requirement; Prediction model; Nonlinear character extraction; Kernel canonical correlation analysis; Support vector machine

Abstract. Scientific irrigation is very important for saving water in agriculture, increasing output and benefit in our country. In this paper a method of nonlinear character extraction based on kernel canonical correlation analysis(KCCA) is presented in which information of soil and environment are input vectors of model. Nonlinear character are extracted by KCCA, then main character variables are determined which reflects the complex relationship between original input and output data and the array dimension of input data is simplified. At last the model based on least squares support vector machine (SVM) were completed. By comparing simulation results, precision and rapidity of the prediction model based on KCCA-SVM are higher than those of CCA-SVM and LS-SCM model. The experimental results show that the method is very effective.

Introduction

Agricultural irrigation is major water user in China, which occupy about 70% of total water requirement. The utilization rate of agricultural irrigation water is only 40%. The appropriate scientific irrigation is to save water and realize increasing production in agriculture. It would become the research focus of agricultural irrigation projects.

Due to irrigation object is a nonlinear and pure time delay system with large inertia, accurate and unified mathematical model can't be established. Most scholars take maximum yield as the objective function, and build model based on the relationship of water demand of crop and crop yields generally[1,2].The method of forecasting irrigation water requirement can be attributed in two categories, one is the method by time series imitation, including neural network model, grey prediction model and combinatorial model[3-6]. Another is through the analysis of water demand of crops, the soil water balance method is researched on the forecast of irrigation water requirement[7]. This model requires a large amount of information of soil and environment as the inputs of the model. So too many factors lead to complexity of model.

In view of the above second kinds of modeling method in which there may be many variables in input space, in this paper kernel canonical correlation analysis method is proposed in which soil and environmental information are the input vector in model. Then the main characteristic variables of the input space are extracted, the secondary factors are eliminated. At last irrigation forecast model is established based on the simplified input space, which can greatly reduce the complexity of the model of irrigation system.

Basic Principle of Kernel Function Method[8,9]

Nonlinear canonical correlation analysis based on the kernel theory utilizes a nonlinear mapping($\varphi(R^n \rightarrow F)$).At first the samples in R^n are mapped into a feature space $F : \varphi\{x\} \in F$, then in the feature space the corresponding linear canonical analysis are completed.

Assume $\{x^1, x^2, \dots, x^N\}$ as N samples of the input variables, $\{y^1, y^2, \dots, y^N\}$ as N samples of the output variables. Accordingly, and $\{\Phi(y^1), \dots, \Phi(y^N)\}$ express the corresponding image in the feature space,

which are treated in centricity, namely $\sum_{i=1}^N \Phi(x^i) = \sum_{i=1}^N \Phi(y^i) = 0$. Assume $\hat{\rho}_F(x, y)$ expresses the experience canonical correlation of x and y , that is based on the experience covariance. It is only need to consider the Φ subspace supported by all sampled data. For any fixed f_1, f_2 , those experience covariance of projection of sampled data in the feature space can be written as:

$$\hat{\rho}_F(x, y) = \text{cov}(\langle \Phi(x), f_1 \rangle, \langle \Phi(y), f_2 \rangle) = \frac{1}{N} \sum_{k=1}^N (\langle \Phi(x^k), f_1 \rangle, \langle \Phi(y^k), f_2 \rangle) \quad (1)$$

Assume S_1 and S_2 express separately the linear space imaged by x^k and y^k . f_1^\perp, f_2^\perp are quadrature component of f_1, f_2 and S_1, S_2 respectively. So f_1, f_2 can be re-expressed as: $f_1 = \sum_{k=1}^N \alpha_1^k \Phi(x^k) + f_1^\perp$ and $f_2 = \sum_{k=1}^N \alpha_2^k \Phi(y^k) + f_2^\perp$, α_1^k, α_2^k are the coefficient vector of feature space.

Therefore, Eq.1 can be written as:

$$\begin{aligned} \hat{\rho}_F(x, y) &= \frac{1}{N} \sum_{k=1}^N \left\langle \Phi(x^k), \sum_{k=1}^N \alpha_1^k \Phi(x^k) \right\rangle \left\langle \Phi(y^k), \sum_{k=1}^N \alpha_2^k \Phi(y^k) \right\rangle \\ &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^N \sum_{j=1}^N \alpha_1^k k_1(x^i, x^k) k_2(y^i, y^k) \alpha_2^j = \frac{1}{N} \alpha_1^T K_1 K_2 \alpha_2 \end{aligned} \quad (2)$$

Here, K_1, K_2 are Gram matrix respectively connected with the data set $\{x^i\}, \{y^i\}$. The typical experience variance are respectively:

$$v_1 = \text{var}(\langle \Phi(x), f_1 \rangle) = \frac{1}{N} \alpha_1^T K_1 K_1 \alpha_1 \quad \text{and} \quad v_2 = \text{var}(\langle \Phi(y), f_2 \rangle) = \frac{1}{N} \alpha_2^T K_2 K_2 \alpha_2 \quad (3)$$

So, get

$$\hat{\rho}_F(K_1, K_2) = \max_{\alpha_1, \alpha_2 \in R^N} \frac{\alpha_1^T K_1 K_2 \alpha_2}{v_1^{1/2} v_2^{1/2}} \quad (4)$$

This is equivalent to implement of KCCA for 2 vectors of N dimension. So KCCA can be converted to the following generalized eigenvalue calculation based on K_1, K_2

$$\begin{bmatrix} 0 & K_1 K_2 \\ K_1 K_2 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \rho \begin{bmatrix} K_1^2 & 0 \\ 0 & K_2^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad (5)$$

If the $\Phi(x_i^k)$ is not be centered, no need to calculate the $\Phi(x_i^k)$, directly from $K_i (i=1,2)$ computing the Gram matrix of centered data: $\bar{K}_0 = N_0 K N_0$. Here $N_0 = I_0 - \frac{1}{N} I$ is the constant singular matrix.

Prediction Model of Kernel Method for Irrigation Water Requirement

Sampled Data Collection. Because irrigation water requirement is affected by soil moisture, soil temperature, air temperature, air humidity, pH, light intensity, soil nutrient, the concentration of CO2 and so on, prediction of irrigation water requirement can be seen as the approximation problem of complex high-dimensional nonlinear function relationship in the time sequence of these parameters. The sample data are pretreated, see Eq.6. Then, the initial time sequence made of these variables constitute as the initial input matrix.

$$X_i^* = \frac{X_i - E(X_i)}{\sqrt{\text{Var}(X_i)}}, i = 1, 2, \dots, n \quad (6)$$

Nonlinear Feature Extraction. Because of large amount of time sequence of these parameters, this will lead to the complex structure of prediction model, and affect prediction time of model. In this paper kernel canonical correlation analysis method is adopted for nonlinear feature extraction. By using the kernel function, the input matrix is transformed into feature space, and the feature extraction is completed in feature subspace with linear CCA method. So the input vectors of prediction model are simplified. The selection of kernel function is related to the accuracy of model. When the structure of system model is unknown, usually Gauss kernel function is better.

Prediction Model on LS-SVM. The input vectors are not be linear correlation by using the method of KCCA. They will be as the sample data, using least squares support vector machine[9], the prediction model of irrigation water requirement will be built.

Simulation Experiment and Results

Before the day of the prediction, soil moisture, soil temperature, air temperature, air humidity, pH, light intensity, soil nutrient, the concentration of CO₂ and other physical parameters are recorded every 2 hour a day. So, 12 groups will be measured every day. Because the weather may be mutated, so it is necessary to consider the conditions of weather. The above data are the input sample data a day.

In this paper 120 groups of sample data are adopted as training samples. In order to have a representative, sample data should be distributed in 12 months. Each set of data includes the soil moisture, soil temperature, air temperature, air humidity, pH, light intensity, soil nutrient, the concentration of CO₂ and other physical parameters every 2 hours a day.

By standardization, sample data constitute the input matrix. The complex correlativity relationship exists among the input data. According to the KCCA principle, the various nonlinear features will be calculated.

For comparison, respectively, the following 3 projects are be adopted in experimentation: (1) experiment was carried out using CCA, and using least squares support vector machine LS-SMV model; (2) using KCCA and LS-SMV model; (3) using LS-SMV model.

Firstly, prediction model of irrigation water requirement is found by training 100 groups of sample based on CCA-SVM, KCCA-SVM and LS-SMV, then tested by another 20 groups of samples for prediction. Model training accuracy and generalization ability are tested by root mean square error (RMSPE). The root mean square error (RMSPE) of predictive value on training data is 2.117%, the RMSPE of predictive value on testing data is 2.554%.

Table 1 shows comparison between the actual water use and predictive value by KCCA model in an irrigation area in 2010. Figure 1 and Table 2 show comparison results among the use of CCA-SVM, KCCA-SVM and LS-SVM three kinds of prediction. The results show that the accuracy of KCCA-SVM model is the best, the accuracy of CCA-SVM and LS-SVM is almost same.

In addition, also taking the time as the metric, calculation volume is compared among three kinds of model mentioned before. Calculation time of CCA-SVM, KCCA-SVM and LS-SVM methods are 3.2s, 3.1s and 4.8s respectively. The data show that the operational time of methods on CCA-SVM and KCCA-SVM are almost same, but less than that of model of LS-SVM.

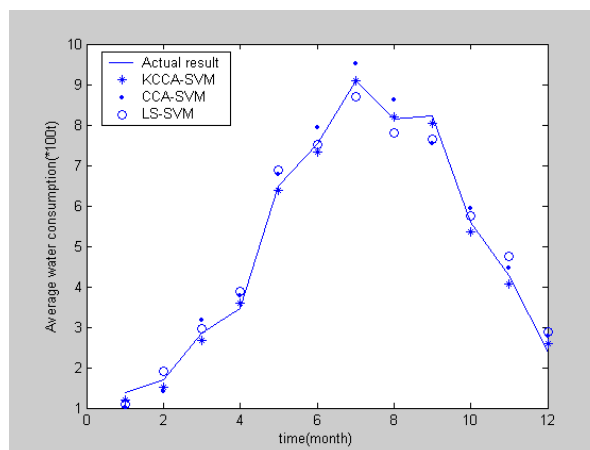


Fig.1 Compared Result of Prediction Value Based on 3 Models

Conclusion

In this paper aiming at the question that too many factor impact on the irrigation water use, this leads to increase of system complexity, so a method for feature extraction of irrigation water requirement based on kernel canonical correlation analysis is presented. The nonlinear feature vector are extracted by KCCA, the complex relationship of original input data are reflected better, and the input data array dimension are simplified. Then based on least square support vector machine regression modeling, prediction of irrigation water requirement is implemented. The simulation results show that the prediction model based on KCCA-SVM has good nonlinear data processing

capacity, high precision of prediction. It is great significance for saving water resources, improving irrigation efficiency by the prediction model.

Table 1 Comparison Between Prediction Data and Real Data in An Irrigation Area in 2010

time	predicting water requirement(10^2)	actual water use (10^2 t)	time	predicting water requirement(10^2 t)	actual water use (10^2 t)
Jan 1	1.378	1.304	Jul 1	8.745	9.003
Feb 1	1.486	1.501	Aug 1	8.052	8.107
Mar 1	2.874	2.765	Sep 1st	7.897	8.034
Apr 1	3.534	3.587	Oct 1	5.656	5.543
May 1	6.325	6.389	Nov 1	4.044	4.158
Jun 1	6.987	7.123	Dec 1	2.356	2.389

Table 2 Comparison Of the Precision of 3 Prediction Methods

time	actual water use(10^2 t)	CCA-SVM		KCCA-SVM		LS-SVM	
		predicting value(10^2 t)	relative error	predicting value(10^2 t)	relative error	predicting value(10^2 t)	relative error
Jan 1	1.304	1.287	0.013	1.378	0.058	1.243	0.047
Apr 1	3.587	3.876	0.081	3.534	0.015	3.321	0.074
Jul 1	9.003	9.145	0.016	8.945	0.006	8.890	0.013
Oct 1	5.543	5.679	0.025	5.656	0.020	5.798	0.046

Literature References

- [1] K.SRINIVA RAJU. Irrigation Planning using genetic algorithm[J]. Water Resources Management, Vol.18(2004) , p.163- 176
- [2] Luo Yi, Lei Zhidong, Yang Shixiu. A conceptual-stochastic model for predicting dynamic variation of water storage in crop rootage[J]. Journal of hydraulic engineering, Vol.31(2008) , p.80-83
- [3] XIE Fang,Tang Deshan. Forecasting Irrigation Water Requirement Based on Least Squares Support Vector Machine[J]. Journal of Anhui Agri.Sci.Vol.38(2010) , p.10273—10275,10288
- [4] Bai Cunyou,Feng Xu, Zhang Shengtang. The Applied Research on Model of Gray Equal Dimension Information in the Forecast of Irrigation Water Use[J]. Journal of Northwest Agriculture and Forestry University: Natural Science Edition, Vol.32(2004) , p 115-118
- [5] Zheng Yusheng,Huang Jiasheng. Prediction of Irrigation Water Use Based on Neural Network[J]. Journal of Irrigation and Drainage, Vol.23(2004) , p.59-61
- [6] Zhao Qiucui, Tang Yanfang, Gu Tuo. Prediction of Irrigation Water Use Based on Parallel Grey Neural Network[J], Journal of Agricultural Engineering, Vol.25(2009) , p.26-29
- [7] YU Huagang, GAo Jun, HUANG Gaoming. Five-Element Planar Cross Array Passive Location Based On Kernel Canonical Correlation Analysis[J]. Systems Engineering and Electronics, Vol.33(2011) , p.1707-1712
- [8] LI De-qi, LIU Chuanling. Feature Extraction Based on Kernel Function and Application of Improved Methods[J].Application Research of Computers, Vol.28(2011) , p 3185-3187
- [9]DENG N Y,TIAN Y J. New Method of Data Mining-Support Vector Machine[M].Beijing:Science Press(2004)