Selection Method and Application of Kernel Parameters

Xunfang LIU
Hunan City University, Yiyang, Hunan, 413000, China

Abstract: The use of relevance vector machine (RVM) Gaussian radial basis kernel function regression model, the relationship between the nuclear parameters and model performance is complex. Aiming at the problem of how to determine the kernel parameters of RVM, put forward a kind of based on the AIC criterion to select kernel parameter of RVM. First, based on Akaike Information Criterion (AIC) thought, come up with a new statistic Q, while the Q as a fitness function. And then use the differential evolution algorithm (Differential Evolution Algorithm, DE) on the kernel parameter optimization, in order to choose determine the kernel parameters. Finally use the algorithm to establish RVM regression model to short-term predict gold price. Experimental results show that the prediction model than the traditional method to establish the model has higher precision and better fitting the generalization ability, and further demonstrate the feasibility and effectiveness of the AIC-based criteria for selecting RVM kernel parameter method.

Keywords: Radial Basis Function, Kernel Parameters, Relevance Vector Machine, Differential Evolution Algorithm, AIC Criterion, Gold Prices

1. Introduction

Gold prices have been proven to be a low-dimensional chaotic time series[2]. Empty space reconstruction phase is the basis of chaotic time prediction, and it can tap hidden in chaotic attractors of evolution. When the phase space reconstruction, the selection of time delay(τ) and embedding dimension (m) is essential and directly affect the prediction accuracy of the subsequent gold price [3]. For the selection of the parameters τ and m, there are mainly two ideas: ①τ and m, respectively, solve alone. Firstly use autocorrelation method, mutual information method to determine τ, then using the G-P method, pseudo-nearest neighbor method to determine m, but only the extraction time autocorrelation linear correlation between sequences. Mutual information method, been able to reflect a variety relationship of holistic systems, but these relationships are only high-dimensional phase space projection in two-dimensional space. It only reflects the independence of the two reconstructed coordinate reconstruction phase space, but can not guarantee that all reconstruction overall independence. In most ways, in order to determine the best τ, you need to determine m, while in order to determine m it needs to determine advance a fixed τ, which will inevitably lead to conflicts[4]. ②τ and m unity solving. Broomhead put forward as the time window method. Kim put forward the C-C method, τ and m unity solved by the relationship between τ and m, so that the gold price time series reconstructed more accurately reflect changes in the gold price trend. So τ , m is often used to solve ideological unity [5]. Advantage of the current gold price prediction algorithm mainly uses neural network algorithm and support vector machines and other algorithms [6,7], because the least squares support vector machine (LSSVM) training speed, good generalization ability.

1.1. Fitness Function

Training of related vector machine is to remove irrelevant points based auto-related decisions (Automatic Relevance Determination, ARD) to obtain sparse model theory. Make the number of correlation vectors obtained by the M training(that is the number of non-zero components ω_i in ω), the obtained RVM model of sparse

$$t = \Phi^* \omega^* + \varepsilon^* \tag{1}$$

Where $\Phi^* = (\phi_{i_1}, \phi_{i_2}, \cdots, \phi_{i_M})$ is the $N \times M$ matrix after Φ deleting not related to the basis vectors in model (1). $\omega^* = (\omega_{i_1}, \omega_{i_2}, \cdots, \omega_{i_M})^T$ is the vector by the components corresponding to a non-zero in w . $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \cdots, \varepsilon_N^*)^T$, $\varepsilon_i^* \sim N(0, \frac{1}{\hat{\beta}})$, $\frac{1}{\hat{\beta}}$ is the variance σ^2 estimated value of noise ε_i , $t = (t_1, t_2, \cdots, t_N)$.

Setting $L = \frac{M}{N}$, then L can approximately represent the relevant vector proportion in the basis vector. Mak R is the correlation coefficient between output vector $t = (t_1, t_2, \cdots, t_N)$ and its prediction vector $\hat{t} = (\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_N)$.

$$R^2=\frac{\left(N\sum_{i=1}^{N}\hat{t}_i t_i-\sum_{i=1}^{N}\hat{t}_i\sum_{i=1}^{N}t_i\right)^2}{\left(N\sum_{i=1}^{N}\hat{t}_i^2-\left(\sum_{i=1}^{N}\hat{t}_i\right)^2\right)\left(N\sum_{i=1}^{N}t_i^2-\left(\sum_{i=1}^{N}t_i\right)^2\right)}$$ set

$SS_{eM}=\sum_{j=1}^{N}\left(t_j-\hat{t}_j\right)^2$ is the residual sum of squares.

Lemma1. In regression model $y=\Phi_q a+\varepsilon$, where Φ_q is $N\times q$ design matrix and the first column of Φ_q is $(1,1,\cdots,1)^T$, $a=\left(a_1,a_2,\cdots,a_q\right)^T$, $\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N)^T$, $\varepsilon_i\sim N(0,\sigma^2)$, $y=\left(y_1,y_2,\cdots,y_N\right)^T$. If $(\Phi_q)^T\Phi_q$ is reversible and $a=\left((\Phi_q)^T\Phi_q\right)^{-1}(\Phi_q)^T y$, then

$\sum_{j=1}^{N}\left(y_j-\hat{y}_j\right)^2=\left(1-R^2\right)\sum_{j=1}^{N}\left(y_j-\bar{y}\right)^2$, where $\bar{y}=\frac{1}{N}\sum_{j=1}^{N}y_j$, $\hat{y}=\left(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_N\right)^T$ is the prediction vector of y, R is the correlation coefficient between y and $\hat{y}$.

Lemma 2. In model (8), $\omega^*=\left((\Phi^*)^T\Phi^*+\frac{1}{\hat{\beta}}A^*\right)^{-1}(\Phi^*)^T t$, where $\frac{1}{\hat{\beta}}$ is the variance σ^2 estimated value of noise ε_i in model(1), A^* is made by A deleting the components tend to ∞.

Proof. Because of ω^* meeting (4), (5) formula, make formula (4) into formula (5) can get it.

Lemma 3. If $N\geq 4$, then exit the positive constant δ, C_1. For each kernel parameter $c<\delta$, make the characteristic root $(\Phi^*)^T\Phi^*$ are no less than C_1.

Proof. For the first column elements of Φ^* is 1 and not all 1, respectively discuss when $c\to 0$, the limited smallest eigenvalue of $(\Phi^*)^T\Phi^*$ and it can be calculated.

The following conventions $N\geq 4$. And then $\frac{1}{\hat{\beta}}\to 0$, by formula (7) we can know $\left\|t-\Phi^*\omega^*\right\|\to 0$, thereby $\min_{\beta}\|t-\Phi\beta\|\to 0$, the σ^2 maximum likelihood estimation $\frac{1}{N}(\min_{\beta}\|t-\Phi\beta\|)\to 0$. The following assumptions model (1) the variance σ^2 of noise ε_i sufficiently small.

Taking into account all the elements of the model establishment RVM thinning process usually located on a threshold value for α_i, the following assumptions the all elements of A^* does not exceed a constant C_2.

Lemma 4. In model (8), when kernel parameter c and $\frac{1}{\hat{\beta}}$ are both sufficiently small, $SS_{eM}\approx\left(1-R^2\right)\sum_{j=1}^{N}\left(t_j-\bar{t}\right)^2$, where $\bar{t}=\frac{1}{N}\sum_{j=1}^{N}t_j$.

Proof. For the first column elements of Φ^* is 1 and not all 1, respectively discussed. By Lemma 1, Lemma 2 and Lemma 3 can get the lemma.

It is similar to the thought of AIC criterion. In model (8), intake $\mathrm{AIC}=N\ln\left(SS_{eM}\right)+2M$. From lemma 4 can get:

conclusion: In model (8), when kernel parameter c and $\frac{1}{\hat{\beta}}$ are both sufficiently small, $\mathrm{AIC}\approx N\ln\left(1-R^2\right)+2M+C$, where $C=N\ln\sum_{j=1}^{N}\left(t_j-\bar{t}\right)^2$.

Since the sample data is given, and therefore the N and C in $N\ln\left(1-R^2\right)+2M+C$ are also given. Thus $N\ln\left(1-R^2\right)+2M+C$ in the model (8) is equivalent to $\ln\left(1-R^2\right)+2\frac{M}{N}$.

Because $\ln\left(1-R^2\right)\approx -R^2-\frac{1}{2}R^4$, $(0<R<1)$, so $\ln\left(1-R^2\right)+2\frac{M}{N}\approx -R^2-\frac{1}{2}R^4+2\frac{M}{N}=2L-R^2-\frac{1}{2}R^4$。

When the statistical error of sample data is small, to take the fitness function for $Q=2L-R^2-\frac{1}{2}R^4$.

1.2. RVM Kernel Parameter Optimization Algorithm based on AIC

Step1: Input sample data. Set the maximum iteration cycles D of DE, population size N_p, scaling factor F and crossover constant CR and set the search range parameters group c. Make iterative algebra $j=0$.

Step2: Randomly generated initial population α_s^0, $(s=1,2,\cdots,N_p)$ within the scope of the parameter c setting. Using RVM obtain prediction value $\hat{t}_i$, $(i=1,2,\cdots,l)$ of t_i. Take the fitness function as $Q=2L-R^2-\frac{1}{2}R^4$. Computing the fitness value of each individual. Record each individual extreme, global extremes and global extreme point.

Step3: Using mutation, crossover and selection of these three operations on the population to be updated, the new fitness value is calculated for each individual populations,

and update each individual extreme, global extremes and global extreme point.

Step4: If $j < D$, then the iterative algebra $j \leftarrow j+1$, turn to step3. Otherwise, the output of the global extreme point c^*.

Step5: Establish RVM model using the parameters c^*.

In this algorithm, the vast majority are concentrated in the calculation processing of fitness, and the time complexity of the RVM trained is $O(N^3)$, so that the time complexity of the algorithm is $O(N_p * N^3 * D)$.

2. Experiments and Analysis

2.1. Predicted Results and Analysis of Different Fitness Function

Mean absolute percentage error currently used (mean absolute percentage error, MAPE) as the fitness function kernel parameters for RVM optimization [9] [10] literature.

The calculation formula is: $\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|t_i - \hat{t}_i|}{t_i}$, which, $\hat{t}_i$ is the actual gold price of the *ith* sample, $\hat{t}_i$ is the precited gold price of the *ith* sample. Use in the training set S in 4.1 and above MAPE as the fitness function established RVM regression model (MAPE-RVM) to predict the test set T in 4.1.

Figure 1 by were obtained fitting the first 1-183 result of the price of gold and section 184-188 of the gold price forecast results by MAPE-RVM model established.

From the Figure 1 can know that the training set performance is better, but known from Figure 1 poor performance on the test set, resulting in a large error when the first 184-188 in the gold price forecast, the prediction accuracy are shown in Table 1 and Table 2. Not only that, if the other samples from 1-188 in five randomly selected for testing set, the rest of the samples for the training set, MAPE-RVM model has a similar situation, like the election of the first 27, 29, 133,154,167 samples for the test set. MAPE-RVM model for the rest of the samples for the training set predictions established for the training set: the mean square error is 1.4637e-006, the average absolute error is 4.5137e-004, and the average relative error (%) is 3.3899e-005. Prediction results on the test set as follows: mean square error is 9.3404e + 003, the average absolute error is 86.1200, the average relative error (%) is 6.09. MAPE is visible to the fitness function of MAPE-RVM model to establish the first 1-188 sample sample space has made this performance, indicating that the model appeared over-fitting, generalization ability is poor.

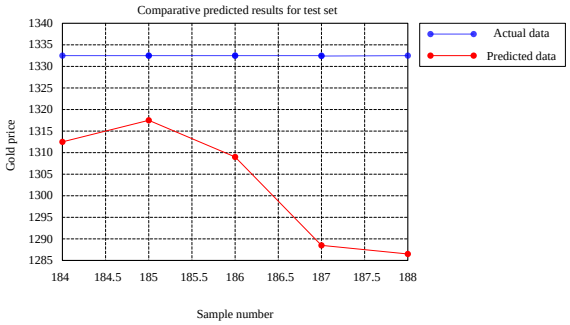

Figure 1. Comparative predicted results for the gold prices of issues No. 184-188.

2.2. Comparative Analysis of Similar Models

Compared with the accuracy and prediction results of differential evolution algorithm based on support vector machine model (DE-SVR), multi-dimensional gray model (GM (1,6)) , we can see that AIC-RVM resulting prediction accuracy is better than the above model. Specific results are shown in Table 1, Table 2. With precision and prediction effect of wavelet neural network model comparison found that the wavelet neural network appeared over-fitting in predicting gold prices, generalization ability is poor. Specific results are shown in Table 2.

Table 1. Accuracy Comparison of Predicted Results for the Gold Prices of Issues No. 1-183.

prediction method	mean square error	mean absolute error	mean relative deviation(%)
DE-SVR	578.4992	20.3887	1.50
GM(1,6)	5.7676e+004	159.2221	11.39
wavelet neural network	59.8627	4.8189	0.36
MAPE-RVM	2.9137e-005	9.9233e-004	7.4437e-005

Table 2. Accuracy Comparison of Predicted Results for the Gold Prices of Issues No. 184-188.

prediction method	mean square error	mean absolute error	mean relative deviation(%)
DE-SVR	1.6555e+003	37.7839	2.91
GM(1,6)	2.9994e+005	543.6824	41.78
wavelet neural network	5.4895e+004	209.1826	16.06
MAPE-RVM	1.0390e+003	29.5498	2.28
AIC-RVM	32.4319	3.2694	0.25

Comprehensive above, we can know that the RVM kernel parameter optimization algorithm based on AIC (AIC- RVM) set up regression model has high precision of fitting results, good generalization ability. Its overall performance is superior to the traditional forecasting model.

3. Conclusion

When the noise variance of the sample data is small, by choosing $Q = 2L - R^2 - \frac{1}{2}R^4$ as a fitness function, the use of differential evolution algorithm optimize the RVM kernel parameter to improve the fitting accuracy and increase sparse of RVM. It reduced the balance of computational complexity, and reduced the man-made interference factors in the process of fitting and over fitting to happen that improve the generalization ability of the model. Using established RVM regression model by this method predicts the gold price. Its simulation experiments show that on the one hand that the regression model generalization ability of RVM model is stronger than the choice of the fitness function used as a mean absolute percentage error. On the other hand the regression model prediction accuracy is better than differential evolution algorithm based on support vector machine model, multidimensional traditional forecasting model gray model. This shows that the idea of using the AIC nuclear parameters to determine a reasonable method is an effective method.

References

[1] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects Across Visual Sensor Networks. Journal of Multimedia, Vol 7, No 1 (2012) pp. 2-8

[2] Phakpoom T. Clustering and Industrial Development: Evidence from Thailand. Nagoya University, 2011.

[3] Kwak S J,Yoo S H,Chang J I. The role of the maritime industry in the Korean national economy: an input-output analysis. Marine Policy, 2011, 29(4): 371-383.

[4] Labaj M. Qualitative input-output analysis and national innovation system in Slovakia. International Journal of Transitions and Innovation Systems, 2011, 1 (2): 105-116.

[5] Lahr M L,Mesnard L D. Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis. Economic Systems Research, 2009, 16 (2): 115-134.