Research on Extreme Financial Risk Early Warning Based on ODR-ADASYN-SVM

Shuanglian Chen[1,a]

[1]Guangzhou University, Guangzhou, Guangdong, 510006, China

[a]15818707232@163.com

Abstract. this paper uses index of Shanghai and Shenzhen 300 as research object, it will combines with ODR, ADASYN and traditional SVM, it puts forward one kind of improved SVM model——ODR-ADASYN-SVM model to predict financial market extreme risk in China, and it also makes evaluation on precision, stability of risk early warning for this model, which has greatly enhanced unbalance sample learning ability of SVM and effectively overcome over-fitting of SMOTE, represents the superior extreme financial risk prediction ability, so it has certain practice and application value.

Introduction

In today of fluctuated and unsafe financial market, establishing scientific risk early warning model can effectively predict and prevent financial market risk, maintain financial economy safety, promote harmony and stability of economy and society, so it has become to be one hot problem that theory circle and practice circle pay high attention to it. Through analysis and cognition of model research for every kind of financial risk early warning in the past, this paper uses Shanghai and Shenzhen 300 index(China securities index 300, CSI300) as research object, it combines with ODR, ADASYN with SVM to construct new ODR-ADASYN-SVM model to predict extreme risk of financial market.

SVM model and computation theory

SVM appeared in the 1990s and it is the neural network model put forward by Vapnik etc to make research on small sample, small probability event. Suppose this sample assemble set (y_1, x_1), (y_2, x_2) …, (y_m, x_m), and $y \in \{-1,1\}$、 $x \in R^n$, input sample data into SVM model on problem optimization, it can be indicated as follows:

$$\min \frac{1}{2} w^T w + c \sum_{i=1}^{n} \xi_i, \quad s \cdot t \cdot \begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \quad （1）$$

In formula （1）, i=1,2, …, m; w indicates weight vector; ξ is slack variable; c＞0 is penalty parameter. It can convert the above-mentioned optimization problems into dual problems through Lagrangian method, its calculation equation is as follows:

$$\max w （\alpha) = -\frac{1}{2} \sum_{i, j=1}^{m} \alpha_i \alpha_j y_i y_j K （x_i, x_j) + \sum_{i=1}^{m} \alpha_i$$

$$s \cdot t \cdot \begin{cases} c \geq \alpha_i \geq 0 \\ \sum_{i=1}^{k} \alpha_i y_i = 0 \end{cases} \quad （2）$$

Here, the same i=1,2, …, m; $\alpha = \{\alpha_1, \alpha_2, …, \alpha_m\}$; K （$x_i$, x_j) is kernel function. When K is positive definite kernel, it has the optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, …, \alpha_m^*)$ and can get the following:

$$\begin{cases} w^* = \sum_{i=1}^{m} \alpha_i^* y_i x_i \\ b^* = y_i - \sum_{i=1}^{m} y_i \alpha_i^* K(x_i, \ x_j) \end{cases} \quad （3）$$

Take this optimal solution to the classified function equation and it can realize classification on this sample through equation (4).

$$f（x）= sign\left[\sum_{i=1}^{m} \alpha_i y_i K（x_i, \ x_j）+ b\right] \quad （4）$$

The established extreme risk early warning model of financial market based on SVM theory

Suppose financial market is i, there are 2 kinds of condition of having extreme risk at time t and without having extreme risk. In calculation, the used sample data is composed of several sample points, sample point is composed of 2 parts, one is characteristic index vector $x_d^i = (x_1^i, \ x_2^i, \ldots, \ x_n^i)$ of n-dimension and d=1.2, …, the other one is corresponding output index, which is state index variable $y^i \in \{+1, -1\}$. Based on this, it can be known that index vector of financial market i at time t constructs one sample point $(x_{t, \ d}^i, \ y_t^i)$. In making research on extreme financial market early warning, after obtaining sample set $(x_{k, \ d}^i, \ y_h^i)$, it can predict the corresponding state index variable at time t according to characteristic index vector at time t.

On establishing early warning model of SVM extreme financial risk, assemble sample data into m time length, its sample data $(x_{j, \ d}^{(i)}, \ y_p^{(i)})$ is used as training ser, $(x_{k-j, \ d}^{(i)}, \ y_{h-p}^{(i)})$ is used as test set and j=1,2, …, m; p=2,3, …, m+1, m＜t-1, it can get the model：

$$y^{(i)} = sgn（f（x）） \quad （5）$$

By combining with calculation theory of, it can get solution and optimal classification function:

$$f(x) = sgn（w^* x + b^*）= sgn（\sum_{j=1}^{m} \sum_{p=2}^{m+1} \alpha_j^* y_p^{(i)} exp（-\gamma \left| x_{j, \ d}^{(i)} - x^{(i)} \right|^2）+ b^*） \quad （6）$$

While in formula （6）, $b^* = y_q^{(i)} - \sum_{j=1}^{m} \sum_{p=2}^{m+1} \sum_{s=1}^{m} y_p^{(i)} \alpha_j^* exp（-\gamma \left| x_{j, \ d}^{(i)} - x_{s, \ d}^i \right|^2）$. This it completes establishment of extreme financial risk early warning model on SVM.

Construction of ODR-ADASYN-SVM extreme financial risk early warning model based on SVM

According to 2 kinds of situations of having extreme financial risk and without having extreme financial risk, suppose S_{maj} is sample set of non-extreme financial risk, S_{min} is sample set of extreme financial risk; x is the sample in S_{maj}. This it can calculate the sample amount of non-extreme financial risk Num：

$$Num = (\left| S_{maj} \right| - \left| S_{min} \right|)\alpha \quad （7）$$

Of which, $\alpha \in [0,1]$, which represents unbalance level of sample, when $\alpha = 1$, which represents that 2 kinds of sample amount reaches complete balance after deleting the designated non-extreme financial risk. And then it gradually obtains the association set of every sample of non-extreme financial risk, it makes classification and uses withx to record classification number, withoutx is the nearest sample of all the samples in association set and deletes sample x, record the added

classification number of in K+1 nearest sample. Through withx-withoutx, it can get the difference value diff, and judges relations of diff and 0 according to the following formula, and it uses this to define nature of sample x.

$$x \in \begin{cases} N & \text{diff} < 0 \\ B & \text{diff} = 0 \\ S & \text{diff} > 0 \end{cases} \quad （8）$$

In the above formula, N represents the noise sample; B and S respectively represent boundary sample and safety sample. According to sample nature, it keeps the boundary sample in S_{maj} and deletes noise sample, while safety sample needs to calculate the Euclidean distance d among the nearest extreme financial risk. Afterwards it makes sequencing from big to small, it deletes from the first one, the amount to non-extreme financial risk until it is reduced to the designated amount $|S_{maj}| - Num$, and it gets sample set of non-extreme financial risk, it is S_{newmaj}。

It calculates the newly generated extreme financial risk sample amount Num_{new}, according to the sample amount of non-extreme financial risk, when the synthetic amount turned into artificial extreme financial risk, 2 kinds of sample amount reaches complete balance.

$$Num_{new} = |S_{newmaj}| - |S_{min}| \quad （9）$$

It finds out K nearest sample for extreme financial risk sample x_i and it can get the nearest non-extreme financial risk proportion r_i according to formula $r_i = \dfrac{\Delta_i}{K}$. So that it can obtain distribution function $\hat{r}_i$ according to formula $\hat{r}_i = \dfrac{r_i}{\sum\limits_{i=1}^{|S_{min}|} r_i}$, at the same time it meets $\sum\limits_{i=1}^{|S_{min}|} \hat{r}_i = 1$. And then it applies SMOTE method to synthesize g_{iex} extreme financial risk sample, make 2 kinds of sample amount reach the same and it is $g_i = \hat{r}_i Num_{new}$。

Through the above-mentioned steps, it can use ODR-ADASYN balance sample of non-balance sample, and then it uses SVM to make training for it, uses SVM model obtained by test sample to make performance test and evaluation.

Demonstration result and analysis of ADR-ADSYN-SVM model

This paper adopts Shanghai and Shenzhen 300 index（CSI300）and samples by 31st, March of 2015 as research object, it lies in that CSI 300 is regarded as the barometer of overall trend of Shanghai and Shenzhen market, which includes about 60% market values, making risk early warning has certain representation and its persuasion is much stronger. It makes calculation on threshold value according to every kind of method, it chooses average value -0.0260 as standard, those exceed standard is extreme financial risk sample, those do not exceed are extreme financial risk sample, they are used to confirm state index variables of all the samples. It makes test on threshold standard value through rear GPD curve and experience fitting effect drawing. From the following figure 1, extreme and non-extreme financial risk sample percentage is 1:20.
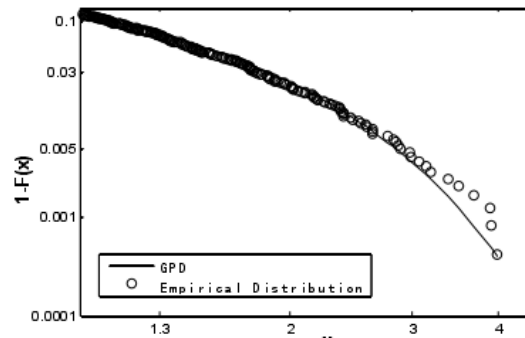
Figure 1 Effect drawing indication diagram of rear experience distribution and GPD distribution fitting for CSI 300 standard

On choosing and extracting characteristic variable, it confirms 8 representative international financial market stock daily benefits as external characteristic index, which is indicated by the following table:

Table 1 Characteristic index of extreme financial risk early warning model

variable	external characteristic index name	variable	Internal characteristic index name
X1	Daily return of HSI	X9	Opening price
X2	Daily return of KOSPI	X10	Opening price
X3	Daily return of TWII	X11	Maximum price
X4	Daily return of GSPC	X12	Minimum price
X5	Daily return of NASDAQ	X13	Change percent
X6	Daily return of Nikkei 225	X14	Change percent
X7	Daily return of DJIA	X15	Volume
X8	Daily return of FISE 100	X16	Turn volume

In order to extract the external characteristic index which can reflect the extreme risk of new financial risk in China, it calculates the low tail dependence coefficient between CSI 300 and every country (area), which is indicated by table 2.

Table 2 Low tail dependence coefficient of CSI 300 and the rest of benefit

	HSI	KOSPI	TWII	GSPC
CSI300	0.3027	0.1155	0.1058	2.4339e-05
	NASDAQ	DJIA	Nikkei225	FTSE100
	4.9092e-05	1.8648e-05	0.0530	0.0153

From the above table, it can obviously see that because HIS, KOSPI, TWII are bigger than 0.1, which indicates that correlation of both is higher, choosing it as the final external risk characteristic index is relatively reasonable. It sets SVM core function as RBF, penalty function C=0.5, core function parameter γ is 0.5, it uses fold cross validation method to make discussion on the nearest parameter of ODR and ADASYN, the result is indicated by the following figure 2 and 3:
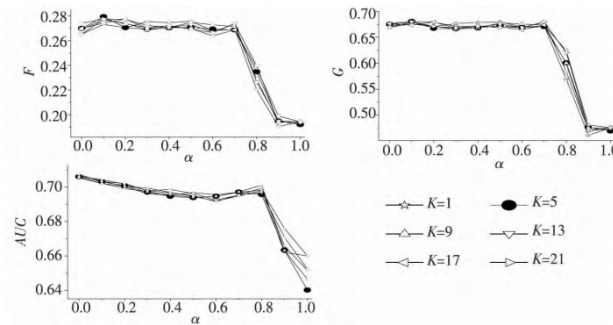
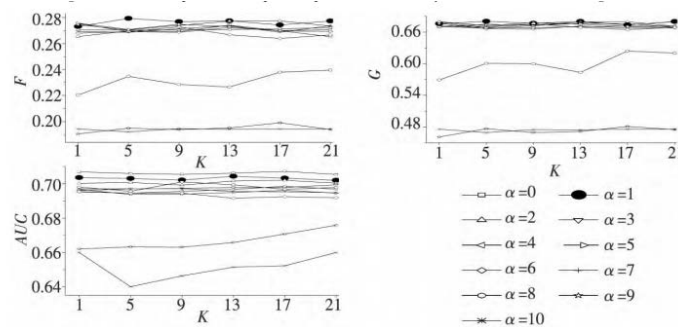Figure 2 Effect of parameter α on model prediction precision under different nearest parameter K

Figure 3 Effect of nearest parameter K on model prediction precision under different parameters α

According to the above figure 2 and 3, it can be known that choosing proper α is very important to construct ODR-ADASYN-SVM extreme financial risk early warning model with the optimal precision, and prediction precision is very high when α is 0.1. Therefore, by adjusting parameter α can get the optimal extreme financial early waning model, prediction performance and stability is obviously superior to SVM and ODR-SVM model.

Conclusion

In a word, ODR-ADASYN-SVM model can effectively enhance the prediction precision of traditional SVM model on extreme risk in Chinese financial market, it introduces into ODR and ADASYN to combine with SVM, which can effectively improve SVM, increase prediction performance and stability, correctly predict extreme risk of Chinese financial market, which indicates high practice value and have certain auxiliary effect on supervision and control of financial market, it is favorable for healthy operation of macro-economy.

References

[1] Tao Xinmin, Tong Zhijing etc. Unbalance Data SVM Classification Algorithm Based on ODR and BSMOTE [J].Control and Decision, 2011,26（10）:1535-1541

[2]Chen Wang, Wei Yu etc. Research on Risk Induction Effect between Chinese Stock Market and Surrounding Stock Market Fluctuation [J].Management Science of China, 2011,19（6）:31-39

[3]Xue Wei. Improved SMOTE Re-sampling Algorithm of Non-balance Data Set [J].Statistics Research, 2012,29（6）:95-98

[4]Xu Guoxiang, Yang Zhenjian. Construction and Application Research of PCA-GA-SVM Model——Demonstration Analysis on Prediction Precision of Shanghai and Shenzhen Index [J]. Asia and Pacific Journal of Finance, 2011（2）:135-147

[5]Lin Yu, Huang Xun, Xu Kai. Research on Extreme Risk Early Warning of Financial Market Based on RU-SMOTE-SVM, 2013（4）:15-20

[6]Chen Ying, Wu Zhiwei, Wang Yang. Multi-market Transaction and Price Discovery of Derivative Securities for Shanghai and Shenzhen 300 index [J]. Journal of Management Science, 2014,17（12）:75-84

[7]Lin Yu, Huang Xun, Chun Weide etc. Research on Early Warning of Extreme Financial Risk Based on ODR-ADASYN-SVM [J]. Journal of Management Science, 2016,19（5）:87-100

[8]Huang Xun. Research on Improved SVM Intelligent Early Warning of Chinese Extreme Financial Risk [D]. Master Thesis of Chengdu University of Technology, 2015