Second International Workshop on Knowledge Discovery and Data Mining

Study on the Risk Prediction of Real Estate Investment Whole Process Based on Support Vector Machine

Wanqing Li
School of Economics and Management
HeBei University of Engineering
Handan, Country
zhao.y001@163.com

Wenqing Meng
School of Civil Engineering
HeBei University of Engineering
Handan, China
chnzhy@sohu.com

Yong Zhao
School of Economics and Management,
HeBei University of Engineering
Handan, China
tiegong2002@163.com

Shipeng Xu
School of Economics and Management
HeBei University of Engineering
Handan, China
x.z521@live.cn

Abstract—**With the rapid development of real estate, the risk of investment is also increasing rapidly. So the risk of predicting and controlling the real estate investment has become the key to the success or failure of the project. In this paper, a support vector machine (SVM) modeling approach for real estate investment risk prediction is proposed at first, which is made use of its merits of structural risk minimization principle, the small study sample and non-linear to analyze the risk factors during investment every stage in real estate projects, then a model based on support vector machines in real estate investment risk is built up, at last, an example is given to prove that this model is effective and practical. All these are used of providing useful help of the future of real estate investment risk control and management.**

Keywords-Support Vector Machine; real estate investment; risk ; prediction

I. INTRODUCTION

As the real estate investment projects has some features such as large, long, complex influence factors, real estate investment risks factors is complex, and risk factors often cause serious consequences. In order to ensure the security of real estate investment and earnings results, it is necessary to do scientific analysis and forecasting for its potential risks during the investment processing, which will make the investment decision-making more scientific and rational, and bring an effective risk prevention and control at the same time in order to reduce the loss brought about by the risks.

Risk analysis and prediction is the first step of project risk management. The main common methods include artificial neural networks and genetic algorithms and so on, and these methods have better self-learning and adaptive capacity, but other issues such as network structure identifying and stocks selecting is obvious, and requiring a large amount of training samples in order to guarantee better results forecast, additionally, it is easy to sink into a local maximum, all of these cause poor generalization. The support vector machine (Support Vector Machine) is based on statistical theory, which is a new machine learning method for small data sample.

II. SUPPORT VECTOR MACHINE AND THE PRINCIPLE OF REGRESSION ALGORITHM

A. *Principles of Support Vector Machines*

Support Vector Machine (Support Vector Machine, referred to as SVM) is the machine learning algorithm based on a statistical learning theory SLT (static learn theory) which was first put forward by Cortes and Vapnik in 1995.It's characteristics are structural risk minimization theory's using, no local minimum problem, no relationship between the complexity of the calculations and the dimensions of the enter sample ,a very strong generalization ability and more obvious advantages in dealing with small samples and high-dimension problems than other algorithms. Thus in recent years, Support vector machines have been widely used to classification and regression problems.

B. *Support Vector Machine Regression (SVR)*

ε-insensitive function and kernel algorithm is the main base of SVR. The using of nuclear function is to avoid possibility of "dimension disaster" by dimension-lifting. In this paper, through the use of insensitive function, non-linear SVR solving process is as follows:

Given a set of training data $\{(x_i, y_i), i = 1,2,\cdots l\}$, in which $x_i \in R^d$ is the ith imported value of study samples. Meanwhile, it is a d-dimensions column vector $x_i = [x_i^{\ 1}, x_i^{\ 2} \cdots x_i^{\ d}]^T$, $y_i \in R$ and is the corresponded target value. First of all, a non-linear ε-insensitive function is defined as:

$$L_\varepsilon(y) = | f(x) - y |_\varepsilon = \begin{cases} 0 \\ | f(x) - y | - \varepsilon \end{cases} \quad (1)$$

$$\text{for} \quad | f(x) - y | < \varepsilon$$
$$\text{otherwise} \quad | f(x) - y | \geq \varepsilon$$

978-0-7695-3543-2/09 $25.00 © 2009 IEEE
DOI 10.1109/WKDD.2009.40

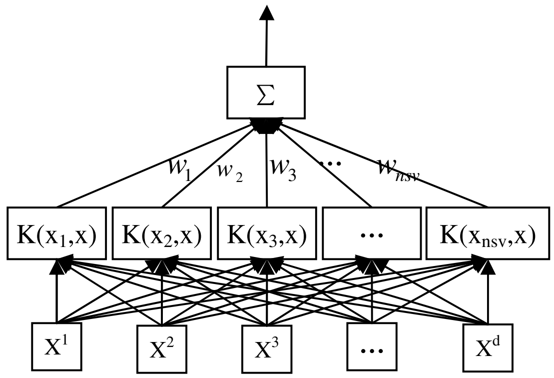

Figure1. SVM predication structure

The following is the standard SVR algorithms learning method for the sample set of the case of non-linear regression to estimation (Figure1):

Considering the non-linear ε -insensitive function:

$$f(x) = \omega \cdot \phi(x) + b \tag{2}$$

$\omega \in R^d$ is weighting vector, $b \in R$ is threshold, $(\cdot)$ represents inner-product. Structural risk minimization principle is used to solve the original loss of function to minimize the risk. That is, under the premise of madding ω and b met the ε -insensitive function, the minimum of $\frac{1}{2}\omega^T\omega$ can be realized as while slack variable ξ_i, ξ_i^* can be introduced. So that the original problem is changed into:

$$\min . \frac{1}{2} \|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{3}$$

$$s.t. \begin{cases} y_i - \omega \cdot \phi(x_i) - b \le \varepsilon + \xi_i \\ \omega \cdot \phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1,2,\cdots,l \end{cases} \tag{4}$$

Lagrange has been through the transformation of its optimization of the dual form:

$$\max_{\alpha,\alpha^*} .[-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i \cdot x_j) - \sum_{i=1}^{l}\varepsilon(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)] \tag{5}$$

With constraints, $$s.t. \begin{cases} 0 \le a_i \le C/l, \quad i = 1,\cdots l \\ 0 \le a_i^* \le C/l, \quad i = 1,\cdots l \\ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \end{cases} \tag{6}$$

Where $K(x_i \cdot x_j) = \phi(x_i) \cdot \phi(x_j)$ is the kernel function and the estimate function is as follow:

$$y = \sum_{i=1}^{nsv} w_i k(x_i, x) + b \tag{7}$$

C is a positive constant, which is called penalty parameter. In order to solve the above problem, making corresponding training sample for $(a_i - a_i^*) \neq 0$ is supporting vector, besides assigning a set of value of $w = \sum_{i=1}^{l}(a_i - a_i^*)\phi(x_i)$, and the calculated formula is gained:

$$f(x) = \sum_{x_i \in sv}(a_i - a_i^*)k(x_i, x_j) + b \tag{8}$$

III. RISK PREDICTING ALGORITHM FOR REAL ESTATE INVESTMENT BASED ON SVM

The basic thought of real estate investment risk forecast model which based on the return model is: taking invests various stages of the real estate the risk target characteristic value as the support vector machines input vector while taking the forecasting of the real estate investment plan risk value as the support vector machines output vector. Using the loss of un-sensitive function with the definition train support vector machines, after compatible study, we can obtain a series of weights and coefficient value of relating different support vector, after the satisfying training requested, the forecast model is established.

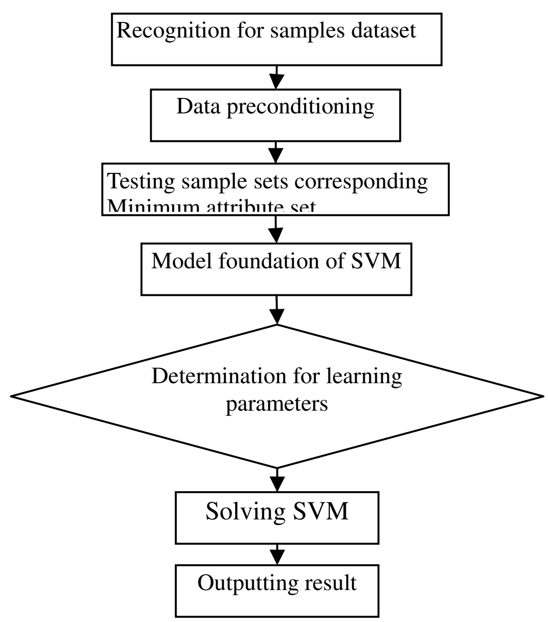

Figure2. Calculation procedure of real estate investment risk based on SVM

Support vector machine learning training samples are usually based on project proposals, project feasibility study report, project evaluation reports, project evaluation report, and other information gathered. If the collected samples of the model dimension of the targets are different and the difference between the values can be great. Features a direct indicator of the value of the input to support vector machine learning, is likely to reduce support vector machine learning performance, even after the completion of the study can not effectively use it to predict. Therefore, when using the application of support vector machines regression for the project investment risk analysis, the first we need is to normalize the indicators of training samples. That is, the linear stretch targets to [0, 1] interval. The following Figure2 are given SVM project investment risk calculation process.

IV. AN EXAMPLE OF PREDICTING FOR PROJECT RISK BASED ON SVM

In order to verify SVM in real estate investment risks in the effectiveness of the application and use, applied to real estate projects in risk prediction with BP neural network risk assessment of the results. Data province select 15 real estate investment projects, in which 12 samples for training, 3 samples for testing.

A. *To determine the factors*

To risk factors as the main basis for the real estate investment projects to be undertaken to consider comprehensive evaluation of the main factors are:

- During the investment decision process, the risk includes development opportunity risk A_1, risk of regional economic environment A_2, risk of regional social environment A_3, risk of project positioning A_4.
- During the land to obtain process, the risks includes risk of market supply and demand B_1, risk of development cost B_2, risk of financing B_3, risk of levy land and remove B_4.
- During the item construction process, the risk includes risk of project quality C_1, risk of project duration C_2,risk of development cost C_3, risk of contracting C_4, risk of project technology C_5,risk of construction claim C_6, risk of natural conditionsC_7,risk of contract mode C_8.

TABLE I. IMPACT OF FACTORS FOR RISK PREDICTION OF REAL ESTATE INVESTMENT WHOLE PROCESS

Impact of Factors for Risk Prediction of Real Estate Investment Whole Process																					
Project code	*Investment decision process*				*Land to obtain process*				*Item construction process*								*Rent and sale management process*				*Risk value*
	A_1	A_2	A_3	A_4	B_1	B_2	B_3	B_4	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	D_1	D_2	D_3	D_4	
1	0.7	0.6	0.7	0.5	0.7	0.6	0.5	0.6	1.0	0.5	0.4	0.3	0.4	0.3	0.5	0.4	0.7	1.0	0.4	0.5	0.565
2	1.0	0.7	0.6	0.5	0.8	0.9	0.7	0.8	0.9	0.7	0.6	0.2	0.5	0.3	0.7	1.0	1.0	0.7	0.3	0.6	0.675
3	0.5	0.3	0.3	0.5	0.9	0.4	0.2	0.3	0.7	0.6	0.4	0.4	0.3	0.2	0.4	0.3	0.7	0.7	0.3	0.3	0.435
4	0.7	0.5	0.7	0.3	1.0	0.5	0.4	0.7	0.6	0.3	0.3	0.1	0.5	0.3	0.7	0.3	1.0	0.8	0.9	0.3	0.545
5	1.0	0.4	0.7	0.6	0.6	0.5	0.3	0.4	0.6	0.5	0.3	0.3	0.4	0.4	0.4	0.3	1.0	0.8	0.3	0.3	0.505
6	0.9	0.6	0.8	0.7	0.8	0.9	1.0	0.5	0.4	0.8	0.7	1.0	0.9	0.3	0.6	0.7	0.5	0.9	0.6	0.7	0.715
7	0.5	0.4	0.7	0.4	0.6	0.3	0.3	0.2	0.7	0.5	0.3	0.2	0.3	0.1	0.4	0.4	0.7	0.7	0.4	0.4	0.425
8	1.0	0.8	0.9	0.8	0.9	0.7	0.9	0.8	1.0	0.7	0.8	0.6	0.8	0.7	0.8	0.7	0.9	1.0	0.6	0.7	0.805
9	1.0	0.3	0.7	0.6	0.6	0.3	0.5	0.3	0.7	0.4	0.5	0.2	0.3	0.2	0.2	0.4	0.7	0.5	0.2	0.3	0.445
10	0.8	0.4	0.6	0.7	0.7	0.5	0.3	0.3	0.7	0.6	0.3	0.3	0.4	0.1	0.2	0.3	0.7	0.8	0.1	0.5	0.465
11	0.7	0.8	0.4	0.1	0.8	0.7	0.7	0.4	0.7	0.8	0.7	0.4	0.6	0.8	0.8	0.9	0.4	0.7	0.1	0.8	0.615
12	0.6	0.3	0.7	0.2	0.1	0.2	0.5	0.5	0.7	0.5	0.5	0.4	0.4	0.3	0.4	0.2	0.4	0.5	0.3	0.2	0.395
13	0.7	0.4	0.6	0.4	0.9	0.3	0.3	0.4	0.9	0.7	0.4	0.3	0.4	0.4	0.5	0.4	0.9	0.9	0.5	0.4	0.535
14	0.5	0.5	0.6	0.3	0.6	0.4	0.4	0.2	0.5	0.4	0.3	0.2	0.3	0.2	0.3	0.3	0.5	0.4	0.2	0.4	0.375
15	0.7	0.6	0.5	0.4	0.8	0.9	0.4	0.3	0.5	0.6	0.7	0.8	0.7	0.8	0.6	0.4	0.5	0.6	0.6	0.3	0.585

- During the rent and sale management process, the risk includes risk of marketing opportunity D_1, risk of sales planning D_2, risk of operating contract D_3 and risk of others such as natural disaster and contingency.

B. *Factors affect the calculation of indicators*

Table 1 is the data that the 15 real estate investment experts have assessed .In the table A_1-D_4 are the risk indicators, as the last item on the expert assessment of the risk assessment value. A_1-D_4 and other effects of the 20 real estate investment risks, the risk indicators of both qualitative factors and quantitative factors, and even the quantitative factors, there is also a great difference in their dimension. As a result, the risk indicators for the A_1-D_4, use of expert scoring. Points: 1.0-0.1 is divided into ten grades 1.0, 0.7, 0.5,0. 3, 0.1 counterpart at risk: higher, high, average and low, lower five grades, and scores of other changes in the appropriate level. When experts examine the full score, the analysis of the rate of investment in real estate industry to submit projects related to the risk, given the risk indicators of the value of the score, the evaluation was to assess the objectives of the indicators on the performance and may be caused by the size of the risks.

C. *Solving Model*

Select the parameters of SVM learning and have a greater ability to predict the impact on the parameters of SVM is a more important issue. SVM includes close to the main parameters of the error tolerance, punishment and C

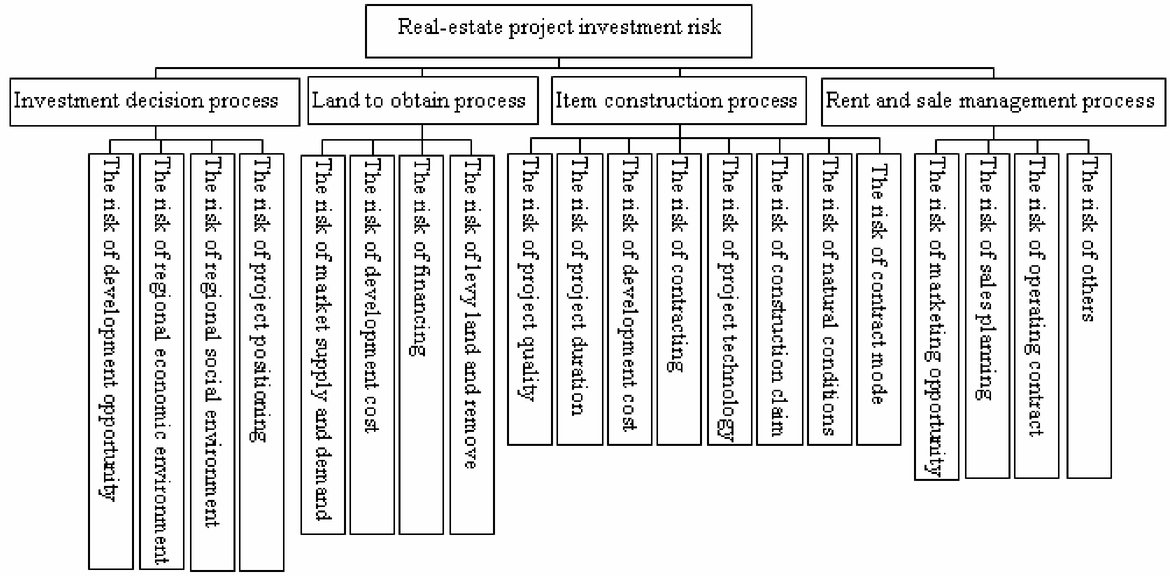

Figure3. Impact factors of risk of real estate investment in all stages

parameters of the nuclear option and its function parameters. In this paper, the nuclear option Gaussian radial basis function kernel (RBF) therefore needs to optimize the parameters. Optimization of these parameters can be tested each other (cross-validation) to determine the way. This article pre-selected sample set of 12 samples as a training sample, 3 samples for testing samples for testing Forecast accuracy of the model approach. First of all, fixed parameters to adjust other parameters, and select one of the least of the parameters as to optimize the value. Finally, by trial and error to determine the final optimal model parameters are as follows: $\sigma = 0.81$, $\varepsilon = 0.24$, $C = 110$, the test results of root-mean-square error $MSE = 0.0157$, the minimum of relative error = 0.0216 and the maximum of relative error = 0.0351. The results show that: support vector machine regression theory to predict the risk of real estate investment, resulting in an effective and workable ways.

V. CONCLUSIONS

Support Vector Machine is a new general machine learning methods which is based on a statistical learning theory framework and can solve in the past a lot of learning of the existence of small samples, had to learn, and so on the actual local minimum problems. This article first supports vector machine technology into real estate investment risks forecasting and tests the effects. The result shows that this method has higher prediction accuracy, and provides a new approach to real estate investment risk control and management.

ACKNOWLEDGEMENT

It is a Doctor project supported by Hebei University of Engineering in Hebei Province

REFERENCES

[1] Vladimir N. Vapnik, “The Nature of Statistical Learning Theory,”. Springer-Verlag, New York,1995,pp.354-432.

[2] Zhang Jin-mu,Wu Bo,Chen Yu and Lin Jian, “Study on the Method of Projection Risk Evaluation Basedon Support Vector Machine,” Optimization of Capital Construction, Vol. 27,No. 6 ,pp.77-94,2006.

[3] Chou Zhicheng,“Modeling by Using Support Vector Machine,”Techniques of Automation &Applications,2007, pp.9-11.

[4] An Wensen,SunYanguang, “A New Method for Constructing Kernel Function of Support Vector Regression,” Information and Control, Vo1.35, No.3, pp.378-381, 2006.

[5] Hu Gen-sheng, Deng Fei-qi,“Support vector regression with piecewise loss function and its application in investment decision Control,”Theory& Applications, Vo1.23, No.2, pp.315-318,2006.

[6] Zhou Xiaojing, Zhang Yaoning, “The whole process of Real Estate Investment Risk Identification,” Economic Research Guide, No.8, Serial No.15.pp.186-187,2007.

[7] Shen Liang , “The Research and Application of Risk Management on the Real Estate Development Projects,”,pp.11-23. 2007.

[8] Zhang Lixia Sift Guoqing, “Project Risk Prediction Based on the Support Vector Machine,”Computer Engineering and Application, Vol.21.pp.224-225,2005.

[9] Zhou Xiaojing ZhangYaoning, “ECONOMIC RESEARCH GUIDE,” Serial No.15, No.8, pp.186-187,2007.

[10] Tang Xiuli, Yang Zijun, Zhao Shiqiang, “The research on identification of real estate investment,”CONSTRUCTION ECONOMY, Vol.7.pp.207-207, 2004.