APPLICATION OF SUPPORT VECTOR MACHINES IN FORECASTING NON-RESIDENTIAL CONSTRUCTION QUANTITYDEMAND IN THAILAND

Wiwat Kittinaraporn[1], Napat Harnpornchai[2], Sutja Boonyachut[3]

1) Ph.D. Candidate, Department of Civil Engineering, Mahanakorn University of Technology, Bangkok, Thailand. Email: kittinaraporn@gmail.com
2) Assistant Professor, Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand. Email: napateconcmu@gmail.com
3) Assistant Professor, Department of Civil Engineering, Mahanakorn University of Technology, Bangkok, Thailand. Email: sutjab@gmail.com

Abstract: This paper deals with the application of a novel neural network technique, so called Support Vector Machine (SVM). The objective of this study is to explore the variable and parameter of forecasting factors in the construction industry to build up forecasting model for construction quantity demand in Thailand. The scope of the research is to study the non-residential demand of construction quantity in Thailand. There are 44 sets of yearly data available, ranging from 1965 to 2009. The correlation between economic indicators and construction demand with the lag of one year was developed by (ApichatBuakla, 2005). The selected variables are used to develop SVM models to forecast the non-residential demand of construction quantity in Thailand. The parameters are selected by using ten-fold cross-validation method. The results are indicated in term of Mean Absolute Percentage Error (MAPE). The MAPE value for the non-residential construction quantity predicted by Epsilon-SVR in corporation with Radial Basis Function (RBF) of kernel function type is 5.90. Analysis of the experimental results show that the support vector machine modeling technique can be applied to forecast construction quantity time series which is useful for decision planning and management purpose.

Keywords: Forecasting, Non-Residential, Construction, Support Vector Machines

1. INTRODUCTION

Construction industry plays an important role for economic and social development of Thailand. It is an industrial strategy for recovering economy and society. It is a fundamental industry in terms of economy development. The construction industry includes the infrastructure, residence, and industry constructions. The construction industry creates a huge amount of revenue for the country in terms of creating different kinds of job as well as construction material industry.

Forecasting is vital in the construction industry because the organization administrators always use the results of the forecasting to support the planning of construction projects in many respects. These include the establishments of organization policy, objectives, strategy, and decision making. In addition, it is useful for optimal and efficient resource allocation. Forecasting can be used as tools in performance evaluation by assessing situations and in creating future requirements which stimulates the organization operations.

There have been several research works on forecasting. (ApichatBuakla, 2005) studied the artificial neural network (ANN) model for predicting the construction quantity in Thailand. It is found that the ANN model yields better predicting results than Multiple Linear Regression. (Tanratanawong& Scott, 2000) developed the ANN model to predict the construction quantity in UK. The model used the quarterly data from 1955-1998 to create the relationship between the annual economic factors and the construction quantities of different types in the following years. There were three ANN models, namely residential construction, non-residential construction, and repairing and maintenance work. It was found that the ANN models yielded better results than the regression model. The results of the study were compared with the prediction results that had been done by two local authorities which had used the Delphi technique and were published as the present references. The ANN-based and the Delphi-based results had similar accuracy for the residual and non-residential constructions. For the repairing and maintenance, the ANN model yielded better results.

It can be seen that previous research works utilized the ANN and linear regression models for prediction. This research applies the SVM to the forecasting of construction quantity. The application of the SVM has not been done in the area of construction management. The SVM model is for the prediction of the future construction quantity in Thailand. The results will be used for creating the database in decision making, planning, and investment of Thailand construction industry. It is thus beneficial for the construction industry. In addition, there will be comparison among the SVM model and other forecasting models for the purpose.

2. DATA AND METHODOLOGY

2.1 Data

(ApichatBuakla, 2005) showed the linear regression model for the non-residential construction using the SPSS program to analyze the regression. The selection of the independent variables is carried out using the stepwise method according to Eq. (1)

$$NRES_{n+1} = ij + a_1X_{n1} + a_2X_{n2} + a_3X_{n3} + \ldots + a_nX_{nk} \qquad (1)$$

Where

$NRES_{n+1}$ is the future annual construction quantity of the non-residential type.

$a_1, a_2, a_3, ..., a_n$ are the regression coefficients.

$X_{n1}, X_{n2}, X_{n3}, ..., X_{nk}$ are the factors that influence the construction quantity of the non-residential type.

The selection of the linear regression model is based on the lowest MAPE.

The data for the input are the data of GDP collected from 1965 to 2009 altogether 44 years.

The analysis of data uses the software DTREG version 10.7.18 (Demonstration Version) http://www.dtreg.com, Phillip H. Sherrod

2.2Theoryof Support VectorMachine

SVM is the technique used for the pattern recognition and classification of data. It is developed by (Vapnik, 1995). The principle of SVM is the construction of hyperplane on the plane of training data to divide data into different groups. In the construction of the hyperplane, the distance between the points that are closest to the hyperplane on both sides will be defined, namely $d_+ + d_-$. Margin is defined as $d_+ + d_-$. The appropriate hyperplane is the one that has the largest margin, as shown in Figure 1. The data that lie on the edge of the margin will be referred to the support vector.

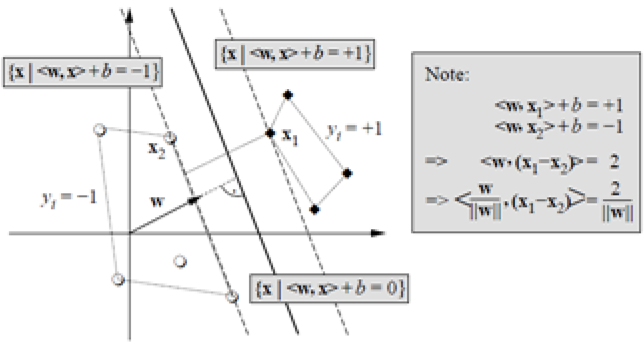

Figure 1.The classification of data by the hyperplane in SVM.

From Figure 1, there are 2 groups of data. The training data include the samples that can be expressed in terms of $\{x_k, y_k\}, k = 1,...,l$ and $x_k \in \Re^n, y_k \in \{-1,+1\}$ in which x_k is the input vector whereas y_k is the class label. The principle of SVM is to construct appropriate hyperplane on the training data plane. The hyperplane is defined by the parameters (w,b). w is the vector that is normal to the hyperplane and b is the constant which defines the position of the vectors that are related to the original position in the input space. The linear hyperplane equation is defined as $(w \cdot x) + b = 0$. To avoid the scale problem, w and b will be defined by the equation $|(w \cdot x) + b| = 0$ for the point that is closest to the hyperplane. Consequently, the hyperplane equation is given by Eq.(2)

$$y_i \left[(w \cdot x_i) + b \right] \geq 1 \ \forall i \quad (2)$$

Accordingly, it is just only the classification of the data using the hyperplane equation. In order to apply the algorithm to the nonlinear dataset, it is necessary to transform the training data to higher dimensional space, namely feature space. Such a transformation is carried out through the nonlinear function. The method is performed in terms a constrained optimization as defined by Eq. (3)

- Maximize

$$W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(x_i,x_j\right)$$

-Subject to (1) $\sum_{i=1}^{l}\alpha_i y_i=0$

(2) $0\le\alpha_i\le C\ \forall i$ (3)

$\alpha_i\ge 0$ is referred to as the Positive Lagrange Multipliers, $K(x_i,x_j)$ is the Kernel function and C is the constant for compensating the error induced during training and model complexity.

A number of the kernel functions can be used, e.g.

Polynomial of degree d

$$K(x_i,x_j)=(\lambda x_i^T x_i+r)^d,\gamma>0 \quad (4)$$

Radial Basis Function (RBF)

$$K(x_i,x_j)=\exp(\gamma\left\|x_i-x_j\right\|^2),\gamma>0 \quad (5)$$

Sigmoid Function

$$K(x_i,x_j)=\tanh(\gamma x_i^T x_j+r) \quad (6)$$

Linear

$$K(x_i,x_j)=x_i^T x_j \quad (7)$$

Consequently, the architecture of the SVM can be depicted in Figure 2. $\hat{y}$ is the estimated output and x is the input. X_i is the support vector, β_i is the weight, and b is bias.

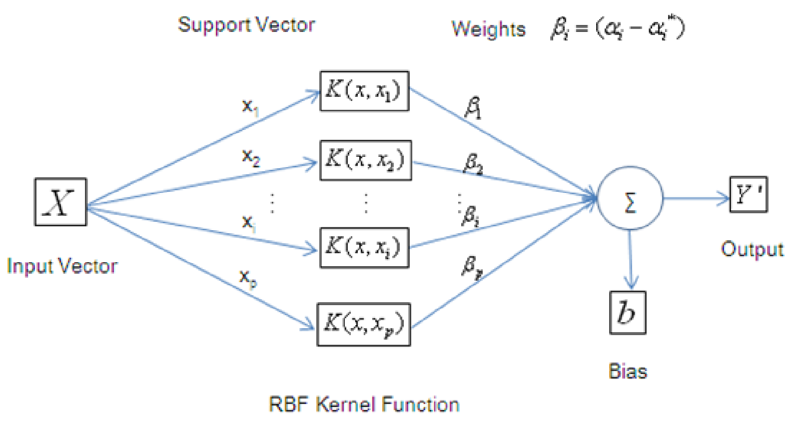

Figure 2 Architecture of SVM.

3. RESULTS AND DISCUSSION

3.1 Testing of Models

The testing of the forecasting model uses the GDP 9 types of data, i.e. T11, T24, T29, T48, T72, T79, T89, T93 and T123 as the input. The 10 float-cross-validation is used in training and testing.

Two types of SVM are used including the Epsilon-SVR and the Nu-SVR. Each type employs 3 kinds of kernel functions including the Radial Basis Function (RBF), Linear and Polynomial degree2.

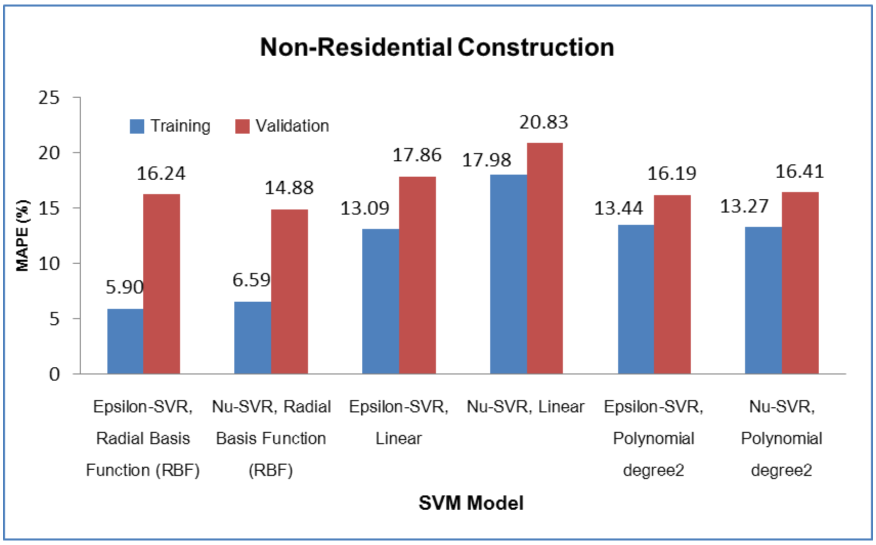

Figure 3 Results from SMV modeling.

Table 1: MAPE of the SVM models for forecasting the non- residential construction quantity.

SVM model	SVM kernel function	MAPE	
		Training	Validation
Epsilon-SVR	Radial Basis Function (RBF)	5.90	16.24
Nu-SVR	Radial Basis Function (RBF)	6.59	14.88
Epsilon-SVR	Linear	13.09	17.86
Nu-SVR	Linear	17.98	20.83
Epsilon-SVR	Polynomial degree2	13.44	16.19
Nu-SVR	Polynomial degree2	13.27	16.41

Table 1 and Figure 3 show that the SVM model for the forecasting of the construction quantity demand of the non-residential type using the Epsilon-SVR and Nu-SVR with the kernel functions of the Radial Basis Function (RBF) is the most efficient.

The next efficient one is the Epsilon-SVR with the kernel function of the Linear and the Nu-SVR with the kernel function of the Polynomial degree2, respectively.

The forecasting by the Epsilon-SVR with the kernel function of the Polynomial degree2 and the Nu-SVR with the kernel function of the Linear are less efficient.

4.2 Forecasting of Construction Quantity Demand

The model that yields least deviation is used for the forecasting construction quantity demand of the non-residential type, i.e. the Epsilon-SVR with the kernel function of the Radial Basis Function (RBF). The results of the forecasting are shown in Figure 4.

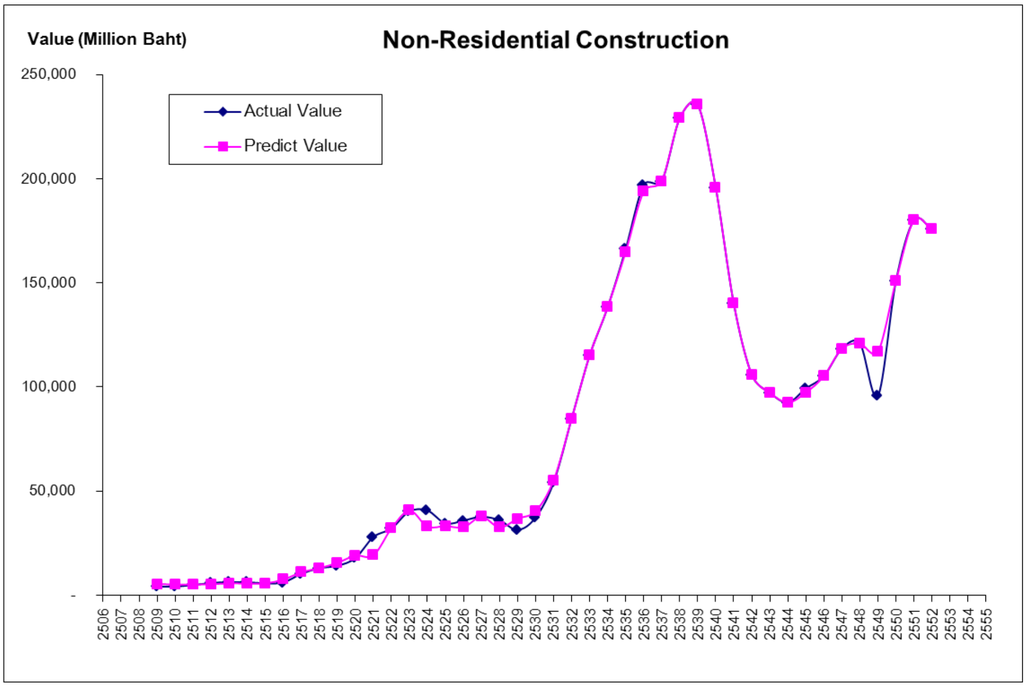

Figure 4 Forecasting of non-residential construction quantity using SVM.

From Figure 4, the forecasting is compared with the real construction quantity data from 1965-2009 altogether 44 years. The forecasting results show that the MAPE is equal to 5.90% which is significantly low.

4. CONCLUSIONS

This paper uses the SVM to predict the non-residential construction quantity demand. The Epsilon-SVR with the kernel function of the Radial Basis Function (RBF) is the best SVM and yields MAPE equal to 5.90%. The information from this research can be used as the guidelines for planning and decision making in construction organizations. The SVM shows its potential in forecasting other construction-related quantities too, e.g. material costs, construction times. In addition, it is expected to be applicable for the quality and safety prediction too.

REFERENCES

Al-Bahar, J.F. (1988). Risk Management in Construction Project: A Systematic Analysis Approach for Construction, Dissertation in partial satisfaction of the requirement for the degree of Doctor of Philosophy: University of California at Berkeley, Berkeley, California.

ApichatBuakla (2005). "Neural network model to forecast Construction output in Thailand", Thesis in partial fulfillment of the requirement for the degree of Master of Civil Engineering: Naresuan University, Phitsanulok, Thailand.

Goh, B.H. (1998). "Forecasting residential construction demand in Singapore: comparative study of the accuracy of time series, regression and artificial neural network techniques." Engineering Construction and Architectural Management, 5(3), 261-275.

Phillip H. Sherrod, "DTREG Predictive Modeling Software." www.dtreg.com. Copyright 2003-2007.

Tanratnawong, S. and Scott, S. (2000). "A neural network model to forecast national construction output.Journal of Financial Management in Construction and Property. (5), 65-77.

Vapnik, V. (1995).The Nature of Statistical Learning Theory, Springer-Verlag, New York.