A New Frontier in Information System Research – The Support Vector Regression Approach

Ching-Chang Lee
Department of Information Management, National Kaohsiung First University of Science and Technology

Shing-Hwang Doong
Department of Information Management, Shu-Te University

ABSTRACT

In this study, we propose a novel application of the Support Vector Regression (SVR) method to model a task variable in the Task-Technology Fit (TTF) theory. The support vector approach learns a parsimonious regression model from the given data to avoid the data over-fitting problem. Founded on the theories of statistical learning, mathematical programming and functional analysis, SVR is shown to outperform the traditional multiple linear regression method from the perspective of regression accuracy. Using a bootstrap procedure, we design a mechanism to extract significant factors from the support vector approach.

Keywords: Bootstrap, Support vector regression, Nonparametric regression, Task-technology fit, Mobile commerce

資訊系統研究方法的新領域—支援向量迴歸方法

李慶章
高雄第一科技大學資訊管理系

董信煌
樹德科技大學資訊管理系

摘要

本研究提出一個創新的支援向量迴歸方法來探討應用任務科技適配理論於資訊系統採用之問題。支援向量迴歸方法可以在給定的資料中產生一個簡潔的迴歸模式，以避免傳統機器學習法中的資料過度學習問題。根基於統計學習、數學規劃及氾函分析理論，支援向量迴歸方法較傳統的多元迴歸方法在迴歸正確性上有較好的成效。本研究中我們也利用拔靴法設計出一個由支援向量方法萃取顯著因子的步驟。

關鍵字：拔靴法、支援向量迴歸、無母數迴歸、任務科技適配、行動商務

1. Introduction

Vapnik and his coworkers proposed a novel method called Support Vector Machine (SVM) to classify data (Vapnik, 1998). The support vector machine is founded on theories from (i) statistical learning; (ii) mathematical programming; and (iii) functional analysis. Recently SVM has been successfully applied to many real world problems such as the hand written digit recognition, speech recognition, bioinformatics, and etc. Because of the high performance, SVM is receiving the attention of many researchers in the field of science and engineering.

In this study, we propose a novel application of SVM (Vapnik, 1995; Scholkopf et al., 2002) in regression analysis to study information system adoption problems. Support Vector Regression (SVR) is shown to outperform the multiple regression method from the perspective of regression accuracy. Using a bootstrap procedure, we design a mechanism to derive significant factors that explain the adoption of mobile commerce in the insurance industry. Our findings from this support vector approach are different from those given by the traditional Multiple Linear Regression (MLR) method. However, it seems that the factors found from this support vector approach can explain the adoption decision better than the factors found from the traditional approach.

This paper is organized as follows. In section 2, we introduce concepts of SVM and SVR, along with a bootstrap procedure to derive significant factors from a SVR model. Two *F*-like statistics will also be introduced to judge the significance level of a predictor variable in the regression model. SVR and MLR will both be applied to study a case of information system adoption using the Task-Technology Fit (TTF) theory. Comparison regarding regression accuracy and the resultant significant factors between both approaches is made in section 3. Finally, we conclude in section 4 with a few remarks regarding the implication of this research.

2. Materials and Methods

2.1 Support Vector Regression

The accuracy of a learnt binary classification rule is usually measured by the expectation error of the rule on test data. In statistical learning theory this error, called the expected risk or actual risk, is shown in equation (1),

$$R[f] = \int_{X\times Y} | f(x) - y | \, dF(x,y) \qquad (1)$$

where x is a multi-dimensional predictor variable, y the corresponding target class, f(x) the predicted class from the learnt classifier f, and F(x,y) a joint distribution of x and y for drawing test data. In addition, let us assume that training data and test data are independently and identically drawn from the input-output space according to the distribution F(x,y). Many learning algorithms including Artificial Neural Networks (ANN) from machine learning and MLR from statistics implement the Empirical Risk Minimization (ERM) principle to learn the model. The empirical risk Remp[f] for a classifier f in equation (2) is given by the fitting error of f on the training data (xi, yi), i =1, …,l.

$$R_{emp}[f] = \frac{1}{l}\sum_{i=1}^{l}|f(x_i) - y_i| \qquad (2)$$

Using statistical learning theory, Vapnik (Vapnik 1995, 1998) proved a type of estimate on the actual risk as follows:

$$R[f] \le R_{emp}[f] + VCConfidence \qquad (3)$$

where VCConfidence is a property of the function family used to learn the model. Basically, the VCConfidence quantity goes up as more complex functions are allowed in the family. Algorithms using ERM do not consider complexity of the function family, so their actual risk is not minimized. As an ERM-based algorithm uses more complex function to model the training data and hence reduce the empirical risk, the VCConfidence may go up substantially so that the upper bound for the actual risk actually increases. On the other hand, a Structure Risk Minimization (SRM) based algorithm such as SVM tries to control both empirical risk and complexity of classifiers at the same time. Thus a trained SVM generally performs better on test data than ANN. The comparison between ERM and SRM is depicted in Figure 1, where the increasing curve is given by the VCConfidence term and the decreasing curse is given by the empirical risk. The sum of them bounds the actual risk.

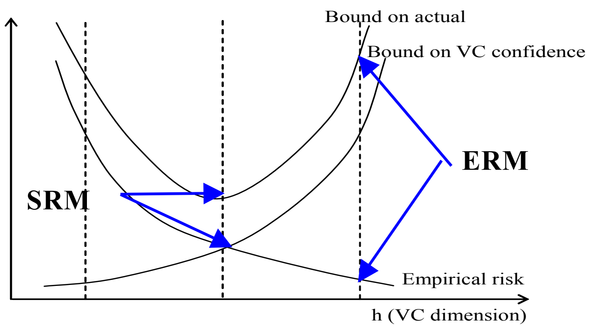

Figure 1: The bound on actual risk (Chin, 1999)

The original binary classification theory of SVM was later extended to solve a regression problem (Vapnik 1998). In this case, we are given a set of training data (xi, yi), i =1, …,l, where xi is a multi-dimensional input vector and yi the corresponding real-valued target value. We would like to find a regression function that predicts well for the test data drawn from the same distribution as the training data. Vapnik introduced a ε-insensitive loss function (ILF) to measure the empirical error to reduce the number of support vectors. In other words, when the difference $|f(x_i) - y_i|$ between a predicted value and the target value is smaller thanε, it will not contribute to the total empirical error according to this ILF. Finding a classifier with the smallest bound on the right hand side of equation (3) is translated into a mathematical programming problem. For theε-SVR problem, this is given by the following optimization problem:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C(\xi + \xi^*)$$

$$\text{subject to} \quad \begin{aligned} & w^T\phi(x_i) + b - y_i \le \varepsilon + \xi_i \\ & y_i - w^T\phi(x_i) - b \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, i = 1, \ldots, l \end{aligned} \tag{4}$$

Here C denotes a trade-off ratio between empirical errors (ξ_i, ξ_i^*) and the model complexity controlled by $\|w\|^2$. The feature map ϕ gives a nonlinear transformation from the input space to the feature space via a kernel function $K(x_i, x_j) = \phi^T(x_i)\phi(x_j)$. The optimization problem in (4) is further transformed into the following Wolf dual problem via the Lagrangian multipliers technique:

$$\max_{\alpha,\alpha^*} \frac{1}{2}(\vec{\alpha} - \vec{\alpha}^*)^T Q(\vec{\alpha} - \vec{\alpha}^*) + \varepsilon \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)$$

$$\text{subject to} \quad \begin{aligned} & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ & 0 \le \alpha_i, \alpha_i^* \le C, i = 1, \ldots l \end{aligned} \tag{5}$$

Here $Q_{ij} = K(x_i, x_j), \vec{\alpha} = (\alpha_1, \ldots, \alpha_l)^T, \vec{\alpha}^* = (\alpha_1, \ldots, \alpha_l)^T$. After solving the Wolf dual problem for the optimizing multipliers, we get a regression function in the following formula.

$$f(x) = \sum_{i=1}^{l}(-\alpha_i + \alpha_i^*)K(x_i, x) + b \tag{6}$$

Support vectors of a SVR model are defined to be those input vectors xi where $\alpha_i \ne 0$ or $\alpha_i^* \ne 0$. Because of the complimentary condition $\alpha_i\alpha_i^* = 0$, these support vectors actually span the regression formula in equation (6), and all other input vectors contribute nothing to the regression function. Scholkopf et al. (1995) showed that the support vectors for a classification problem represent a small fraction of the original data set and support vectors from different kernels overlap with a high percentage. In this sense, a support vectors set is a stable characteristic of the data (Scholkopf et al., 1995). For the SVR case, using the Kuhn-Tucker theorem (Vapnik, 1995; Scholkopf et al., 2002) one can show that the predicted value at a support vector deviates from the given target with a distance greater than or equal toε. In addition to the linear loss function in equation (4), a quadratic loss function such as $\xi_i^2 + (\xi_i^*)^2$ has also been considered in SVR literature.

A kernel function defines a nonlinear mapping from the input space to the feature space. With this nonlinear mapping facility, SVR has powerful regression capacity to model many real world applications. Because of the kernel, SVR can easily carry out the inner product operation of two feature vectors without explicitly using the feature map. Kernels commonly used in the SVR literature include the following functions.

1. Polynomial kernel:

$$K(x_i, x_j) = (\gamma x_i^T x_j + \theta)^d \tag{7}$$

2. Radial basis function kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \qquad (8)$$

3. Sigmoid kernel:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + \theta) \qquad (9)$$

Most researchers use the Radial Basis Function (RBF) kernel because its function can substitute other kernel functions (Keerthi and Lin, 2003; Lin and Lin, 2003). In this research, we adopt polynomial function and RBF as kernel for the SVR approach. Hyper-parameters γ, θ, d for the kernels, and C that decides the trade-off between regression accuracy and model complexity, will specify the setting for a SVR approach completely.

2.2 Extracting significant factors from SVR

Traditional MLR from statistics will not only give the regression coefficient but also the significance level of each predictor variable in the form of p-value. In this approach, it is assumed that the true output is corrupted by a random noise ε:

$$f(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon \qquad (10)$$

The significance level of a predictor variable xi is derived from a hypothesis testing with the following null hypothesis.

$$H_0 : \beta_i = 0 \qquad (11)$$

Under some normality assumption on the noise, it is shown that the above hypothesis testing can be examined using a t-test (Devore, 2004).

2.2.1 Bootstrap for SVR

SVR is commonly classified as a nonparametric approach in statistics (Green and Silverman, 1994; Sprent and Smeeton, 2002; Hastie et al., 2003) and very few literatures about testing the significance of predictor variables are available for this approach. Ait-Sahalia et al. (2001) used a Nadaraya-Watson kernel regression estimator to discuss the significant factors in a kernel regression problem. Green and Silverman (1994) discussed the model fit problem for a smoothing regression problem, i.e. a roughness penalty was charged to the prediction function in addition to the prediction accuracy. However, these two types of regression are different from the support vector regression. The kernel regression approach is more like a local modeling technique for regression analysis, while the smoothing regression approach is closely related to splines based regression.

As indicated above, significant factors extraction in MLR is resolved by testing the null hypothesis in equation (11). This problem can also be resolved by using the Sum of Squared Errors (SSE) approach as follows. Let SSEf denote the SSE of the full model with all predictor variables in regression modeling and SSEr the SSE of the reduced model with one predictor variable removed in regression modeling. If the left out variable is significant in the full model, we expect to see that the difference (SSEr – SSEf) is substantially large. For MLR, this is translated into the following f-statistic:

$$f = \frac{SSE_r - SSE_f}{SSE_f /(l - k - 1)} \quad (12)$$

Under the normality assumption on the noise, the random variable f has an F distribution of degree of freedom 1 in the numerator and degree of freedom (l – k - 1) in the denominator. Here l denotes the number of cases (training examples) and k the number of predictor variables.

For the SVR approach, we would like to use a similar idea to test the significance level of a predictor variable. However, since the distribution of the test statistic in equation (12) is not known for the support vector approach, we need to look for help from distribution free statistics. Bootstrap technique (Hastie et al. 2001, Devore 2004) is a common technique used in distribution free problems. Indeed, Friedman (2003) suggested a bootstrap procedure for testing the significance level of variables in a general regression setting. We will follow his approach for extracting significant factors in the SVR approach. Thus, a null hypothesis is posited as follows:

H0: the reduced model provides the same explanatory power as the full model

In order to test this hypothesis, a statistic t similar to (12) will be formulated and we assume that smaller values of t represent greater likelihood of H0. That is, SSE of the reduced model does not deviate too much from SSE of the full mode when H0 is likely to be true. We now produce P data sets by randomly permuting the output values of the original data set, i.e. we bootstrap P data sets such as $\{(x_i, y_{\pi(i)}), i = 1, \ldots, l\}$ from the original data set $\{(x_i, y_i), i = 1, \ldots, l\}$ with $\pi(i)$ being any permutation function on the set of integers {1,…,l}. For each of these data sets, a t statistic will be calculated and if the t value from the original data set is greater than or equal to the 1 - α quantile of {ti, i =1, …,P}, then we may reject H0 with significance level α. In other words, we will say that the left out variable is significant with level α in the SVR model. This is valid for any number of random permutations P, but the power increases with increasing P (Friedman 2003). The test statistic we will use for the support vector approach is listed below:

$$t = \frac{SSE_r - SSE_f}{SSE_f} \quad (13)$$

2.2.2 Effective number of parameters

Using the bootstrap procedure, we need not worry about the degree of freedom problem since we are comparing the order of test statistics from the same formula. Therefore we leave out the degree of freedom term in equation (13). On the other hand, one may ask if a simple formula like the F-statistic exists for the SVR approach. Then, this will involve the degree of freedom problem for the SSE random variables. A common approach to define this degree of freedom is to subtract the number of parameters from the number of total cases. Hastie et al. (2001) provides a formula to compute the effective number of parameters for general regression modeling.

Let us stack the given targets y1,…,yl into a vector $\vec{y}$, and similarly for the predicted targets into a vector $\vec{y}^P$. Now suppose we can write a fitting formula relating these targets as $\vec{y}^P = S * \vec{y}$, where S is an l by l matrix dependent on the input vectors xi but not on the targets yi. Then the effective number of parameters is defined as d(S) = trace(S), the sum of the diagonal elements of S. Now for the SVR using quadratic ε-insensitive loss function and ε is set to 0, we get the so-called ridge regression (Cristianini et al. 2000):

$$\underset{w,\xi_i}{Min}\ \lambda\|w\|^2 + \sum_{i=1}^{l}\xi_i^2$$

$$\text{subject to} \qquad y_i - w^T\phi(x_i) = \xi_i, i = 1,\ldots,l \qquad (14)$$

For this ridge regression, it can be shown that $\vec{y}^P = S * \vec{y}$ holds with

$$S = K(K + \lambda I)^{-1} \qquad (15)$$

Here K is the kernel matrix $K_{ij} = K(x_i, x_j)$ and the negative one exponent denotes the inverse matrix of (K + λI). The λ in equation (14) is equivalent to 1/C for the C in equation (4). Therefore, we can calculate the effective number of parameters for this particular SVR.

2.2.3 *F*-like test statistics

For a ridge regression, we introduce two F-like statistics to extract significant factors from the SVR approach. These two statistics use different formulae to interpret the degree of freedom for SSE.

- Test statistic 1: The full model has *k* parameters and the reduced model has *k* - 1 parameters, where *k* is the number of predictor variables. In other words, one predictor variable is counted as one parameter in SVR. Then, the first test statistic is given by

$$f_1 = \frac{SSE_r - SSE_f}{SSE_f/(l-k)} \qquad (16)$$

  And, the rejection region is $f_1 \geq F(\alpha;1,l-k)$ where $F(\alpha;1,l-k)$ denotes the *F* critical value with significance level α.

- Test statistic 2: The full model has v_1 effective number of parameters and the reduced model has v_2 effective number of parameters, where v_1 and v_2 are computed as the trace of matrix in equation (15). The full model will use all predictor variables to compute the kernel matrix while the reduced model will use the reduced set of variables to compute this matrix. The test statistic is given by

$$f_2 = \frac{(SSE_r - SSE_f)/(v_1 - v_2)}{SSE_f/(l-v_1)} \qquad (17)$$

  And the rejection region is $f_2 \geq F(\alpha;v_1-v_2,l-v_1)$. We will compute two types of kernel matrix by using the polynomial kernel and the RBF kernel in the experiment section. The kernel matrix K is an l x l positive definite matrix for both the full and reduced models with l equal to the number of cases. The matrix inversion in (15) is solved with a library from the IMSL Fortran library.

2.2.4 Significant Factor Extraction

In order to use a bootstrap procedure with the test statistic in equation (13) to extract significant factors from SVR, P = 100 data sets will be created from the original data set by randomly permuting the output variables. The test statistics for these P data sets are computed, and so is the one for the original data set. We will set the significant level α equal to 0.05 in this study. When the test statistic for the original data set is greater than or equal to 95% of the P test statistics created earlier, we reject the null hypothesis and conclude that the left out variable is a significant factor. Test statistics f1 and f2 in equations (16) and (17) are also calculated for the original data set and compared to the respective F distribution in order to extract significant factors. We then compare the results for significant factors extraction from the bootstrap and the F- statistics approaches.

3. Experiments and Results

3.1 Data Preparation

Data collected from study of Lee et al. (2006) is used to examine the performance of the SVR approach explained in the last section. We briefly introduce the data set in the following.

Lee et al. (2006) conducted an empirical study of mobile commerce in insurance industry based on the Task-Technology Fit (TTF) model (Goodhue and Thompson, 1995). According to TTF, the existence of a fit among task, technology and the user promotes the willingness of the user to use the technology and the user's work performance. In the case of mobile technology adoption in insurance industry, the particular TTF model is described in Figure 2.

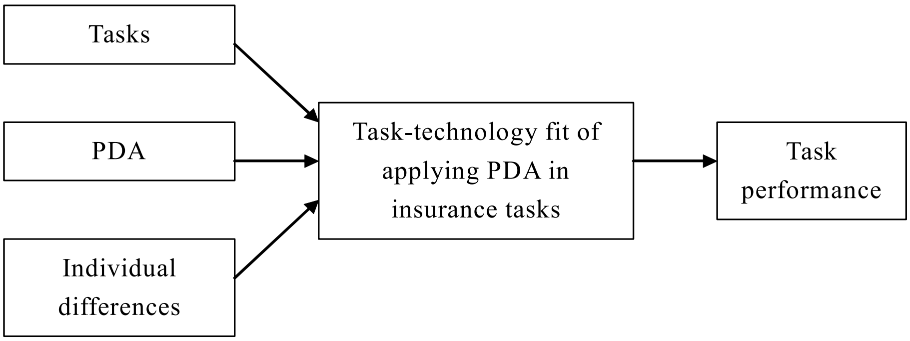

Figure 2: Task technology model for insurance industry (Lee et al., 2006)

In order to compare the SVR approach with the traditional MLR approach, we will use the second part of the TTF model as the research model (Figure 3). The eight indicators of TTF play the role of predictor variables and the three task indicators are the outcome variables. The three major tasks for insurance agents are (i) recruitment of new insurances; (ii) post-contract customer services; and (iii) supply of other information and services. The eight TTF indicators are explained below.

Goodhue and Thompson (1995) developed eight factors to measure the fit between tasks and information technology. These indicators include

1. Data Quality:
   a. Currency – The data meets the requirements of the task.
   b. Right data – to store the required data for task.
   c. Right level of detail – The stored data are correct with sufficient details.

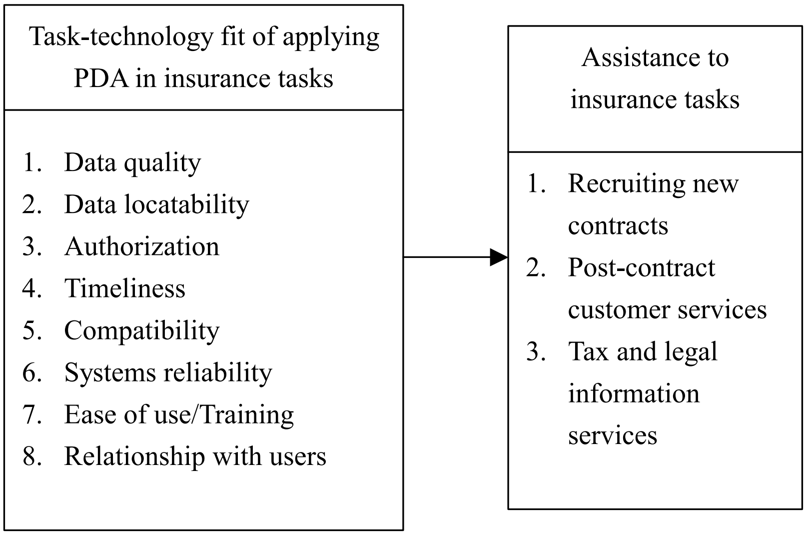

Figure 3: Research model for comparison (Lee et al., 2006)

2. Data Locatability:
   a. Locatability – The user can easily locate required data.
   b. Meaning – Each item of the data has clear definition and easy to use for users.
3. Authorization: The users are properly authorized to download data relevant to the task from corporate databases.
4. Production timeliness: The system can provide relevant information to the task in a timely manner.
5. Compatibility: The data are consistent with each other when they come from two or more different sources.
6. Systems reliability: The user can reliably depend on the system to complete the task without system problem and system breakdowns.
7. Ease of use/Training: It is easy to learn how to use the system and it is convenient to use the system to access data, including the ease of use of hardware and software, and easy to obtain relevant training.
8. Relationship with users:
   a. IS understanding of business – The system fits the users' daily requirements and corporate goals.
   b. IS interest and dedication – The data of the information system supports customers' requirements.
   c. Responsiveness – The information system can supply the needed information whenever the users need help in performing their work.
   d. Consulting – The company provides good technical support to information systems

users.

e. IS performance – The information system provides its users solution to their work requirements.

3.1.1 Samples & Measurement Methodology

In the study of Lee et al. (2006), 450 questionnaires are sent randomly and 274 are returned. Excluding incomplete and inconsistent questionnaires, there are 238 final useful samples. The questionnaire consists of questions related to the predictor and outcome variables in the regression problem. The measurement method is described as follows:

User's Task Performance

The effect of PDA technology on the insurance agents' performance is measured by the respondents' self assessment of how useful adopting PDA technology it is to assist them in performing the three major types of insurance tasks. The user's task performance is measured in a 5-point Likert scale. These task assessment indicators become the outcome variables in the three regression problems.

PDA Task-Technology Fit

A questionnaire is developed to assess the task-technology fit of PDA technology in the three major tasks of insurance agents based on the eight factors of task-technology fit model of Goodhue and Thompson (1995). The task-technology fit instrument uses a 5-point Likert scale. The descriptive statistics of these TTF indicators are shown in Table 1. These eight TTF indicators are the predictor variables in the three regression problems described above.

3.1.2 Reliability and Validity

Cronbach's α were used to measure the reliability of research instrument. In practical application, the value of Cronbach's α should exceed 0.5, preferably more than 0.7 (Nunally, 1978). A Cronbach's α higher than 0.7 is considered high reliability, while a value lower than 0.35 is deemed not reliable. The Cronbach's α of our instrument ranges from 0.6347 to 0.9412, indicating a medium high to high reliability.

To ensure content validity, the questionnaire design was based on well-established and validated instruments in the literature. In terms of construct validity, the factor loadings of the eight task-technology fit factors of the Goodhue and Thompson (1995) questionnaire were all above 0.5. In summary, both the content and construct validity of the research instrument have been achieved.

Table 1: Descriptive statistics for TTF indicators

Indicator	Min value	Max value	Average	Std. Dev.
TTF 1: Data quality	2	5	3.83	.59
TTF 2: Data locatability	2	5	3.86	.60
TTF 3: Authorization	2	5	4.11	.63
TTF 4: Timeliness	2	5	3.88	.67
TTF 5: Compatibility	2	5	3.69	.69
TTF 6: System reliability	2	5	3.64	.70
TTF 7: Easy of use/Training	2	5	3.65	.60
TTF 8: Relationship with users	2	5	3.64	.59

3.2 Experimental results

3.2.1 Regression Accuracy

The first comparison between the traditional method and our support vector approach is to compare the regression accuracy. MLR is commonly used to derive a model when the output is continuous. We use the MLR as the traditional approach and compare its performance with the new support vector approach.

We randomly partitioned the data set into a training set and a test set. A regression model was trained on the training set via the traditional approach or the support vector approach, and the model was validated on the test set. For the support vector approach, we adopted a quadratic polynomial kernel with $\gamma= 0.125$, $d = 2$, and $C = 0.01$. A linear insensitive loss function was used and the insensitive accuracy ε was set to 0.01. We used LIBSVM (Chang and Lin, 2001) to implement the ε-SVR. Table 2 lists results from both MLR and SVR regression techniques. The rate column indicates the percentage of the training data in the original data set. Thus, the first row is the result for a randomly selected training set that accounts for 70% (167 cases) of the original data set (238 cases). The MSE (mean squared error) column measures averaged squared errors on the test set, and the correlation column indicates the correlation coefficient between the predicted values f(xi) and the given values yi on the test cases. One can see that MSE from SVR is smaller than that from MLR, and predicted values from SVR approach are more closely correlated to given targets than MLR approach. This shows that SVR fulfills its mission of prediction by controlling the bound on the actual risk well.

Table 2: Comparison of generalization error for Task 1 regression

	Multiple regression		**Support vector regression**	
Rate	**MSE**	**Correlation**	**MSE**	**Correlation**
0.70	0.3096	0.5623	0.2696	0.6133
0.75	0.4021	0.5066	0.3729	0.5654
0.80	0.4059	0.4174	0.3363	0.4868
0.85	0.4464	0.4448	0.4260	0.5279
0.90	0.3987	0.4832	0.3695	0.5439

3.2.2 Factor Extraction

Table 3 lists the factor extraction information for regressing three major tasks with respect to eight TTF indicators by MLR. It can be seen that data quality is significant in the regression model for all three tasks, while compatibility and relationship with users factors are also significant with $p\text{-value} < 0.05$ for task 3.

Table 3: Multiple regression of tasks w.r.t. TTF

β **coefficients**	**Task 1**	**Task 2**	**Task 3**
TTF 1: Data quality	**.474***	**.582***	**.540***
TTF 2: Data locatability	.174	..064	.105
TTF 3: Authorization	.020	.106	-.153
TTF 4: Timeliness	.115	..18	.006
TTF 5: Compatibility	-.007	-.056	**.267***
TTF 6: System reliability	-.060	.004	.000
TTF 7: Easy of use/Training	..055	.022	-.013
TTF 8: Relationship with users	.112	.148	**.248***

*: significant with p-value < 0.05

Regarding the factor extraction problem with the SVR approach, we will use both bootstrap procedure and the two F-statistics proposed in section 2 to discover significant factors. For this study, we consider factors obtained with the bootstrap procedure are the correct ones for the SVR approach. In order to use the two F-statistics, we need to assume a quadratic loss function withεequal to 0 for SVR, i.e. a Ridge regression. Both the polynomial kernel and RBF kernel will be used to calculate SSE of the full model and the reduced model. Hyper-parameters γ, θ, d and C in equations (4), (7), (8) and (9) are commonly determined via a cross validation method. The open source software LIBSVM (Chang and Lin, 2001) provides a cross validation option when it is used to train a model. Hyper-parameters giving the highest cross validation accuracy are used to compute the SSE for the SVR approach. No matter which predictor variable is removed, we use the same hyper-parameters found for the full model to compute the SSE of the reduced model.

Tables 4 and 5 show results for the Task 1 regression problem using the polynomial and RBF kernels respectively. An 'X' mark in the bootstrap column indicates that the test statistic (13) for the original data set using the corresponding row factor as the left out variable for the reduced model is greater than or equal to 95% of the test statistics obtained on the P=100 randomly permuted data sets. This bootstrap procedure is considered the standard way to extract factors for SVR. In certain experiments where the focal test statistic is around 94%~96% of the test statistics from permuted data sets, we redo the experiments with P=200 to increase the power of the bootstrap approach. For the polynomial kernel, both test statistics in (16) and (17) extracted the same significant factor as the bootstrap method. However, for the RBF kernel test statistic f2 extracted the same factors as the bootstrap method, and the f1 statistic missed one significant factor (Data locatability).

Table 4: p-value for Task 1 model (polynomial kernel)

TTF indicators	**Bootstrap**	**p-value f_1**	**p-value f_2**
Data quality	X	**.000***	**.000***
Data locatability		.189	.052
Authorization		1.000	1.000
Timeliness		.215	.079
Compatibility		1.000	1.000
System reliability		1.000	1.000
Ease of use/Training		.634	.285
Relationship with users		1.000	1.000

*: significant with p-value < 0.05

Table 5: p-value for Task 1 model (RBF kernel)

TTF indicators	**Bootstrap**	**p-value f_1**	**p-value f_2**
Data quality	X	**.000***	**.000***
Data locatability	X	.062	**.022***
Authorization		.479	.291
Timeliness		.128	.061
Compatibility		.701	.501
System reliability		.667	.485
Ease of use/Training		.251	.117
Relationship with users		.588	.365

*: significant with p-value < 0.05

Tables 6 and 7 list the significant factors extracted for Task 2 with the polynomial and RBF kernels respectively. Again, they show that test statistic f2 extract more consistent factors with the bootstrap method than the f1 statistic. The f1 statistic claims too many factors as significant than the bootstrap method does.

Table 6: p-value for Task 2 model (polynomial kernel)

TTF indicators	**Bootstrap**	**p-value f_1**	**p-value f_2**
Data quality	X	**.000***	**.000***
Data locatability		**.039***	.060
Authorization	X	**.000***	**.002***
Timeliness		.122	.272
Compatibility		.215	.389
System reliability		1.000	1.000
Ease of use/Training		1.000	1.000
Relationship with users	X	**.001***	**.004***

*: significant with p-value < 0.05

Table 7: p-value for Task 2 model (RBF kernel)

TTF indicators	Bootstrap	p-value f_1	p-value f_2
Data quality	X	**.000***	**.000***
Data locatability	X	**.003***	**.028***
Authorization	X	**.000***	**.010***
Timeliness		**.011***	.112
Compatibility		**.022***	.193
System reliability		**.014***	.180
Ease of use/Training		.117	.487
Relationship with users	X	**.000***	**.002***

*: significant with p-value < 0.05

Tables 8 and 9 show results for the Task 3 regression problem. In this case, test statistics f1 and f2 are even in extracting significant factors for the task. The f2 statistic over-claims a significant factor (relationship with users) than the bootstrap method in the polynomial kernel case, and the f1 statistic fails to recognize this factor as significant in the RBF kernel case.

Table 8: p-value for Task 3 model (polynomial kernel)

TTF indicators	Bootstrap	p-value f_1	p-value f_2
Data quality	X	**.000***	**.000***
Data locatability		1.000	1.000
Authorization		1.000	1.000
Timeliness		1.000	1.000
Compatibility	X	**.003***	**.001***
System reliability		1.000	1.000
Ease of use/Training		1.000	1.000
Relationship with users		.094	**.029***

*: significant with p-value < 0.05

Table 9: p-value for Task 3 model (RBF kernel)

TTF indicators	Bootstrap	p-value f_1	p-value f_2
Data quality	X	**.000***	**.000***
Data locatability		.493	.399
Authorization		.698	.655
Timeliness		.537	.500
Compatibility	X	**.001***	**.001***
System reliability		.162	.164
Ease of use/Training		.196	.157
Relationship with users	X	.053	**.045***

*: significant with p-value < 0.05

We further summarize the result in Tables 10 and 11. Each 'X' mark in a column for a row means the corresponding row factor is significant (p-value < 0.05) with the specific test procedure for the specific task. Table 10 summarizes the result for the polynomial kernel, and table 11 for the RBF kernel. From these two tables, we can see that test statistic f2 is more consistent with the bootstrap procedure in extracting significant factors. This may be due to the fact that the effective number of parameters from equation (15) is closer to the real number of parameters for our SVR models.

Table 10: Significant factors from SVR (polynomial kernel)

TTF indicators	Task 1			Task 2			Task 3		
	BS	f_1	f_2	**BS**	f_1	f_2	**BS**	f_1	f_2
Data quality	X	X	X	X	X	X	X	X	X
Data locatability					X				
Authorization				X	X	X			
Timeliness									
Compatibility							X	X	X
System reliability									
Ease of use/Training									
Relationship with users				X	X	X			X

BS: bootstrap; significant with p-value < 0.05

3.3 Comparing the Factors

Both MLR and SVR yield significant factors affecting the adoption of mobile commerce in our study of insurance industry. However, they offer different choices of the factors in the three major tasks of insurance agents. We discuss the difference in the following.

Table 11: Significant factors from SVR (RBF kernel)

TTF indicators	Task 1			Task 2			Task 3		
	BS	f_1	f_2	**BS**	f_1	f_2	**BS**	f_1	f_2
Data quality	X	X	X	X	X	X	X	X	X
Data locatability	X		X	X	X	X			
Authorization				X	X	X			
Timeliness					X				
Compatibility					X		X	X	X
System reliability					X				
Ease of use/Training									
Relationship with users				X	X	X	X		X

BS: bootstrap; significant with p-value < 0.05

3.3.1 Task 1: Recruiting new contracts

The significant factor identified by MLR is data quality (Table 3), and the significant factors identified by the bootstrap procedure of SVR are data quality for the polynomial kernel and data quality and data locatability for the RBF kernel (Tables 10 and 11). MLR provides a linear model to fit the given data, while SVR with polynomial and RBF kernels provide another means with different explanatory power to fit the data. A kernel function determines how the input will be mapped (nonlinearly) into a feature space. Different kernels mean different feature maps and provide different regression accuracies. For most cases, the RFB kernel provides better regression accuracy than the polynomial kernel, which predicts better than MLR in turns. Therefore, the significant factors extracted from the RBF kernel seem to explain the regression model better than the polynomial SVR and the MLR methods.

3.3.2 Task 2: Post-contract customer services

Again, the significant factor identified by MLR is data quality (Table 3), and the polynomial SVR selects data quality, authorization and relationship with users indicators as the significant factors (Table 10). Using the bootstrap procedure, the RBF SVR adds the data locatability indicator to the list of significant factors (Table 11). This seems quite reasonable. The post-contract customer service is the most important reason that insurance agents need the mobility of a PDA to service their customers well. Customer may change the payment method, beneficiaries, and other terms of an insurance contract at anytime and anyplace, and the satisfaction and trust of the insured derived from such services will result in continuance of the existing contract and opportunity of new contracts in the future. Rules of thumb tell us that the other two major tasks of insurance agents can be helped by a rich personal network or well designed web system. On the other hand, many post-contract customer services are conducted face to face with the customers at anytime and anyplace, so insurance agents need professional and efficient technology to support the services. Task 2 seems to be the most demanding task among the three for the PDA mobile technology, and our SVR approach picks out more factors as significant than the MLR approach.

3.3.3 Task 3: Tax and legal information services

The MLR method captures data quality, compatibility and relationship with users as the significant factors for this task (Table 3); the bootstrap RBF SVR agrees with MLR in selecting the significant factors, while the polynomial SVR drops the relationship factor from the list of significant factors (Tables 10 and 11). In this case, we tend to believe that MLR and RBF SVR have selected the reasonable factors for further investigation.

4. Discussion and Conclusions

In this study, we propose a novel application of SVM in regression analysis to model task variables in the TTF theory. SVM is founded on theories from (i) statistical learning; (ii) mathematical programming; and (iii) functional analysis. It has been used successfully in science and engineering to solve many real world problems, but the application of SVM in information systems study seems to be scarce. The support vector approach learns a parsimonious regression model from the given data to avoid the data over-fitting problem. Since it uses the SRM approach, a model trained by SVM generally performs better than ERM based methods such as ANN or MLR. We have shown that the SVR approach provides a higher accuracy level than the MLR approach in a case study of mobile technology adoption in insurance industry.

We have also designed a bootstrap procedure to extract significant factors from the SVR approach. Our method compares SSE of the full model using all predictor variables and SSE of the reduced model using all but one predictor variable in regression modeling. As a common belief in statistics, if the left out variable is significant, then the difference between these SSEs should also be significant. We proposed equation (13) as the test statistic and used a bootstrap procedure to test the null hypothesis. Bootstrap procedure is appropriate here since the probability distribution of the test statistic (13) is not known. This support vector approach has yielded different significant factors from those given by the MLR method. Significant factors extracted from different kernel functions (e.g. polynomial and RBF) in SVR also differ. This is reasonable since different kernels provide different explanatory power of the fitted data. In other words, MLR, SVR with polynomial kernel and SVR with RBF kernel provide three different models to fit the data and they offer different explanatory power of the data. In our experience and also the experiences of many other researchers (Keerthi and Lin 2003, Lin and Lin 2003), SVR with a RBF kernel seems to explain the data better than SVR with a polynomial kernel, which again explains the data better than a MLR approach in turns. Therefore, significant factors extracted from the bootstrap procedure of a SVR with RBF kernel should be more meaningful than the ones extracted by other methods.

In addition to the bootstrap procedure, we also experimented two F-like test statistics for extracting significant factors. It is found that, for the case study of mobile technology adoption in insurance industry, test statistic f2 extracts more consistent factors with the bootstrap procedure when RBF kernel is used. This may be due to the fact that the effective number of parameters used in f2 is closer to the real number of parameters. Using f2, one needs not bootstrap many data sets to get a set of test statistics for comparison. On the other hands, we have to find the inverse of a high dimensional matrix in order to get the effective number of parameters.

In summary, we suggest future study of information systems to use SVR with RBF kernels, and a bootstrap procedure or test statistic f2 to extract significant factors. This is because SVR with RBF kernels can explain the fitted data better, and bootstrap procedure or test statistic f2 extracts consistent factors in our case study. Though this finding is limited to the data set of our case study of mobile technology, we believe it should generally hold true for a fair class of data sets. Another issue in MLR involves the interaction factors between predictor variables. The reason why we did not consider stepwise regression in SVR as commonly used in MLR to discover interactions between predictor variables is because the interaction factors have been taken care of in the kernel of SVR.

Our study in this paper has made the following contributions to the academia and the industry. First of all, we adopt a new regression method from science and engineering to study regression problems in information systems research. The SVR approach yields better regression accuracy than traditional MLR approach, and thus explains the data better. Secondly, we provide a novel approach to derive significant factors from SVR that affect an information technology adoption. This study of factors extraction from SVR seems to be very scarce in the literature. Our study has moved the understanding of SVR a further step forward.

There are many methods for feature selection in the field of data mining. For example, one may consider the problem as a combinatory optimization problem that selects the best combination of features from the pool of available features. The goodness of a feature set is judged by its explanatory power such as R2 - the coefficient of determination, and the search toward a better feature set is guided by heuristic procedures like genetic algorithm or simulated annealing. The limitation of this research is: (1) We only compared the traditional multiple linear regression approach with the various support vector regression based methods. The support vector regression approach is a late favorite in the field of machine learning, and yet using the technique to select features is scarce. We combine the support vector technique with a bootstrap method because the bootstrap method is a well-established procedure in statistics. (2) We tested our idea only on the data set collected from Lee et al. (2006). More verification may be performed on empirical data collected from future information system studies.

Acknowledgement

This research is partly supported by grants from National Science Council (NSC) of Taiwan under contract number: NSC-94-2213-E-366-013 & NSC 94-2416-H-327-006. We also appreciate the valuable remarks from the anonymous reviewers.

References

1. Ait-Sahalia, Y., Bickel, P. and Stoker, T., "Goodness-of-fit tests for kernel regression with and application to option implied volatilities", ***Journal of Econometrics*** (105), 2001, pp. 363-412.
2. Chang, C. C. and Lin, C. J., "LIBSVM: a library for support vector machines, 2001". Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
3. Chin, K. K., "Support Vector Machines applied to Speech Pattern Classification", Master's thesis, Engineering Department, ***Cambridge University***, 1999.
4. Cristianini, N. and Shawe-Taylor, J., "Support Vector Machines and Other Kernel Based Leaning Methods", ***Cambridge University Press***, 2000.
5. Devore, J., "Probability and Statistics for Engineering and the Science", ***Thomson Brooks/Cole***, Belmont, CA, 2004.
6. Friedman, J., "On Multivariate Goodness-of-Fit and Two-Sample Testing", ***PSYSTAT 2003***, SLAC, Stanford, CA, Sept. 8-11, 2003.
7. Goodhue, D. L. and Thompson, R. L., "Task-Technology Fit and Individual Performance", ***MIS Quarterly*** (19:2), 1995, pp. 213-236.
8. Goodhue, D. L., "Development and Measurement Validity of a Task-Technology Fit Instrument for User Evaluations of Information Systems", ***Decision science*** (29:1), 1998, pp. 105-137.
9. Goodhue, D. L., and Thompson, R. L., "Task-Technology Fit and Individual Performance," ***MIS Quarterly*** 19(2), 189-211, 1995.
10. Hastie, T., Tibshirani, R. and Friedman J., "The Elements of Statistical Learning", ***Springer-Verlag***, New York, 2001.
11. Keerthi, S. S. and Lin, C. J., "Asymptotic behaviors of support vector machines with Gaussian kernel", ***Neural Computation*** 15(7), 2003, pp. 1667-1689.
12. Lee, C.C, Cheng, K.S. and Cheng, H.H., "Empirical Study of Mobile Commerce in Insurance Industry: Task-Technology Fit and Individual Differences," ***Decision Support Systems***, 2006, in Press.
13. Lin, H. T. and Lin, C. J., "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods", ***Technical Report***, Department of Computer Science and Information Engineering, National Taiwan University, 2003.
14. Nunnally, J., "Psychometric Theory, 2nd ed.", ***McGraw Hill Book Co***, New York,

1978.
15. Scholkopf, B, Burges, C. and Vapnik, V., “Extracting Support Data for a Given Task”, ***Proceedings of First International Conference on Knowledge Discovery and Data Mining***, AAAI Press, Menlo Park, CA, 1995, pp. 252-257.
16. Scholkopf, B., and Smola, A. J., “Learning with kernels”, ***The MIT Press***, Cambridge. MA, 2002.
17. Sprent, P. and Smeeton, N., “Applied Nonparametric Statistical Methods”, third edition, ***Chapman & Hall/CRC***, Boca Raton, FL, 2002.
18. Vapnik V, "Statistical Learning Theory", ***Wiley***, New York, 1998.

Appendix: Questionnaire problems used in study of Lee et al. (2006). All problems assume a 5-point Likert scale response model for 1 = Strongly disagree, 2 = Disagree, 3 = Fair, 4 = Agree, and 5 = Strongly agree

- Self assessment

Task 1: The PDA mobile commerce system provides sufficient information for me to recruit new contracts
Task 2: The PDA mobile commerce system provides sufficient information for me to conduct post-contract customer service
Task 3: The PDA mobile commerce system provides sufficient information for me to offer other services such as tax and legal information to customers

- TTF indicators

Data quality

1. The data provided by the PDA mobile commerce system can satisfy my work needs
2. The PDA mobile commerce system has stored the necessary task related data
3. The PDA mobile commerce system has stored correct data with the right granularity for the task

Data locatability

1. It is easy to find the data that I need for the task in the PDA mobile commerce system
2. Data in the PDA mobile commerce system are clearly defined

Authorization

1. I am authorized to download work related data from the company database to my PDA

Compatibility

1. Data from different sources for the PDA mobile commerce system are compatible with each other

Reliability

1. The PDA mobile commerce system is reliable and functions correctly

Ease of use/Training

1. The PDA is easy to use
2. The PDA mobile commerce system is easy to use
3. I can always get the PDA related training whenever I need it
4. I can always get the PDA mobile commerce system related training whenever I need it

Relationship with users

1. The PDA mobile commerce system satisfies my daily work needs, and is consistent with the company goals
2. The PDA mobile commerce system can support my customers' requests
3. Whenever I have special requests or need help for the task, the PDA mobile commerce system always provides me with the needed information
4. The company provides good support for using the PDA technology
5. The company provides good support for using the PDA mobile commerce system
6. The PDA mobile commerce system provides me with the needed solution plans for the task