INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 13, Number 4, Pages 648–655

THE SVM-BASED PREDICTION OF PERIODICAL CIRCULATION AND PROCUREMENT COSTS IN A UNIVERSITY LIBRARY

SHILIAN CAI

Abstract. In order to effectively use books and reasonably distribute purchasing funds, the Support Vector Machines (SVM) method is used in this paper to establish a mathematics model for the related historical data of the library at Beijing University of Civil Engineering and Architecture. The book circulation and the allocation proportion of purchasing funds in the future are predicted based on the model. It is shown that the SVM method is feasible in predicting the book circulation and allocation proportion of purchasing funds with high non-linearity even if the size of a sample is small.

Key words. Support Vector Machines, book circulation, purchasing fund, algorithm, prediction.

1. Introduction

The book circulation is an important technical index to reflect the availability of library resources. The book circulation can be affected by the readership, the types of library books, and the requirements of readers for different literatures, etc. The prediction and analysis of the book circulation can provide scientific evidences for the further development of library resources, the mining of the library potential, the improvement of book quality, the improvement of services for teaching and research, and the implementation of quantitative management. The analysis of circulation variation and influence factors helps to plan book volumes for borrowing, design library facilities, administer literature, and adjust readers' behaviors.

There is a common problem in the budget allocation of all kinds of literatures in university libraries. It is important to solve the problem for the literature, information, and resources in the libraries. At present, most libraries schedule budget or expenditures through adjusting costs according to the plan in recent years and the available funds in the current year. This is a qualitative method based on past experience. There has been some quantitative research in this aspect; however, the research focused on the allocation ratio of the literature budgets or expenditures among different majors or departments when the available funds are known.

The total amount and the distribution ratio of collection expenditure are subjected to many complicated factors. These factors can be readers' demands, the environment and conditions of the library, and human factors, etc. All the various factors must be considered comprehensively in order to quantitatively analyze the total amount and the distribution ratio of collection expenditure and provide reference data for the resource allocation of the library.

Because there is some uncertainty in the borrowing time and the number of book circulation, the relationship between the book circulation and the related major factors is highly non-linear. In addition, the recodes and data in the library are incomplete. Even if they are complete, the time-series chain is short and there

Received by the editors October 15, 2013 and, in revised form, February 1, 2014.
The Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry No. 383(2005) and North China Electric Power University (2005).

are not enough samples for modeling. Therefore, it is difficult to effectively simulate book circulation through a traditional method. For example, traditional statistics methods can be used when there are enough sample data and the prior distribution of the sample is known (cf. [11]). However, it is not easy for these conditions or criteria to be met in practical applications; therefore, the results based on these methods are not ideal. The neural network method can solve non-linear questions (cf. [6]), but its applications are confined due to the uncertainty of its structure and the possible local minimization. Also, the learning algorithm in the neural network is able to make the experience risk (not the expectation risk) minimized. There is no substantive breakthrough in principles compared with the traditional least square method (cf. [10]), which makes it difficult to extend its applications.

The Support Vector Machines (cf. [3],[5],[9]), a new method, is based on the statistical learning theory which was proposed by Vapnik, et al. (cf. [11]). It achieves actual risk minimization through the structural risk minimization; therefore, a better learning effect can be obtained even if the sample size is small. This method introduces the concept of structure risk as well as uses the Kernel mapping idea. Compared with traditional methods, the advantages of the Support Vector Machines are not only overcoming the large sample requirement problem, but also solving the dimension disaster and the local minimization problem. In addition, it has a strong function in solving non-linear problems. Research in SVM has not been conducted for a long time. It has attracted researchers in China in recent years. Although SVM has a solid foundation in theory, there are many problems in applications and these problems should be solved. Its theory will also be further developed and extended with more and more applications.

The organization of this paper is as follows: the next section introduces some basic concepts and methods briefly; then, the linear SVM method and the non-linear SVM method are used to simulate and predict the book circulation and the collection expenditure in the university library; finally, numerical simulation results are analyzed.

2. Preliminaries

Regression problem: Let training set be as follows:

$$T = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \in (R^m \times R)^n,$$

where $x_i \in R^m,\ y_i \in R,\ i = 1, 2, ..., n$. Find $f(x),\ x \in R^m$, such that we can get y from $y = f(x)$ to any $x \in R^m$.

2.1. Linear SVM.

Let above real-valued function: $f(x) = \omega \dot{x} + b$, satisfying the following constraint condition:

$$\omega \cdot x_i + b - y_i \leq \varepsilon, \qquad i = 1, 2, \cdots, n, \tag{1}$$

$$y_i - \omega \cdot x_i - b \leq \varepsilon, \qquad i = 1, 2, \cdots, n. \tag{2}$$

Therefore, we introduce the the following object function (cf. [7])

$$\phi(\omega, \xi) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*), \tag{3}$$

where C is a positive constant and known as penalty factor; ξ, ξ^* are known as the upper and lower specification limits of relaxation variable respectively. We adopt the following ε insensitive function:

$$\beta_\varepsilon(y) = \begin{cases} 0, & \text{when} |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon, & \text{other,} \end{cases}$$

and then the following linear convex quadratic programming is obtained(cf. [7, 12])

$$\begin{cases} \min\limits_{\omega, b, \xi_i, \xi_i^*} \phi(\omega, \xi) = \frac{1}{2}||\omega||^2 + C\sum\limits_{i=1}^{n}(\xi_i + \xi_i^*) \\ s.t \quad [(\omega \cdot x_i) + b] - y_i \le \varepsilon + \xi_i, \quad i = 1, 2, \cdots, n \\ \qquad y_i - [(\omega \cdot x_i) + b] \le \varepsilon + \xi_i^*, \quad i = 1, 2, \cdots, n \\ \qquad \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \cdots, n. \end{cases} \tag{4}$$

In order to get the dual problem of the above problem, the following Lagrange function is introduced:

$$\begin{aligned} L(\omega, b, \xi^{(*)}, \alpha^{(*)}, \eta^{(*)}) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i \xi_i + \eta_i^* \xi_i^*) \\ - \sum_{i=1}^{n} \alpha_i(\varepsilon + \xi_i + y_i - (\omega \cdot x_i) - b) \\ - \sum_{i=1}^{n} \alpha_i^*(\varepsilon + \xi_i^* + y_i - (\omega \cdot x_i) - b), \end{aligned}$$

where $\alpha^{(*)} = (\alpha_1, \alpha_1^*, \cdots, \alpha_n, \alpha_n^*)^T$, $\eta^{(*)} = (\eta_1, \eta_1^*, \cdots, \eta_n, \eta_n^*)^T$ are Lagrange multiplier vectors.

As a result, the dual problem of convex quadratic programming (4) is: (cf. [7])

$$\begin{cases} \min\limits_{\alpha^{(*)} \in R^{2n}} -\frac{1}{2}\sum\limits_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) \\ \qquad + \sum\limits_{i=1}^{n}[\alpha_i(y_i - \varepsilon) - \alpha_i^*(y_i + \varepsilon)] \\ s.t \quad \sum\limits_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0, \quad 0 \le \alpha_i, \le \alpha_i^*, \quad i = 1, 2, \cdots, n. \end{cases} \tag{5}$$

Lagrange multiplier α_i α_i^* can be obtained by solving (5), then the coefficient of regression equation $f(x) = \omega \cdot x + b$ is as follows:

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i. \tag{6}$$

Because not all α_i, α_i^*, are zero, by using $Karush - Kuhu - Tucker$ condition we get:

If $0 < \alpha_i < C$, we have $\xi_i = 0$, then $[(\omega \cdot x_i) + b] - y_i = \varepsilon + \xi_i$, b is solved;

If $0 < \alpha_i^* < C$, we have $\xi_i^* = 0$, then $y_i - [(\omega \cdot x_i) + b] = \varepsilon + \xi_i^*$, b is solved.

2.2. Non-linear SVM.

Procedures similar to the linear case can be used. First, the data can be mapped to the high dimensional *eigenspace*; then a linear recurrence is curried out in the *eigenspace*. The dimension problem can be avoided by using the kernel function $K(x, y)$ (cf.[8]).

Consequently, the following non-linear planning problem is obtained(cf. [1, 2]):

$$(7)\qquad \begin{cases} \min\limits_{\alpha^{(*)}\in R^{2n}} -\dfrac{1}{2}\sum\limits_{i,j=1}^{n}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(x_i,x_j) \\ \qquad\qquad +\sum\limits_{i=1}^{n}[\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)] \\ s.t \qquad \sum\limits_{i=1}^{n}(\alpha_i^*-\alpha_i)=0, \quad 0\le\alpha_i, \le\alpha_i^*, \quad i=1,2,\cdots,n. \end{cases}$$

Function $f(x)$ can be directly expressed as follows:

$$(8)\qquad f(x)=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)K(x,x_i)+b.$$

According to $Kuhn-Tucker$ theorem (cf.[12]), we have

$$(9)\qquad \begin{cases} \alpha_i[\varepsilon+\xi_i-y_i+f(x)]=0, & i=1,2,\cdots,n, \\ \alpha_i^*[\varepsilon+\xi_i^*-y_i+f(x)]=0, & i=1,2,\cdots,n. \end{cases}$$

By simple calculation, we get

$$(10)\qquad \begin{cases} \varepsilon+\xi_i-y_i+f(x)=0, & \alpha_i\in(0,C), \\ \varepsilon+\xi_i-y_i+f(x)=0, & \alpha_i^*\in(0,C). \end{cases}$$

b can be obtained from (10).

2.3. Algorithm implementation.

The algorithm implementation of SVM are studied in this section, we discuss dual problem (5) as an example. The non-linear SVM is similar to the linear SVM except the kernel function is used in the inner product operation.

First, problem (5) is expressed as the following compact form

$$(11)\qquad \begin{cases} \min\limits_{\alpha} d(\alpha)=\dfrac{1}{2}\alpha^T H\alpha-e^T\alpha \\ s.t. \qquad q^T\alpha=0, \qquad\qquad 0\le\alpha\le Ce, \end{cases}$$

where $e=(1,1,\cdots,1)^T$, $q=(-1,1,\cdots,-1,1)^T$, $\alpha=\alpha^{(*)}$,

$$H=\begin{pmatrix} -1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}\begin{pmatrix} x_1\cdot x_1 & \cdots & x_1\cdot x_n \\ x_2\cdot x_1 & \cdots & x_2\cdot x_n \\ \cdots & \cdots & \cdots \\ x_n\cdot x_1 & \cdots & x_n\cdot x_n \end{pmatrix}\begin{pmatrix} -1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}^T.$$

Then, the sequential minimal optimization algorithms solving the convex quadratic programming (11) are as follows:

Algorithm 1: Sequential minimal optimization algorithm (cf. [3])

(i) Select the required precision ε and $\alpha^0=(\alpha_1^0,\alpha_2^0,\cdots,\alpha_{2n}^0)=0$, let $k=0$;

(ii) Select subset $\{i,j\}$ from set $\{1,2,\cdots,n\}$ as efficient set B by means of the feasible approximate solution α^k;

(iii) Solve the optimization problem related to efficient set B, the new feasible approximate solution α^{k+1} is got;

(iv) If the precision of α^{k+1} is satisfied according to ε, operation stops and we obtain approximate solution $\bar{\alpha} = \alpha^{k+1}$; otherwise, let $k = k + 1$ and return to (2).

Algorithm 2: Selection of the efficient set B (cf. [10]**)**

(i) Let α^k be the approximate solution of problem (11), then the gradient of the object function $d(\alpha) = \frac{1}{2}\alpha^T H\alpha - e^T\alpha$ at α^k is as follows:

$$\nabla d(\alpha^k) = H\alpha^k - e. \tag{12}$$

(ii) Compute

$$\begin{aligned} i &= \arg\max_t \{-y_t[\nabla d(\alpha^k)]_t | t \in I_{up}(\alpha^k)\}, \\ j &= \arg\max_t \{-y_t[\nabla d(\alpha^k)]_t | t \in I_{low}(\alpha^k)\}, \end{aligned}$$

where

$$\begin{aligned} I_{up} &\equiv \{t | \alpha_t < C,\ y_t = 1 \ or\ \alpha_t > 0,\ y_t = -1\}, \\ I_{low} &\equiv \{t | \alpha_t < C,\ y_t = -1 \ or\ \alpha_t > 0,\ y_t = 1\}. \end{aligned}$$

(iii) Let $B = \{i, j\}$.

Algorithm 3: Solve the optimization problems in the efficient set B (cf. [10]**)**

(i) The optimization subproblem according to the efficient set

Let B be the set of the subscripts of all training points that are included in the efficient set, then we can write α for the following form by changing the order of the components of the vector $\alpha = (\beta, \gamma)^T$, where β is the component in which the subscripts are included in set B, γ is the remainder of vector α except the component β . Therefore, Y and H can be expressed as

$$Y = \begin{pmatrix} y_{11} \\ Y_{22} \end{pmatrix}, \qquad Y = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}.$$

And then, the subproblem according to the efficient set B is expressed as

$$\begin{cases} \min\limits_{\alpha} w(\alpha) = \dfrac{1}{2}\beta^T H_{11}\beta - \beta^T(e - H_{12} \cdot \gamma), \\ \quad s.t. \qquad \beta^T y_{11} + \gamma \cdot y_{22} = 0, \qquad 0 \le \beta \le Ce. \end{cases} \tag{13}$$

Without loss of generality, let $\beta = (\beta_1, \beta_2)$, so the problem (13) is written as the following optimization problem with two variables:

$$\begin{cases} \min\limits_{\alpha} w(\beta_1, \beta_2) = \dfrac{1}{2}h_{11}\beta_1^2 + \dfrac{1}{2}h_{22}\beta_2^2 - \beta_1\beta_2 h_{12} \\ \qquad\qquad\qquad -(\beta_1 + \beta_2) + y_1 p_1 \beta_1 + y_2 p_2 \beta_2, \\ \quad s.t. \quad \beta_1 y_1 + \beta_2 y_2 = constant, \quad 0 \le \alpha_i \le C, \quad i = 1, 2, \cdots, n, \end{cases} \tag{14}$$

where

$$\begin{aligned} h_{i,j} &= y_i y_j (x_i, x_j), \qquad i, j = 1, 2, \\ p_1 = \sum_{i=3}^{n} y_i \alpha_i (x_i, x_1), &\qquad p_2 = \sum_{i=3}^{n} y_i \alpha_i (x_i, x_2). \end{aligned}$$

(ii) Let $(\beta_1^{old}, \beta_2^{old})^T$ be the feasible point of the problem (14)

Compute

$$\beta_2^{unc} = \beta_2^{old} + \frac{y_2 E}{\kappa},$$

TABLE 1. Readership, collection of books and circulation in 2000-2012.

Time	Undergraduate	Postgraduate	Faculty	Collection of books(volume)	Circulation(volume)
2000	950	50	150	28	98000
2001	1007	58	170	30	101211
2002	1336	64	183	35	141311
2003	1476	101	201	37	201855
2004	1380	103	263	39	217893
2005	1037	143	340	50	253959
2006	1041	215	425	60	299526
2007	1038	205	402	60	273343
2008	1512	260	429	64	238606
2009	1510	300	463	69	196627
2010	1510	362	446	80	188134
2011	1610	385	567	80	170312
2012	1720	430	600	120	170111

TABLE 2. Comparison of the prediction and realistic circulations (linear model).

Time	Real	Prediction	Absolute Error	Relative Error
2011	21503	22103	600	2.7%
2012	20709	23652	2943	14.21%

where

$$\begin{cases} E = p_1 - p_2 + \alpha_1^{old} y_1 (h_{11} - h_{12}) + \alpha_2^{old} y_1 (h_{12} - h_{22}) - y_1 + y_2, \\ \kappa = h_{11} + h_{22} - 2h_{12}. \end{cases}$$

(iii) Compute β_2^{new} by β_2^{unc}

$$\beta_2^{new} \begin{cases} V, & if \quad \beta_2^{unc} > V; \\ \beta_2^{unc}, & if \quad U \leq \beta_2^{unc} \leq V; \\ U, & if \quad \beta_2^{unc} < U, \end{cases}$$

where U and V are decided according to the following methods

$$\begin{cases} U = max(0, \beta_2^{old} - \beta_1^{old}), \\ V = max(C, C + \beta_2^{old} - \beta_1^{old}), \end{cases} \qquad when\ y_1 \neq y_2;$$

$$\begin{cases} U = max(0, \beta_2^{old} + \beta_1^{old} - C), \\ V = max(C, \beta_2^{old} + \beta_1^{old}), \end{cases} \qquad when\ y_1 = y_2.$$

The optimization solution of problem (14) is composed of β_2^{new} and β_1^{new} , i.e. $(\beta_1^{new}, \beta_2^{new})$.

3. Numerical modeling and analysis

The data in this paper are from the university library. For a quick convergence, the data are normalized by means of the following formula:

$$x_i^{'} = \frac{x_i - x_{min}}{x_{max} - x_{min}}. \tag{15}$$

After a normalized treatment of data, the numerical modeling is carried out. Finally, realistic simulation results are obtained by the inversion of formula (15). Let the parameter $\varepsilon = 0.05$ in this paper.

3.1. The modeling and analysis of book circulation.

We extract data (Table 1) according to the main factors which affect the book circulation.

We build the SVM model using the data of 2000-2010, and then the data of 2011-2012 are used to test the numerical modeling results. Algorithms 1-3 are used for writing program code. Numerical modeling is conducted for the linear and

Table 3. Comparison of the prediction and realistic circulations (non-linear model).

Time	Real	Prediction	Absolute Error	Relative Error
2011	21503	22021	518	2.4%
2012	20709	21374	665	3.2%

Table 4. The readership and the book purchasing fund 2000-2012 (ten thousand yuan).

Time	Undergraduates	Postgraduate	Faculty	Books	Database	Magazines
2000	950	50	150	14	75	15
2001	1007	58	170	17	78	17
2002	1336	64	183	35	80	34
2003	1476	101	201	59	95	32
2004	1380	103	263	83	100	42
2005	1037	143	340	170	221	43
2006	1041	215	425	113	212	44
2007	1038	205	402	42	112	44
2008	1512	260	429	66	119	44
2009	1510	300	463	122	122	50
2010	1510	362	446	128	171	53
2011	1610	385	567	217	211	55
2012	1720	430	600	243	216	81

Table 5. Predicted results (linear model).

Time		Book	Database	Magazines
2011	real	217	211	55
	prediction	192	197	52
	absolute error	-25	-14	-3
	relative error	-11.5%	-6.6%	-5.5%
2012	real	243	216	81
	prediction	212	201	73
	absolute error	-31	-15	-8
	relative error	-12.8%	-6.9%	9.9%

non-linear models. For the non-linear model, the kernel function is obtained as follows:

$$K(x, x^{'}) = [(x \cdot x^{'}) + 1]^2. \tag{16}$$

Simulation results are listed in Tables 1 and 2.

From Tables 2 and 3, we find that good results can be achieved by the SVM method (linear or non-linear) for small sample size case. Clearly, the predicted results of the non-linear model are better than those of the linear model.

3.2. The simulation and analysis of literature funds.

Similar to the book circulation case, literature funds are simulated. We get Table 4 from the statistical data of the university library according to the main factors related to literature funds.

The linear and non-linear SVM models are used to simulate the collection expenditure of books, database and magazines at the university, respectively.

Similar to Section 3.1, we build the SVM model using the data of 2000-2010, and the data of 2011-2012 are used to test the numerical modeling results. Predicted results are shown in Tables 5 and 6. The kernel function is still the one that is expressed in equation (16).

From Tables 5 and 6, we find that good results can be achieved by using the SVM method (linear or non-linear) for small sample size. In general, the predicted results of the non-linear model are better than those of the linear model except the prediction of magazines.

Table 6. Predicted results (non-linear model).

Time		Book	Database	Magazines
2011	real	217	211	55
	prediction	198	202	51
	absolute error	-19	-9	-4
	relative error	-8.7%	-4.3%	-7.3%
2012	real	243	216	81
	prediction	219	210	71
	absolute error	-24	-6	-10
	relative error	-10%	-2.8%	-12.4%

4. Conclusions

The appropriation expenditure of books and literatures is affected by different factors, and the relationship between them is non-linear. It is difficult to use conventional method to simulate the relationship. The *SVM* method can be used to simulate and predict the book circulation and the collection expenditure and good prediction results can be obtained even if sample sizes are small. The methods and results in this paper can be used to provide scientific evidence for the staffing and the budget allocation in a university library.

References

[1] Bennett. K, B1 Lie. J., A support vector machine approach to decision trees. Rensselaer Polytechnic Institute, Troy, NY: R. P.I. Math Repot., pp.97-100, 1997.

[2] Bradley J. P. S., O. L. Mangasarian, Massive data discrimination via Linear Support Vector Machines. http://citeseer. nj. nec. com/bradley98massive.html.

[3] Cortes C. Vapnik V. , Support vector machine. Machine Learning, 20, pp.273-297, 1995.

[4] Dricker H, Wu D, Vapnik V., Support vector machine for span categorization. Trans on Neural Neworks, 10(5), pp.1048-1054, 1999.

[5] Maller K-R, Smola A J, Ritsch G etc, Predicting time Series with support vector Machines. In: Proc. of ICANN' 97. Springer lecture notes in computer science, pp.999-1005, 1997.

[6] Mukherjee S, Osuna E, Girosi F, Nonlinear prediction of chaotic time series using support vector machines. In: Proc. of NNSP'97, 1997.

[7] Pedroso J. P., N. Murata, Support vector machines for linear programming: motivation and formulations. http://eiteseer. nj. Dec. com/pedros099support. html.

[8] Scholkopf B., R. Williamson, A Smola, etc, Support vector method for noveity detection. http://citeseer. nj. nec. com/400144. html.

[9] Smola A J, Scholkopf B., A tutorial on support vector regression. NeuroCOLT TR NC-TR-98-030, Royal Holloway College University of London, UK, 1998.

[10] Suykens J A K, Vandewalle J, De Moor B., Optimal control by least squares vector machines. Neural Networks, 14, pp.23-35, 2001.

[11] Vapnik V N., Statistical learning theory. New York: Wiley, 1998.

[12] Weston J., C. Watkins, Multi-class support vector machines. CSD-TR-98-04. Royal Holloway University of London, 1998.

Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Email: caishilian@bucea.edu.cn