A New Model for Libraries Efficiency Evaluation

Jianchao Kuang [1], Dong Wu[2], Lu Zhao[2], Zhong Wang[3], Xin Luo[3]
[1.]*College of Commerce, Chengdu University of Technology, Chengdu, China*
[2.]*College of Information Management, Chengdu University of Technology, Chengdu, China*
[3.]*College of Energy resources, Chengdu University of Technology, Chengdu, China*
Email: kjc@cdut.edu.cn, wd@cdut.edu.cn

Abstract: How to accurately evaluate library's efficiency is important to library management. However the accuracy of conventional models is low. This paper proposes a PLS-SVM model to solve this problem. SVM is very good in global optimization and generalization and is suitable for classifying different parents. While, PLS is integrated with SVM to treat the characters of samples, in order to overcome multi-linear correlation among variables and reduced the dimension of input variables at the same time. Applying the new model to evaluate the library efficiency of 14 college and universities, the accuracy achieved 100%. Therefore, the model can evaluate the efficiency of library accurately with results consistent with reality, and also can provide useful reference for the similar research in other regions.

Keywords: Library Efficiency; Evaluation Model; PLS-SVM

1 Introduction

Contemporarily, in an age of information, digitalization and intellectualization, the library as a knowledge reservoir, plays an imperative role in dissemination of knowledge and cultivating talents. China's lack of effectiveness of some of the concept of the library, formed a collection of books as the center, re-possession of light used in the passive mode of service. The concept of means and services as a result of the backwardness of the old management style, not the high quality of service, resulting in a low utilization rate library, the lower efficiency. It becomes a primary research subject that how to use the limited resources to meet the needs of readers, exert its function, increase the efficiency, that makes university library have sustained development.

The evaluation of library system is a value judgmental process, which is grounded on the objectives of the library as well as all relevant standards. Within the aim to draw an objective, fair and reasonable evaluation thoroughly by systematically analyzing the library work, researches of library benefits had been conducted centrally in recent past ten year in China. [1] For instance, Zheng, Gang (1978) established college Library Evaluation Indexes. [2] Sun Wei-qing (2004) expatiated on the library benefits theories, criteria and conversion efficiency. [3] Based on the fuzzy mathematics model theory, Chen Fang-fang (2008) invented a mathematics model for quantitative evaluation on user benefits of books. [4] Wang Ai-xia (2008), many research studies have been carried out on college index system Evaluation of University Libraries Management with analytical hierarchy process (AHP). [5] Researches and practices of library evaluations are highly valued in Chinese academic world, in contrast with the underestimated mathematic method. Thus, the evaluation of library merely remains on a qualitative analysis stage, whereas lack of quantitative analysis. As a result, qualitative and quantitative analysis should integrate simultaneously in order to examine the process and result of library activities. It will make a huge improvement for a much better, objective and scientific evaluation of library.

The present paper is financially supported the Key Project of Department of Education of Sichuan Province (Grant No. 08ZB069)

2 PLS-SVM Model of Library Efficiency

2.1 Partial Least Squares (PLS) Algorithm

Partial Least Squares is a new type of multivariate data analysis methods, is the integration of regression modeling, principal component analysis and canonical correlation analysis. [6] This method is applicable to more than the dependent variable of the multi-variable linear regression modeling, and can effectively address the serious variables related issues.

Assume dependent variable assemblage $Y=\{y^1, y^2, y^3, \ldots y^n\}$ and independent variable assemblage $X=(x^1, x^2, \ldots, x^n)$. Also $y \in R^n$, $x^i \in R^n$, i=1,2,…,n. y and x are standardized random variable, which is their average value is zero and variance equal one. Firstly, t^1 and u^1 are selected from independent variable assemblage and dependent variable assemblage respectively in PLS, that has to satisfy the following two requirements: [7]

(1) Maintain the original aberrance data as many as possible in order to achieve the biggest variance, i.e. $D(t^1) \rightarrow \max$, $D(u^1) \rightarrow \max$, $D(t^1)$、$D(u^1)$ is the variance

of t^1 and u^1.

(2) Maximum correlation between t^1 and u^1, i.e. $r(t^1,u^1)\rightarrow\max$, $r(t^1,u^1)$ is the correlation coefficient between t^1 and u^1.

Extract the first principal component and then do the regression analysis between X and Y to t^1. If the accuracy satisfied, then no further calculation is required. Otherwise, continue to do the 2nd round regression analysis on the data left from the previous calculation and so forth until the required accuracy will be satisfied. Lastly, extract amount of m components from X t^1, $t^2 \ldots t^m$, principal component of t^1, $t^2 \ldots t^m$ can be presented by X, then revert the regression equation between X and Y.

2.2 Principles of SVM

SVM (Support Vector Machine, SVM) is proposed by the 90' of 20th century. It is a statistical learning theory developed on the basis of the machine learning method to the principle of structural risk minimization instead of big samples experience risk minimization. The basic idea is the definition of the appropriate inner product kernel function; through the nonlinear transform input space will be mapped to a high-dimensional space, and then the characteristics of this new space to strike the optimal separating hyper plane. [8] SVM overcome the problems of general machine learning (artificial neural network). Because the samples are few so general machine learning appears less prone to "study" issues such as local optimization. SVM has self-organization, self-learning and associative memory function, and can deal with the situation of non linear sub.

If the training sample is linear inseparable, then the corresponding optimal question is:

$$\begin{cases} \min\limits_{w,\xi_i,\xi_i^*,b} & \frac{1}{2}(w\cdot w)+C\cdot\frac{1}{l}\sum\limits_{i=1}^{l}(\xi_i+\xi_i^*) \\ s.t. & (w\cdot\phi(x_i)+b)-y_i\le\varepsilon+\xi_i \\ & y_i-(w\cdot\phi(x_i)+b)\le\varepsilon+\xi_i^* \\ & \xi_i,\xi_i^*\ge 0 \end{cases} \quad (1)$$

ξ_i is a training sample of separating hyper plane deviation, ξ_i =0 linear separable. C is the defined penalty coefficient, which used for controlling the balance between samples and generalization capability. The bigger the C, the more inaccurate the penalty is.

Input Lagrange multiplier, the question will be transformed as the following dual form:

$$\begin{cases} \max\limits_{\alpha,\alpha^*} & \sum\limits_{i=1}^{l}\left[\alpha_i^*(y_i-\varepsilon)-\alpha_i(y_i+\varepsilon)\right]-\frac{1}{2}\sum\limits_{i=1}^{l}\sum\limits_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_i^*)K(x_i,x_j) \\ s.t. & \sum\limits_{i=1}^{l}(\alpha_i-\alpha_i^*)=0,\quad 0\le\alpha_i,\alpha_i^*\le C/l,\quad i=1,\cdots l \end{cases} \quad (2)$$

$K(x_i,x_j)=\phi(x_i)\cdot(x_j)$ is kernel function

The solution for question 2 is $(\bar{\alpha},\bar{\alpha}^*)$, then

$$f(x)=w\cdot\phi(x)+b=\sum_{SV}(\bar{\alpha}-\bar{\alpha}^*)K(x_i,x)+\bar{b} \quad (3)$$

The formula for $\bar{b}$ is:

$$\bar{b}=y_i-\sum_j(\bar{\alpha}-\bar{\alpha}^*)K(x_i,x_j)-\varepsilon,\quad \alpha_i\in(0,C/l) \quad (4)$$

$$\bar{b}=y_i-\sum_j(\bar{\alpha}-\bar{\alpha}^*)K(x_i,x_j)+\varepsilon,\quad \alpha_i^*\in(0,C/l) \quad (5)$$

Common choices of kernel function are defined as follows (Wang, 2008):

1) Linear kernels:
$$K(x,y)=x^Ty$$
2) Polynomial kernels:
$$K(x,y)=[s(x\cdot y)+c]^d$$
3) Gaussian RBF kernels:
$$K(x,y)=\exp\left\{-|x-y|^2/\delta^2\right\}$$
4) Two nuclear sigmoid neural network:
$$K(x,y)=\tanh\left[s(x\cdot y)+c\right]$$

2.3 Efficiency Evaluation of Libraries Based on PLS-SVM

SVM model of the traditional indicators is difficult to deal the issue of inter-related, thus affecting the accuracy of evaluation. PLS-SVM model use the least square method to the original targets, extract the final component for the input of partial least-squares SVM. The model not only overcomes a number of linear relationships between variables, remove the original correlation between indicators of the impact of the model, but also reduce the variables to improve the efficiency of the training. The Steps of efficiency evaluation of libraries model based on PLS-SVM are:

(1) Determine the classification of the effectiveness of the library. From an economic point of view to examine efficiency is the ratio of outputs and inputs. Library is an intellectual investment, but also the spirit of the production of the service sector; it is the community's input in the library of human, material and financial resources for investment. Its output is the library for the commu-

nity or provides readers with useful knowledge, intelligence and information, that is, the extent to meet the needs of the community. To facilitate the understanding, through expert's points, based on experience categories, the library effectiveness is divided into three levels: relatively high (Ⅰ), general (Ⅱ), is relatively low (Ⅲ).

(2) Select input variables and collect samples. Because of all the different dimensionless parameters, in order to avoid a nuclear function, including calculation of plot difficulties caused by the calculation, use pre-need for Pretreatment normalized so that all parameters are in [0, 1] between.

(3) Select the appropriate kernel function; put the data into the PLS-SVM model for training. In this paper, Gaussian kernel function (radial basis function or RBF) nuclear is used, C and δ is adjusted to find the best parameters, and then train the whole training set.

(4) Test the model. Under normal circumstances, if the model identifies rate reach 90%, then it is a reasonable model that can be used further for prediction and evaluation.

(5) Use the PLS-SVM evaluation model to evaluate the overall efficiency of other libraries.

3 Applications

University Library data in this article is used as an example to prove that the method is effective. A total of 14 samples are used to test the PLS-SVM model. [9] Data are shown in **Table 1**.

According to the PLS theory, standardize 14 samples in table 1 and extract 6 main elements in table 2. Choose RBF kernel function and then utilize software libsvm-2.88 to discipline the first 10 data (A-J) in **table 2**. In order to prevent the over-fitting problems, use intercross verification and grid searching to gain the optimal parameter. The result shown: when C=45.2548339959, δ =0.0078125, the accuracy of intercross verification can achieve 80%(**graph 1**). Then test the last four samples in **Table 2** into the model, the predicted accuracy achieve 100% (**graph 2**).

In addition, this article will also compared the partial least-squares support vector machine model results, the simple support vector machine results and the problem of discriminate analysis results, the results of the comparison in **table 3**.

Table 3.

School code	Level	Model		
		Discriminate Analysis	SVM	PLS-SVM
K	Ⅱ	Ⅲ	Ⅱ	Ⅱ
L	Ⅱ	Ⅱ	Ⅱ	Ⅱ
M	Ⅲ	Ⅲ	Ⅲ	Ⅲ
N	Ⅱ	Ⅱ	Ⅰ	Ⅱ
Prediction accuracy		75%	75%	100%

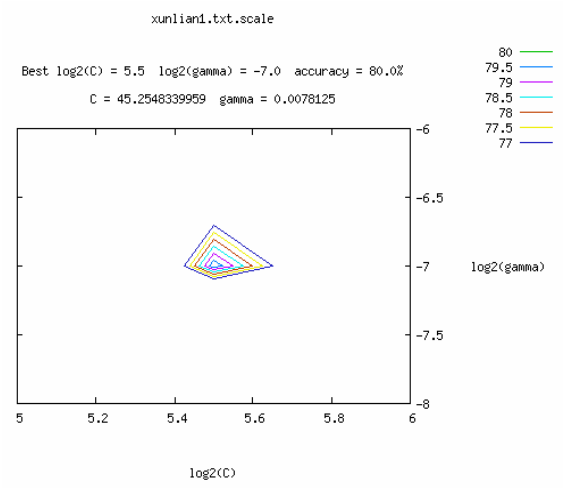

Figure 1. Parametric Selection Chart

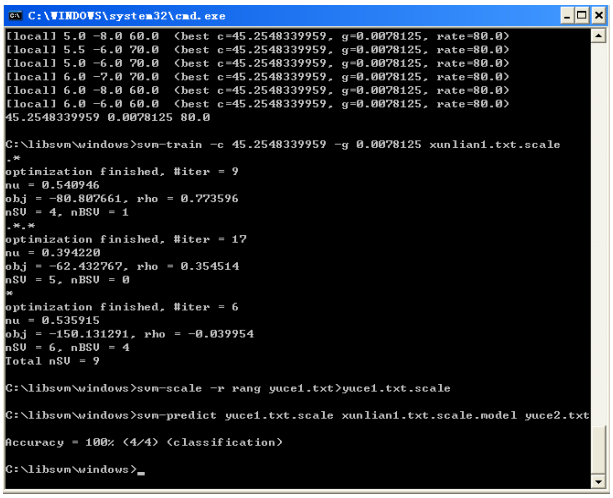

Figure 2. Results map

4 Conclusions

Based on partial least squares support vector machine method can predict the overall efficiency of library prediction and evaluation, the method for small samples of library materials, the other existing methods have the advantage of unparalleled. It has a good generalization ability, was able to effectively close the sample point, the prediction model was reasonable and objective response to the actual overall efficiency of the library. In this paper, experimental results show that in the evaluation of library effectiveness, partial least squares support vector machine support vector machines is better than simple methods and discriminate analysis methods. However, due to so many factors that affect the effectiveness of libraries, it is difficult to choose the best evaluation and assessment for the library, so that may affect the prediction.

978-1-935068-12-9 © 2010 SciRes.

Table 1. Classification of 14 samples

School code	Financial input X1i		Material input X2i				Human input X3i			The utilization rate output X4i				Users scale output X5i		Level
	X_{11}	X_{12}	X_{21}	X_{22}	X_{23}	X_{24}	X_{31}	X_3	X_{33}	X_{41}	X_{42}	X_{43}	X_{44}	X_{51}	X_{52}	
A	689	615	0.98	145	8	5849	192	119	133	107	896715	7590	24287	11165	13909	I
B	440	410	0.76	107	32	2704	155	114	125	84	1058863	5901	21342	13399	20205	I
C	181	172	1.46	271	106	3962	73	29	40	97	376696	2283	4729	19590	2506	III
D	274	272	1.02	140	21	3588	105	62	83	100	504413	20	253	19323	4946	II
E	332	359	0.84	126	3	1608	78	36	33	92	444045	61	1637	12141	4596	III
F	638	325	1.11	62	4	2961	118	74	103	93	552338	310	2560	18968	11392	II
G	314	275	1.38	117	10	6070	103	65	67	98	836476	0	5337	29812	3692	II
H	245	214	1.44	85	2	4132	144	70	69	103	479832	0	0	25602	6304	II
I	241	261	1.16	137	12	3800	54	28	28	105	380000	88	0	16529	0	III
J	217	203	1.38	105	6	6642	167	98	112	98	886136	0	1076	45682	9205	II
K	216	209	1.32	162	6	4433	114	71	65	98	729431	0	5345	30475	9199	II
L	275	185	0.82	72	12	3419	226	89	99	90	495764	226	630	33206	7204	II
M	217	212	0.79	177	30	1652	30	16	12	98	353838	0	88	10469	171	III
N	133	192	0.85	127	12	2300	132	59	73	94	1468885	0	104	41880	4892	II

Source from: Sorted by author

PS: A-N is the library codes. $X_{11:}$ annual average library operation costs. $X_{12:}$ annual average literature purchase expenditure. $X_{21:}$ library area measurement. $X_{22:}$ literature possessions. $X_{23:}$ quantity of newly purchased books. $X_{24:}$ quantity of seats. $X_{31:}$ full time staff number. $X_{32:}$ number of staff with bachelor degree or higher. $X_{33:}$ number of staff with medium professional title and above. $X_{41:}$ weekly opening hours. $X_{42:}$ number of external borrowed books. $X_{43:}$ annual Interlibrary Loan. $X_{44:}$ amount of transited literature. $X_{51:}$ number of full time bachelor and college students. $X_{52:}$ number of master and doctor students.

Table 2. After extracting the partial least squares data component

School code	P_1	P_2	P_3	P_4	P_5	P_6	Level
A	2.50	-0.33	-1.59	-0.12	0.30	-0.25	I
B	2.19	-0.19	-0.49	0.69	0.01	0.07	I
C	0.04	-0.18	-1.67	0.79	-0.26	-0.03	III
D	0.63	0.29	-0.94	0.00	0.00	-0.19	II
E	0.48	-0.14	-0.35	-0.12	0.09	-0.06	III
F	1.23	0.30	-0.66	-0.23	-0.32	0.24	II
G	0.77	0.70	-1.41	0.03	0.12	0.23	II
H	0.65	0.74	-1.26	-0.18	-0.17	-0.03	II
I	0.11	0.12	-1.13	-0.28	0.16	-0.09	III
J	1.09	1.33	-1.45	0.27	-0.02	0.04	II
K	0.71	0.64	-1.23	0.23	0.07	0.12	II
L	1.08	0.84	-0.48	0.24	-0.27	-0.52	II
M	-0.10	-0.30	-0.55	0.03	0.19	-0.23	III
N	0.70	0.88	-0.48	0.61	0.67	-0.12	II

Acknowledgements

The authors would like to thank the anonymous reviewers and editors for their helpful comment.

References

[1] Zhang Jian. The Theory and Methods of Library Evaluation[M]. *Southwest Jiaotong University Press*, Chengdu:2007 (Ch)

[2] Zheng Gang. Establish Index System of the University Library Evaluation[J]. *Libraries for the New Millennium*, 2008 (Ch)

[3] Sun Wei-qing. The Brief Talk on Benefits of Libraries[J]. *Journal of the Sichuan Society for Library Science*, 2002(04) (Ch)

[4] Chen Fang-fang. A Mathematical Model for Evaluating Library Efficiency of the Use[J]. *Scientific and Technical Information (Academic Research)*, 2008(06) (Ch)

[5] Wang Ai-xia. The Theory and Evaluating Target T System of University Library Management[J]. *Modern Information*, 2008(12) (Ch)

[6] Wang Hui-wen. Partial Least-squares Regression Method and Its Application[M]. *National Defense Industry Press*, Beijing: 1999

(Ch)
[7] Bastien P,Vinzi V E,Tenenhaus M. PLS Generalized Linear Regression[J].*Computational Statistics & Data Analysis*, 2005, 48:17-46
[8] Liu De-jun. Ran Zai-qun. Application of Support Vector Machine in Logging Analysis of Qyia sag in Daqing Oilfield[J]. *Geophysical Prospecting for Petroleum*, 2007(03) (Ch)
[9] Jin Ting. A Study on Benefit Evaluation of Libraries Based on DEA[D]. *Chengdu University of Technology*, 2009 (Ch)
[10] Zhong Bo, Liu Zhao-ke. Support Vector Machine Multi-classified Method Based on Partial Least-squares[J]. *Computer Engineering and Applications Computer Engineering and Applications*, 2007 (17) (Ch)
[11] Hong L., Hou J., Li Z. H. MINITAB Statistical Analysis Textbook[M]. *Publishing House of Electronics Industry*, Beijing: 2007
[12] Hsu C.W., Chang C.C., Lin C.J. A Practical Guide to Support Vector Classification. Http://www.csie.ntu.edu.tw/~cjlin, 2008
[13] Jiang Z.Q., Fu H.G., Li L.J. Support Vector Machine for Mechanical Faults Classification[J]. *Journal of Zhejiang University Science*, 2005, 6A (5), 433-439
[14] Vapnik V.N. The Nature of Statistical Learning Theory [M]. *Springer-Verlag*, New York, 1995
[15] Vapnik V.N. An Overview of Statistical Learning Theory[M]. *IEEE Trans. on Neural Networks*, 1999, 10(5): 988-999
[16] Virtual Computational Chemistry Laboratory. Partial Least Squares Regression (PLSR). Http://146.107.217.178/lab/pls/m_description.html
[17] Wang H.W., Wu Z.B., Meng H. Partial Least-squares Regression—Linear and Nonlinear Methods[M]. *National Defense Industry Press*, Beijing: 2006
[18] Wang L., Mu Z.C., Guo H. Application of Support Vector Machine in the Prediction of Mechanical Property of Steel Materials[J]. *Journal of University of Science and Technology*, Beijing: 2006, 13(6): 512-51.

978-1-935068-12-9