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:

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

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

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