行政院國家科學委員會專題研究計畫成果報告

支撐向量法與統計學習理論
Support Vector Machines and Statistical Learning Theory

計畫編號：NSC 89-2213-E-002-106
執行期限：89 年 8 月 1 日至 90 年 7 月 31 日
主持人：林智仁　　台灣大學資訊工程系
e-mail: cjlin@csie.ntu.edu.tw
http://www.csie.ntu.edu.tw/~cjlin

一、中文摘要

支撐向量法(Support vector machines)是一個非常新而且有前景的模型辯識方法。 經過 Bell Labs 人員在 90 年代初期的研究，目前 Support vector machines (SVM) 已成為 Machine Learning 與 Pattern Recognition 中一重要主題

經過今年與去年國科會計畫的研究，我們已成為世界上 SVM 軟體的主要提供者。研究單位與工業界已有超過上千位使用者。在本報告中我們將敘述到目前為止我們的成果。

另外我們也應用 SVM 到各種實際問題。我們使用此軟體參加 IJCNN Challenge 2001(IJCNN 是世界上最大的 Neural networks 會議之一) 獲得第一名及英國郵局 2000 年贊助的光學字型辨識 (Optical Character Recognition) 比賽獲得第一名。另有國外多所學校已使用我們的軟體來教學。

關鍵詞：支撐向量法, 模型辨識, 統計學習, 最佳化問題

Abstract

The support vector machine (SVM) is a new and very promising technique for classification and pattern recognition. After the development at Bell Labs in early 90s, it has now been an important research field in machine learning and pattern recognition.

After the efforts in the past two years, currently our group is one of the major players on SVM software development. There are thousands of users worldwide. We also apply SVM to some practical data classification problems. For example, using SVM our group is the winner of IJCNN Challenge 2001, a competition sponsored by one of the largest Neural Networks conference in the world. In this report we will describe our existing results in detail.

Keywords: pattern recognition, support vector machines, statistical learning theory, large-scale optimization

二、Introduction

Given some data with known classes (categories), a data classification problem is to generate a model based on existing information. Then for new data with unknown classes, this model can predict which classes they are in. Some examples are object recognition, text categorization, and handwritten digit recognition, etc.

Existing data classification methods are considered topics of pattern recognition and machine learning. For example, decision tree, neural networks, and Bayesian networks. Recently a new and promising method called support vector machine (SVM) [1,9] was developed in former AT&T Bell Labs and has drawn wide attentions in machine learning community.

Given l training data in two classes, the support vector machine tries to find a separating hyperplane between these two classes of data. Though this idea for linear separable data has been proposed 30 years ago, only recently researchers realized how to generalize it for more complicated data. To be more precise, for data which are not linear separable, we can map them to a higher dimensional space. Then on this higher space, there may have a linear separating hyperplane to separate those data.

To find the separating hyperplane, the support vector machine requires the solution of a quadratic programming problem with l variables. However, the Hessian matrix in the objective function is fully dense so there are even difficulties to store the problem itself. Therefore, traditional optimization methods cannot be applied to solve this problem.

To solve large-scale classification problems, after 1997, some researchers have proposed methods to conquer this memory difficulty. A major method is called the decomposition method where the basic idea is to fix most variables and only work on few variables in each optimization iteration. That is, a sub-problem is solved so the memory is enough to store the data of this sub-problem. This is like that if you want to minimize a function with ten variables, you use an iterative process where in each iteration, eight variables are fixed and a sub-problem with two variables is solved. Thus we call these two variables the working set while the other eight variables are in the non-working sets. In each iteration, there will have a new working set. Then each variable is updated in some iterations so finally the problem is minimized.

In an earlier project (NSC 89-2213-E-002-013) we have preliminarily studied the theoretical convergence of the decomposition method [3]. In addition, a new working set selection strategy was developed. This new decomposition method becomes the foundation of the software BSVM [4].

In this project (NSC 89-2213-E-002-106), we extend earlier works and have the following achievements:

1. **More complete studies on the convergence of decomposition methods [7,8].**

A major drawback of the convergence proof in [3] is that it does not apply to any practical implementations. In other words, it proves only the convergence of some theoretical algorithms.

The main breakthrough is in [7] where we prove the asymptotic convergence of the popular software SVMlight [6]. The proof also applies to the integrated SVM library LIBSVM developed by our group [2].

The remaining issue of the proof in [7] is that it requires an assumption on problems whose kernel matrix (Hessian of the objective function) is not positive definite. Note that elements of the kernel matrix are inner products of training data after they are mapped into higher dimensional spaces. For some simpler situations this issue is solved in the paper [8] where no assumption is need for the convergence proof.

2. **Implementation of an integrated library for support vector machines. [2].**

After developing the software BSVM, some researchers not in the fields of pattern recognition and machine learning complained that our software is a little too complicated for them to use. Indeed many users of support vector machines do not know much about optimization. Motivated by users' comments, we designed a simple SVM software: LIBSVM [2].

LIBSVM was released in April 2000. At that time we still worked on the previous project. For the current project we extend it to a real library. Some particular features are as follows:

- It now solves different SVM formulations: C-Support vector classification and regression, nu-Support vector classification and regression, and one-class classification.

- It can efficiently solve multi-class classification. We have conducted a complete comparison on different multi-class methods for SVM [5].
- Currently it is the only SVM software that provides easy cross validation for model selection.
- It handles unbalanced data by a weighted SVM.
- It contains an easy-to-use GUI demonstrating SVM classification and regression.

From April 2000 to October 2001 there are more than 5,500 downloads of LIBSVM. With the help of users from different organizations we now have interfaces to other software:

- Matlab interface developed at Ohio State University
- R (a popular statistics software) interface developed at Vienna University of Technology.
- Python interface developed at HP Labs.

三、Conclusions

In the future we will try to keep the leading position in SVM software design. In addition, we will extend SVM to different applications by collaborating with researchers in different research areas.

四、Acknowledgements

Students who join this projects are Chih-Chung Chang, Chih-Wei Hsu, Jen-Hou Lee, and Kuan-Min Lin.

六、Bibliography

[1] C. J. C. Burges "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery,* 2(1998) 121-167.

[2] C.-C. Chang and C.-J. Lin. "LIBSVM: introduction and benchmarks," (2000) Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[3] C.-C. Chang, C.-W. Hsu and C.-J. Lin. "The analysis of decomposition methods for support vector machines," *IEEE Transactions on Neural Networks*. 11(2000), 1003-1008.

[4] C.-W. Hsu and C.-J. Lin. "A simple decomposition method for support vector machines," *Machine Learning*. 46(2002), 291-314. Software available at http://www.csie.ntu.edu.tw/~cjlin/bsvm

[5] C.-W. Hsu and C.-J. Lin. "A comparison of methods for multi-class support vector machines," To appear in IEEE Transactions on Neural Networks, 2002.

[6] T. Joachims. Making large-scale SVM learning practical. In B. Scholkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods - Support Vector Learning, Cambridge, MA, 1998. MIT Press.

[7] C.-J. Lin. "On the convergence of the decomposition method for support vector machines," *IEEE Transactions on Neural Networks* 12(2001), 1288-1298.

[8] C.-J. Lin. "Asymptotic convergence of an SMO algorithm without any assumptions," To appear in IEEE Transactions on Neural Networks, 2002.

[9] V. Vapnik. Statistical Learning Theory. John Wiley, New York, 1998.