Support Vector Machines Summary

Stephen Sinclair MUMT611, Winter 2006

28th March 2006

1 Introduction

Support Vector Machines (SVM) are a type of classification algorithm based on determining the optimal division between two sets of feature vectors. In the most simple form of SVM, this division is linear, while an extended form of SVM utilizing a "kernel function" allows non-linear classification. One characteristic of SVM classifiers is that they can operated efficiently on data with large feature sets, otherwise described as data with high dimensionality. This is useful for many pattern recognition tasks; in particular, musical applications based on spectral features tend to have high degrees of dimensionality.

2 Implementing SVM

As stated by Busuttil (2003), the SVM concept was developed in the 1970's by Vladimir Vapnik at Bell Laboratories. However, it began to receive much more attention in the last decade or so, due to the increasing importance of classification in dealing with large databases of information.

Additionally, some optimizations for training algorithms allowed SVM classifiers to be reasonably implemented. Bell Labs published an optimal margin training algorithm based on numerical methods of polynomial order in the number of training patterns, an efficiency comparable to other classifier algorithms (Boser et al. 1992).

A few years later, a thorough discussion of implementation issues in SVM theory was published by MIT's Artificial Intelligence Laboratory (Osuna et al. 1996). The report discusses the linear separable and non-separable cases, as well as non-linear approaches using the kernel trick, providing proofs with reference to structural risk minimization techniques. An extension for handling weighted training data is provided, as well as some example applications and experimental results on face recognition.

Schölkopf et al. (1996) described a technique for efficiently taking advantage of certain apriori knowledge, called invariance transformations. This consists of some transformation on the training data which is known not to change the labeling of the data. The technique described takes advantage of how SVM classifiers work by applying the transformations only to the support vectors identified in the original training set.

The following year, Osuna et al. (1997) published an improved algorithm for training SVM classifiers. The problem, they state, is that previous algorithms assumed small numbers of support vectors. For large numbers of support vectors, computation time became too long. The proposed algorithm decomposes the quadratic programming problem into subproblems which are more easily computed.

Burges and Schölkopf (1997) also published some improvements for SVM in terms of accuracy and speed. In fact, these improvements incorporate both the subproblem optimizations and invariance transformations into a single implementation. The increase in speed was claimed to be 22-fold, with greater accuracy than the original training method.

In 1998, Burges published an exhaustive tutorial on support vector machines, summarizing previous work on the subject and presenting the theory and implementation from a fresh perspective, complete with original proofs and some analytical examples (Burges 1998). This tutorial proved a good introduction to the topic, though rather thoroughly embedded with complicated mathematics. A reader is recommended to read Busuttil's 2003 overview of SVM before delving into this 40-page document (Busuttil 2003).

3 Improvements

Later work tended to concentrate on improving the generalization accuracy of SVM classifiers, as well as improving training times without sacrificing the former. Kwok (1999) published a method of moderating the outputs from SVM classifiers in terms of Bayesian confidence levels. The argument is that SVM classifiers, once trained, are considered able to classify without regard to confidence, when in reality this is inappropriate. Moderating the outputs with an "evidence framework" is a way to increase generalization.

Williamson et al. (1999) published a method for the use of entropy numbers in choosing an appropriate kernel function. It was an attempt to explain kernel function choice by more analytical means rather than previous ad-hoc or empirical methods. The entropy numbers associated with mapping operators for Mercer kernels is discussed.

In (Chapelle and Schölkopf 2002), it was discussed that previous work on invariance transformations was mostly appropriate only for linear SVM classifiers. For non-linear SVM classifiers, it discusses an analytical method of utilizing a kernel PCA map for incorporating invariance transformations.

In a more recent paper, Tsang et al. (2005) discussed a way to take advantage of the approximations inherent in kernel classifiers, by using the Minimum Enclosing Ball algorithm as an alternative means of speeding up training. Training time had previously been reduced mostly by modifying the training set in some way. Their final classifier, which they called the Core Vector Machine, converged in linear time with space requirements independant from number of data points.

References

Boser, B., I. Guyon, and V. Vapnik. 1992. A training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *Annual ACM Workshop on COLT*, Pittsburgh, PA, 144–52. ACM Press.

Burges, C. and B. Schölkopf. 1997. Improving the accuracy and speed of support vector machines. *Advances in Neural Information Processing Systems 9*, 375.

Burges, C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2*(2), 121–67.

Busuttil, S. 2003. Support vector machines. In *Proceedings of the Computer Science Annual Research Workshop*, Villa Bighi, Kalkara. University of Malta.

Chapelle, O. and B. Schölkopf. 2002. Incorporating invariances in non-linear support vector machines. In T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems*, Volume 14, 609–16. Cambridge, MA: MIT Press.

Kwok, J. 1999. Moderating the outputs of support vector machine classifiers. *IEEE-NN 10*(5), 1018.

Osuna, E., R. Freund, and F. Girosi. 1996. Support vector machines: Training and applications. Technical Report AIM-1602, Massachisetts Institute of Technology Artificial Intelligence Laboratory.

Osuna, E., R. Freund, and F. Girosi. 1997. An improved training algorithm for support vector machines. In J. Principe, L. Gile, N. Morgan, and E. Wilson (Eds.), *Neural Networks for Signal Processing VII: Proceedings of the 1997 IEEE Workshop*, New York, NY, 276–85. IEEE.

Schölkopf, B., C. Burges, and V. Vapnik. 1996. Incorporating invariances in support vector learning machines. In *Proceedings of the International Conference on Artificial Neural Networks*, Volume 1112, 47–52.

Tsang, I., J. Kwok, and P.-M. Cheung. 2005. Core vector machines: Fast svm training on very large data sets. *Journal of Machine Learning Research 6*, 363–92.

Williamson, R., A. Smola, and B. Schölkopf. 1999. *Entropy numbers, operators and support vector kernels*, pp. 127–44. Cambridge, MA: MIT Press.