Summary of Presentation on Support Vector Machine

This presentation is an overview about one of the most popular supervised pattern classifier the support vector machine (SVM), including its theory, implementation, and its application in music information retrieval (MIR).

Common supervised Bayesian classifiers often suffers from small sample-size effect or peak effect, that is to say, overly small or large sample-size results significant classification errors (Support vector classifiers: a first look). This is because these kinds of classifiers always try to estimate the probability densities of practical test data-set using finite sample sets, which can only generate inaccurate estimates. Therefore, the approaches that only take into account empirical risk (ER) are not adequate for generalization (Vapnik 1995).

Guided by Vapnik's suggestion of avoiding solving a more general problem as an intermediate step when dealing with an explicit problem, artificial neural network (ANN) came up as rescues to the problematic classification issues, which has greatly changed the old scene with improved generalization. Although ANN has largely avoided the estimation of probability densities, however, the lack of strong theoretical support makes ANN hard to control in terms of its convergence and generalization. In fact, the design of ANN highly depends on empirical knowledge of the designers (Vapnik 1995).

Vapnik and Cervonenkis took a thorough study into the theoretical issues of the validity of ER-based classification and came up with the VC theory, which successfully describes the complexity and the ability of generalization of a supervised classifier including ANN using a term VC dimension (Vapnik 1995).

Based on VC theory, Vapnik proposed the brand new classifier SVM, which focuses on the explicit problem of general classification: To find an optimal hyperplane that can separate samples of two classes with least structural risk (SR). Through utilizing the concept of margin, SVM can define a hyperplane "separating two (linearly) separable classes in such a way that its margin is largest" (Support vector classifiers: a first look), which ensures the SR to be as small as possible with a fixed ER. The problem can be boiled down to dot product based quadratic programming (Support vector classifiers: a first look).

In a linearly non-separable case, SVM uses a kernel function to map the original input data to a higher dimensional space in which the samples become linearly separable. In this way, SVM effectively controls the VC dimension to be finite with the decision function and its complexity completely denotable via the support vectors. The training cost of SVM thus does not depend on the dimensionality of feature space (Support vector classifiers: a first look).

As a classifier, SVM has many great advantages, largely due to its strong mathematical basis, including its excellent ability of generalization, its insensibility to “curse of dimensionality”, versatile choices of kernel function and so on. However, its design still calls for empirical knowledge when it comes to the choices of parameters. Therefore, it is still not an ideal candidate that can serve as a fairly generalized classifier (Support vector classifiers: a first look).

SVM has been successfully applied to domains such as speech recognition and face recognition. In MIR, there are only a few reported applications (Li and Guo 2000; Ferguson 2004), in which SVM has shown its nice performance. The potential of SVM as a practical classifier in MIR has yet to be tested in the future.

Bibliography

Ferguson, R. 2004. *Automatic Segmentation in Concert Recordings*. M. S. thesis, McGill University.

Li, S., and G. Guo. 2000. Content-Based Audio Classification and Retrieval Using SVM Learning. *Proceeding of the Special Session on Multimedia Information Indexing and Retrieval*: 156-63.

Tax, D., D. Ridder, and R. Duin. 1997. *Support vector classifiers: a first look* [online]. Delft, The Netherlands: Delft University of Technology, Faculty of Applied Physics, Pattern Recognition Group, 1997 [cited 16 March 2005]. Available from World Wide Web: (http://www.ph.tn.tudelft.nl/Research/neural/feature_extraction/papers/asci_97/a97.html)

Vapnik, V. 1995. *The nature of statistical learning theory*. Berlin: Springer-Verlag.