ELSEVIER

Neurocomputing 47 (2002) 303–304

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Book reviews[1]

Advances in Kernel Methods. Support Vector Learning, B. Scholkopf, C.J.C. Burges, A.J. Smola (Eds.); MIT Press, Cambridge, 1999, pp. 376+vii. ISBN 0-262-19416-3.

Support vector machines (SVM) have emerged as an interesting and conceptually appealing architecture of pattern recognition. They originated from early concepts developed by Vapnik and Chervonenkis. The essence of functioning of SVMs is in classifying patterns via hyperplanes. These hyperplanes are constructed in such a way that they separate the patterns belonging to different classes. The hyperplanes are made optimal in terms of an error margin and as such they solve a certain quadratic optimization problem. The solution vector is expressed via a subset of the training patterns called support vectors. The development of SVMs is inherently associated with a construction of a new high-dimensional space formed on a basis of the original data space encountered in the classification problem. This space, referred to as a feature space (being some other dot space) is formed by some nonlinear mapping. Importantly, the resulting dot products are evaluated by kernels (such as e.g., polynomials). With regard to the fundamentals of SVMs arise two questions: (a) how to construct a function mapping the original data space into a highly dimensional feature space, and (b) what properties should the linear function (hyperplane) have to provide with good approximation capabilities? Interestingly, there are strong theoretical results that help answer these questions. In particular, it is known that the generalization ability of the learning machine depends on the capacity of a set of functions (more specifically, the Vapnik–Chervonenkis, or briefly VC capacity of such set) rather than on the dimensionality of a space itself. The design of the SVMs translates into quadratic programs so an efficient solution to the classification problem calls for a comprehensive way of handling highly dimensional tasks of quadratic programming. Subsequently, this has been a major pursuit and an ongoing challenge in the construction of SVMs. The optimization problem becomes crucial when it comes to noisy data and then the scaling process becomes a burning issue. The applications of this class of classifiers are numerous and diversified.

This edited volume is devoted to the fundamentals, algorithms, and practice of support vector machines. In part, it resulted from the workshop at the annual Neural Information Processing Systems (NIPS) conference held in Breckenridge, Colorada in December 1997. The editors divided the book into four sections that reflect a way of presentation of the entire material. It starts from the theoretical part, moves to implementations, elaborates on applications, and finally includes a series of extensions of the existing fundamentals.

The first section consists of seven chapters authored by experts in the field. The first contribution written by Vapnik is an interesting and authoritative overview of the area. It lucidly delivers a conceptual background of SVMs and explains ways of improving generalization abilities through test points exploited during training (a process called here a transduction). The paper authored by Barlett and Shawe-Taylor concentrates on the generalized performance of SVMs; they also discuss effects of large margins of SVMs. The related presentation deals with the Bayesian voting schemes and large margin classifiers that are discussed by Cristianini and Shawe-Taylor. They reveal interesting links between machine learning and Bayesian methods.

The issue addressed by Opper is concerned with several into ways of coming up with a refinement of the VC dimension with intent of forming tighter bounds on the generalization error of the classifiers. This refinement is accomplished in the setting of statistical mechanics. Williamson et al. discuss the capacity control and kernel mapping. The primary goal there is to select the right kernel for a problem at hand.

The implementation section of the volume is mainly directed to various ways of efficient handling a problem of quadratic programming (QP) being the optimization heart of the SVM design. The three

[1] Material to be included in this section can be submitted to: Prof. K.J. Cios, Computer Science and Engineering Dept., University of Colorado at Denver, Campus Box 109, 1200 Larimer Street, Denver, CO 80217-3364, USA. E-mail: kcois@carbon.cudenver.edu

PII: S0925-2312(01)00684-1

contributions (by Kaufmann, Joachims, and Platt) cover algorithms of QP reviewed in relation to the training purposes of the classifiers including chunk training (based on ideas of Osuna et al.) and a sequential minimal optimization.

The application section of the volume (consisting of three papers authored by Mattera and Haykin, Muller et al., and Kressel) shows how SVMs are used in the dynamic reconstruction of chaotic systems, prediction of time series, and pattern classification (namely, optical character recognition). An interesting idea of utilizing various loss functions (especially those developed in the spirit of Huber's robust statistics) puts these application problems in a new and practically appealing setting.

The last section is devoted to the extension of the available algorithms and consists of five contributions. In general, they dwell on the existing well-known and widely used algorithmic developments such as principal component analysis and vector regression. First, Osuna and Girosi discuss ways of reducing the run-time complexity in SVMs. Stitson et al. look at the support vector regression with the use of ANOVA decomposition kernels. An important and practically omnipresent task of probability density estimation via support vectors is analyzed by Bennet. Finally, Scholkopf et al. introduce an idea of kernel principal component analysis (PCA). This nonlinear form of the PCA exploits Mercer kernels. It is shown how to apply this generalization to feature extraction.

Overall, the book is a timely and comprehensive compendium of knowledge about this interesting and promising category of the classification algorithms. Well-balanced theoretical and experimental aspects of the methodology and algorithms will please researchers as well as practitioners in the area.

Witold Pedrycz[a,b]
[a]*Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alta., T6G267 Canada*
[b]*Systems Research Institute, Polish Academy of Sciences Warsaw, Poland*
E-mail address: pedrycz@ee.ualberta.ca