

Comments on: Support vector machines maximizing geometric margins for multi-class classification

Shigeo Abe

Published online: 15 August 2014
© Sociedad de Estadística e Investigación Operativa 2014

The support vector machine (SVM) is one of the most popular classification algorithms. In the SVM, the minimum distance from the separating hyperplane to training samples of one class is called margin and the SVM is trained so that the margin is maximized under the constraint that the margin of one class is the same as that of the other. (The resulting separating hyperplane is called optimal separating hyperplane.)

Much research has been conducted to improve the generalization ability of the SVMs. Especially, because SVMs are basically binary classifiers, the extension to multi-class problems is not unique. Thus several multi-class SVMs have been developed.

Among them, divide-and-conquer types of SVMs are widely used: one-against-all (OAA) SVMs and one-against-one (OAO) SVMs. The OAA SVM consists of n binary classifiers that separate one class from the remaining classes, where n is the number of classes, and classifies a sample into the class with the maximum decision output. The OAO SVM consists of $n(n - 1)/2$ binary classifiers, which separate one class from another, and classifies a sample by voting.

Another approach is all-together (AT) SVMs, in which the decision functions that separate any one class from others are determined simultaneously.

The conventional wisdom is that the generalization abilities of AT SVMs are not so good compared to those of OAA or OAO SVMs although the computational complexity of AT SVMs is much higher.

A series of research by Tatsumi and Tanino is an attempt to defy this conventional wisdom. They notice that by the conventional AT SVM, unlike the binary SVM, margin

This comment refers to the invited paper available at doi:[10.1007/s11750-014-0338-8](https://doi.org/10.1007/s11750-014-0338-8).

S. Abe (✉)
Kobe University, Rokkodai, Nada, Kobe, Japan
e-mail: abe@kobe-u.ac.jp

maximization does not lead to maximizing the distance between the hyperplanes and the training samples (geometrical margin). And they formulate the multi-objective multi-class SVMs (MMSVMs) that maximize the geometrical margins. Through long and complicated theoretical analysis, they have developed the MMSVM-OA, which is based on the OAA SVM.

In training an MMSVM-OA, first an OAA SVM is trained. Then the parameter values for the MMSVM-OA are determined, and the MMSVM-OA is trained based on the decision functions determined by the OAA SVM. Computer experiments show the effectiveness of the MMSVM-OA over the OAA SVM and the conventional AT SVM.

Performance of multi-class SVMs can be measured by training time, classification time, and the generalization ability. Let me discuss their approach by these measures.

As for the training time, training of the OAA SVM becomes the overhead for the MMSVM-OA. Comparing the OAA and OAO SVMs, usually training of the OAO SVM is faster, although their generalization abilities are comparable. So to speed up training, is it possible to develop MMSVM-OO? Or is there any other way to speeding up training?

Directed Acyclic Graph (DAG) SVMs (Platt et al. 2000; Kijirikul and Ussivakul 2002), which are a variant of OAO SVMs, can speed up classification. Is it possible to incorporate this type of tree architecture to speed up classification?

In the literature several approaches have been discussed to improve generalization ability of SVMs. The optimal separating hyperplane is optimal when the distributions of class data are unknown. However, it is not optimal when they are known. For instance, if each class in a two-class problem obeys a different Gaussian distribution, the optimal separating hyperplane is no longer optimal. To improve generalization ability in such a situation, the Mahalanobis metric is introduced into SVMs (Lanckriet et al. 2002; Huang et al. 2006; Shivaswamy and Jebara 2007; Yeung et al. 2007; Xue et al. 2011; Abe 2012). Is it possible to introduce the Mahalanobis metric into MMSVM?

In the computer experiments, the best cross-validation (CV) accuracies are used to compare generalization abilities of the classifiers. The CV accuracies are often used in the literature, but they may introduce some bias in comparison.

To make comparison fair, in training the SVM, special care should be taken to avoid using the information of the test data set. Comparison with the best CV accuracy violates this condition: because the best CV accuracy is selected from among the CV accuracies for different hyperparameter values, the associated trained SVM implicitly uses the information of the test data set.

In Table 7, MMSVM is inferior to OAA, especially for the vehicle and glass data sets. Why does this happen?

Because of the kernel expansion (6), the solution of MMSVM is not sparse. But because MMSVM-OA is based on the solution of OAA, the solution is sparse. Therefore, MMSVM-OA is faster in training and classification than MMSVM. Does the comparison of variables shown in Table 4 reflect this fact?

In conclusions, Tatsumi and Tanino's approach is a new and innovative way of improving the generalization ability of multi-class SVMs, and I believe that the research will extend in many directions.

References

- Abe S (2012) Training Mahalanobis kernels by linear programming. In: Villa AEP et al (eds) ICANN 2012. Lecture notes in computer science, vol 7553. Springer, Berlin, pp 339–346
- Huang K, Yang H, King I, Lyu MR (2006) Learning large margin classifiers locally and globally. In: Proceedings of the twenty-first international conference on machine learning (ICML 2004), pp 1–8
- Kijisrikul B, Ussivakul N (2002) Multiclass support vector machines using adaptive directed acyclic graph. In: Proceedings of the 2002 international joint conference on neural networks (IJCNN '02), vol 1, Honolulu, Hawaii, pp 980–985
- Lanckriet GRG, El Ghaoui L, Bhattacharyya C, Jordan MI (2002) A robust minimax approach to classification. *J Mach Learn Res* 3:555–582
- Platt JC, Cristianini N, Shawe-Taylor J (2000) Large margin DAGs for multiclass classification. In: Solla SA, Leen TK, Müller K-R (eds) *Advances in neural information processing systems*, vol 12. MIT Press, Cambridge, MA, pp 547–553
- Shivaswamy PK, Jebara T (2007) Ellipsoidal kernel machines. In: *Proceedings of the eleventh international conference on artificial intelligence and statistics (AISTATS 2007)*
- Xue H, Chen S, Yang Q (2011) Structural regularized support vector machine: a framework for structural large margin classifier. *IEEE Trans Neural Netw* 22(4):573–587
- Yeung DS, Wang D, Ng WWY, Tsang ECC, Wang X (2007) Structured large margin machines: sensitive to data distributions. *Mach Learn* 68(2):171–200