A New Approach for Applying Support Vector Machines in Multiclass Problems Using Class Groupings and Truth Tables

Mauricio Kugler, Hiroshi Matsuo and Akira Iwata
mauricio@kugler.com, matsuo@nitech.ac.jp, iwata@nitech.ac.jp

Department of Computer Science and Engineering
Nagoya Institute of Technology
Gokiso-cho, Showa-ku, 466-8555, Nagoya, Japan

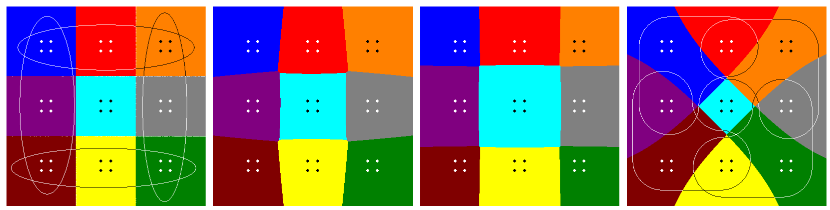
The Support Vector Machines (SVMs) had been showing a high capability of complex hyperplane representation and great generalization power. These characteristics lead to the development of more compact and less computational complex methods than the One-versus-Rest (OvR) and One-versus-One (OvO) [1] classical methods in the application of SVMs in multiclass problems. This paper proposes a new method for this task, named *Truth Table Fitting Multiclass SVM* (TTF-MCSVM), in which less SVMs are used than other classical methods. The main objective of this research is the development of an efficient method to be applied in problems with very large number of classes, like in the recognition of East Asian languages characters (e.g. Japanese and Chinese kanji).

The TTF-MCSVM is based on the combination of many simple binary SVMs, like the OvR and OvO. The N classes are divided in M combinations of two groups, where M is:

$$M = \lceil \log_2 N \rceil \tag{1}$$

The two groups obtained in each combination will correspond to two pseudo-classes, and these two pseudo-classes will form a grouping. As the M groupings are binary, they can be applied to M simple and independent binary SVMs. This approach can be also represented by the encoding scheme described in [2], in which large encodings were used to increase the classifiers performance. However, the focus in the present work is the opposite, i.e. to use the smallest possible encoding size, resulting in less SVMs than other classical methods. If the groupings are consistent (i.e. each class has a different associated pseudo-class combination), the M groupings are arranged in a Boolean truth table, where the inputs are the results of all SVMs and the output is the winner class, which is chosen by the overlap of the decision hyperplanes of the SVMs. During the training phase, each SVM is trained independently, and the margin is maximized to divide the two pseudo-classes of that SVM. On the test phase, the results of the M SVMs are applied on the truth table and, if the groupings are correct, it will be possible to choose the winner class.

A basic example is shown in Fig. 1. The same 9 class problem was solved using the TTF-MCSVM (Fig. 1(a), with the groupings drawn in the figure), OvR (Fig. 1(b)) and OvO (Fig. 1(c)) methods. All of them used gaussian kernel with σ=1.5 and C=100. Using the TTF-MCSVM method, less classifiers were used,



(a) (b) (c) (d)

Fig. 1. Nine classes problem: (a) TTF-MCSVM (4 SVMs, 16 SV, Grouping Set A), (b) OvR (8 SVMs, 47 SV), (c) OvO (28 SVMs, 108 SV), (d) TTF-MCSVM (4 SVMs, 20 SV, Grouping Set B)

less support vectors (SV) were found and a better generalization was obtained when compared to the other methods. However, a different grouping set, such as the one shown in (Fig. 1(d)), can clearly result in a smaller margin.

Table 1 shows the comparison of three UCI benchmark problems, *iris*, *wine* and *glass*, solved by the TTF-MCSVM (with random groupings), OvR and OvO methods. The results shows the smaller number of support vectors for the best performance parameters combinations, calculated over the average of a 10-fold cross validation (p means d for polynomial kernel and σ for gaussian kernel).

Table 1. UCI benchmark experiments comparison

Problem	**TTF-MCSVM**				**OvR**				**OvO**			
(N, samples)	M	CR%	#SV	p, C	M	CR%	#SV	p, C	M	CR%	#SV	p, C
Polynomial Kernel												
iris (3, 150)	2	99.33	31	2, 1.0	3	99.33	52	2, 0.9	3	94.67	95	2, 0.03
wine(3, 178)	2	98.82	50	1, 0.2	3	98.82	106	2, 0.07	3	98.82	40	1, 0.2
glass (6, 214)	3	66.19	304	2, 0.2	6	69.52	844	3, 0.004	15	70.95	502	2, 0.02
Gaussian Kernel												
iris (3, 150)	2	98.67	19	4, 80	3	98.67	43	3, 60	3	96.00	76	1, 1
wine(3, 178)	2	98.82	49	20, 100	3	99.41	87	10, 9	3	99.41	86	10, 3
glass (6, 214)	3	75.71	363	0.9, 4	6	75.71	603	1, 6	15	76.19	560	1, 4

The TTF-MCSVM performance shows no statistically significant difference from the OvR and OvO methods, using a smaller number of classifiers and, in almost all cases, considerably less SV. In the case of the *glass* problem, optimized groupings sets could increase the performance. The small number of classifiers, specially when the kernel matrix is calculated in advance, leads to smaller training time. Future works include the study on the groupings optimization and experiments dealing with more complex real world problems.

References

1. Jason Weston and Chris Watkins. Support vector machines for multi-class pattern recognition. In *Proceedings of the 7th European Symposium On Artificial Neural Networks*, pages 219–224, Bruges, 4 1999.
2. Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.