J. Astrophys. Astr. (2016) 37: 9

Stellar Spectral Classification with Minimum Within-Class and Maximum Between-Class Scatter Support Vector Machine

Liu Zhong-Bao
School of Computer and Control Engineering, North University of China, Taiyuan 030051, China.
e-mail: liu_zhongbao@hotmail.com

Received 12 November 2015; accepted 29 December 2015
DOI: 10.1007/s12036-016-9374-0

Abstract. Support Vector Machine (SVM) is one of the important stellar spectral classification methods, and it is widely used in practice. But its classification efficiencies cannot be greatly improved because it does not take the class distribution into consideration. In view of this, a modified SVM-named Minimum within-class and Maximum between-class scatter Support Vector Machine (MMSVM) is constructed to deal with the above problem. MMSVM merges the advantages of Fisher's Discriminant Analysis (FDA) and SVM, and the comparative experiments on the Sloan Digital Sky Survey (SDSS) show that MMSVM performs better than SVM.

Key words. Stellar spectral classification—support vector machine (SVM)—Fisher's discriminant analysis (FDA)—class distribution.

1. Introduction

Support Vector Machine (SVM) is one of the automatic spectral classification method and it is widely used in practice (Li *et al.* 2015; Qu *et al.* 2003). SVM is applied to classify the spectra from the RAdial Velocity Experiment (RAVE) and the Sloan Digital Sky Survey (SDSS) (Re Fiorentin *et al.* 2008). SVM is applied to classify the spectra to separate quasi-stellar objects from variable stars, non-variable stars and microlensing events (Kim *et al.* 2011). Isomap+SVM is used to classify the stellar spectral data in Bu *et al.* (2014). SVM tries to find an optimal separating hyperplane with the maximum margin, but it does not take the class distribution into consideration, and therefore, its classification efficiencies cannot be greatly improved. In order to solve the above problem, Minimum within-class and Maximum between-class scatter Support Vector Machine (MMSVM) is proposed in this paper. MMSVM tries to merge Fisher's Discriminant Analysis (FDA) and SVM where a modified class of SVM has been constructed, inspired by the optimization

of the Fisher's discriminant ratio. It is believed that the proposed classifiers have the advantages of both SVM and FDA.

Recently, stellar spectral subclasses classification based on Fisher criterion and manifold learning is analyzed by Liu & Song (2015), in which Modified Discriminant Analysis (MDA) based on Fisher Criterion and Manifold Learning is proposed to feature extraction, and then the traditional classifier SVM is used to classify the stellar spectra. The above classification method uses similar concepts with the proposed MMSVM method in this paper. However, there are huge differences between these two methods. In the method proposed by Liu & Song (2015), MDA is firstly used to extract the feature of the stellar spectra, and then SVM is applied in the low-dimensional subspace for classification. In this paper, PCA is firstly used to reduce the dimension of the stellar spectra and the proposed MMSVM is applied to classification. The data distribution is taken into consideration in these two methods and the purpose is to improve the classification efficiency. The data distribution is reflected in the MDA feature extraction method proposed by Liu and Song, while the data distribution is reflected in the process of classification in the proposed MMSVM method.

The rest of the paper is organized as follows. FDA and SVM are introduced in section 2, MMSVM is proposed in section 3; experiments are provided in section 4 and section 5 conclude our work.

2. Review of FDA and SVM

Here we will review FDA and SVM. In the binary classification, assume the training dataset $X = [x_1, x_2, \ldots, x_N]$, where $x_i (i = 1, \ldots, N)$ is the training sample, N is the data size. Two different classes c_{+1} and c_{-1} with number N_{+1} and N_{-1} respectively and the label $y_i \in \{+1, -1\}$.

2.1 *Review of FDA*

Fisher's Discriminant Analysis, also known as Fisher's Linear Discriminant Analysis (FLDA), is an important feature extraction method (Liu & Song 2015). The Fisher's discriminant ratio can be described by the following optimization equation:

$$J(W_{\text{opt}}) = \max_{W} \frac{W^T S_{\text{B}} W}{W^T S_{\text{W}} W} \tag{1}$$

where W is the projection matrix, S_{B} and S_{W} are respectively called the between-class scatter and the within-class scatter, which are defined as follows:

$$S_{\text{B}} = N_{-1}(m - m_{-1})(m - m_{-1})^T + N_{+1}(m - m_{+1})(m - m_{+1})^T, \tag{2}$$

$$S_{\text{W}} = \sum_{x \in c_{-1}} (x - m_{-1})(x - m_{-1})^T + \sum_{x \in c_{+1}} (x - m_{+1})(x - m_{+1})^T, \tag{3}$$

where m is the total mean vector of the dataset Xm_{+1} and m_{-1} are the mean sample vectors for the classes c_{+1} and c_{-1} respectively.

It can be seen from the above analysis that (1) FDA aims to maximize the ratio of the between-class scatter to the within-class scatter, thus maximizing the class discrimination; (2) FDA tries to keep the global structure of the training dataset during the process of feature extraction.

2.2 *Review of SVM*

The geometry explanation of SVM is to find a hyperplane with maximal classification margin to separate two classes (Bovolo *et al.* 2010). Let the hyperplane $w^T x + b = 0$ and the margin be $2/\|w\|$. The optimal problem of SVM can be summarized as follows:

$$\min J(w, b, \xi_i) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i, \tag{4}$$

subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \text{for } i = 1, 2, \ldots, N, \tag{5}$$

$$\xi_i \geq 0, \qquad \text{for } i = 1, 2, \ldots, N, \tag{6}$$

where the regularization parameter C is a constant to trade off the two goals. The larger the C, the more the error term is emphasized. Small C means that large classification margin is encouraged. The parameter ξ_i is the relaxation parameter which allows the existence of classification errors.

3. MMSVM

Although SVM performs well in practice, it does not take the class distribution into consideration, and therefore, its classification cannot be greatly improved. In view of this, Fisher's Discriminant Analysis is introduced to SVM and MMSVM is proposed. MMSVM takes both the samples in the boundaries and the distribution of the classes into consideration, and fully utilizes the advantages of FDA and SVM. The optimization problem of MMSVM is described as follows:

$$\min J(w, b, \xi_i) = \frac{1}{2} w^T (S_{\mathrm{W}} - S_{\mathrm{B}}) w + C \sum_{i=1}^{N} \xi_i, \tag{7}$$

subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \text{for } i = 1, 2, \ldots, N,$$

$$\xi_i \geq 0, \qquad \text{for } i = 1, 2, \ldots, N,$$

where S_{W} and S_{B} are respectively the within-class scatter and the between-class scatter. The definitions of w, C, ξ_i are the same with SVM. In equation (7), $\frac{1}{2} w^T (S_{\mathrm{W}} - S_{\mathrm{B}}) w$ tries to find the optimized projection, in which S_{W} and S_{B} reflect the data distribution according to their definitions. $S_{\mathrm{W}} - S_{\mathrm{B}}$ means that MMSVM tries to make the samples in the same class as close as possible and the samples with different classes far away from each other in the process of classification. $C \sum_{i=1}^{N} \xi_i$ denotes

the soft margin (Tefas *et al.* 2001), which can improve the robustness of the proposed classifier MMSVM. The solution of the MMSVM takes into consideration both the samples in the boundaries and the distribution of the classes and gives a robust solution.

The above optimal problem can be transformed to the following dual form based on Lagrangian theorem.

$$\max_{\alpha} \alpha^T 1 - \frac{1}{2}\alpha^T Q\alpha,$$

subject to

$$\alpha^T Y = 0, \qquad \alpha \geq 0,$$

where

$$Q = \frac{1}{2}[y_i y_j x_i^T (S_W - S_B)^{-1} x_j + \mu_{ij}],$$

$$\mu_{ij} = \begin{cases} \frac{1}{C} & i = j \\ 0 & i \neq j \end{cases}, Y = [y_1, y_2, \ldots, y_N]^T, 1 = [1, 1, \ldots, 1]^T.$$

The corresponding decision function is similar to SVM (Bovolo *et al.* 2010), but the definition of w is different, which can be computed by the above Lagrangian theorem and it can be obtained that $w = \sum_{i=1}^{N} \alpha_i y_i x_i (S_W - S_B)$,

$$f(x) = \text{sign}(w^T x + b) = \text{sign}\left(\frac{1}{2}\sum_{i=1}^{N} \alpha_i y_i x_i^T (S_W - S_B)^{-1} x + b_0\right).$$

The optimal threshold b_0 can be found by exploiting the fact that for all support vectors x_i with $0 < \alpha_i < C$, their corresponding slack variables ξ_i are zero, according to the KKT condition (Fletcher 1987). For any $x_i \in S = \{x_i | \alpha_i > 0, i = 1, 2, ..., N\}$, the following equation holds:

$$y_i \left(\frac{1}{2}\sum_{j=1}^{N} \alpha_j y_j x_j^T (S_W - S_B)^{-1} x_i + b_0\right) = 1.$$

Averaging over these patterns yields a numerically stable solution for the bias term,

$$b_0 = \frac{1}{N}\sum_{i \in S}\left(y_i - \frac{1}{2}\sum_{j=1}^{N} \alpha_j y_j x_j^T (S_W - S_B)^{-1} x_i\right).$$

4. Experimental analysis

We will investigate the performance of MMSVM compared with SVM in this section. The experimental datasets are from Sloan Digital Sky Survey (SDSS), Data Release 8 (Almeida & Prieto 2013). The datasets consist of 4 subclasses of K-type spectra: K1-type, K3-type, K5-type and K7-type, whose signal-to-noise ratios (SNRs) with $10 < \text{SNRs} < 20$, 3 subclasses of F-type spectra: F2-type, F5-type and

Table 1a. The total number of K starts with $10 < \text{SNRs} < 20$.

Stellar subclass type	K1	K3	K5	K7
Number	5505	6092	4597	4476

Table 1b. The total number of F starts with $50 < \text{SNRs} < 60$.

Stellar subclass type	F2	F5	F9
Number	1416	8156	13785

Table 2a. The comparative experimental results on the K-type dataset.

Training size	Test size	SVM	MMSVM
6201	14469	0.5801	0.6781
8268	12402	0.5966	0.6776
10335	10335	0.7198	0.8043
12402	8268	0.7891	0.8602
14469	6201	0.8704	0.9199
Average classification accuracy		0.7112	0.7880

Table 2b. The comparative experimental results on the F-type dataset.

Training size	Test size	SVM	MMSVM
7007	16350	0.5102	0.5368
9343	14014	0.6012	0.7222
11679	11678	0.6333	0.7708
14014	9343	0.7003	0.7484
16530	7007	0.7438	0.8710
Average classification accuracy		0.6378	0.7298

F9-type, whose SNRs with $50 < \text{SNRs} < 60$, are listed in Tables 1a, b. These data with the wavelength coverage from 3800 Å to 9000 Å have been shifted to a common rest-frame and normalized to a constant total flux.

The classification process is composed of the following steps:

Step 1. The experimental dataset is divided into two parts, one for training and the other for test.
Step 2. Principal Component Analysis (PCA) (Deeming 1964) is used to reduce the dimension of the stellar spectra.
Step 3. The test dataset is mapped to the low-dimensional subspace.
Step 4. SVM and MMSVM are respectively applied in the low-dimensional subspace for classification.

In our experiments, 30%, 40%, 50%, 60%, 70% of K-, F- spectral datasets are respectively used for training, and the remaining are used for test. The original data are projected to 5-dimensional space. Table 2a, b give the comparative experimental results.

It can be seen from Tables 2a, b that the classification accuracies of SVM and MMSVM rise with the size of training datasets. In the view of the classification results, MMSVM performs better than SVM. From average performance, we can see the classification accuracy of MMSVM is obviously higher than that of SVM.

5. Conclusions

SVM is widely used in automatic stellar spectra classification. One of its serious shortages is that it does not take the class distribution into consideration and therefore its classification efficiencies are limited. In view of this, MMSVM is proposed in this paper. In MMSVM, the between-class scatter and the within-class scatter in FDA, which describe the data distribution, are introduced to SVM. Comparative experiments on the SDSS dataset show that MMSVM outperforms SVM.

Acknowledgements

The author is very grateful to the anonymous referee for many useful comments and suggestions. This work is supported by Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2014142, 20130609).

References

Almeida, J. S., Prieto, C. A. 2013, *AJ*, **763**, 1.
Bovolo, F., Bruzzone, L., Carlin, L. 2010, *IEEE TIP*, **19(11)**, 2983.
Bu, Y. D., Chen, F. Q., Pan, J. C. 2014, *New Astronomy*, **28**, 35.
Deeming, T. J. 1964, *MNRAS*, **127**, 493.
Fletcher, R. 1987, *Practical Methods of Optimization,* 2nd ed. New York: Wiley.
Kim, D. W., Protopapas, P., Byun, Y. I. *et al.* 2011, *APJ*, **735(2)**, 68.
Li, H., Chung, F. L., Wang, S. T. 2015, *Applied Soft Computing*, **36**, 228.
Liu, Z. B., Song, L. P. 2015, *PASP*, **127**, 789.
Qu, M., Shih, F. Y., Jing, J. *et al.* 2003, *Journal of Solar Physics*, **217**, 157.
Re Fiorentin, P., Bailer-Jones, C. A., Beers, T. C. *et al.* 2008, *American Institute of Physics Conference Series*, **1082**, 76.
Tefas, A., Koreopoulos, C., Pitas, I. 2001, *IEEE Trans. Pattern Anal. Mach. Intell.*, **23(7)**, 735.