J. Astrophys. Astr. (June 2017) 38:21
DOI 10.1007/s12036-017-9441-1

Classification of Stellar Spectra with Fuzzy Minimum Within-Class Support Vector Machine

LIU ZHONG-BAO[1,*], SONG WEN-AI[1], ZHANG JING[1] and ZHAO WEN-JUAN[2]

[1]School of Software, North University of China, Taiyuan 030051, China.
[2]School of Information, Business College of Shanxi University, Taiyuan 030031, China.
*Corresponding author. E-mail: liu-zhongbao@hotmail.com

MS received 14 October 2016; accepted 28 February 2017; published online 19 June 2017

Abstract. Classification is one of the important tasks in astronomy, especially in spectra analysis. Support Vector Machine (SVM) is a typical classification method, which is widely used in spectra classification. Although it performs well in practice, its classification accuracies can not be greatly improved because of two limitations. One is it does not take the distribution of the classes into consideration. The other is it is sensitive to noise. In order to solve the above problems, inspired by the maximization of the Fisher's Discriminant Analysis (FDA) and the SVM separability constraints, fuzzy minimum within-class support vector machine (FMWSVM) is proposed in this paper. In FMWSVM, the distribution of the classes is reflected by the within-class scatter in FDA and the fuzzy membership function is introduced to decrease the influence of the noise. The comparative experiments with SVM on the SDSS datasets verify the effectiveness of the proposed classifier FMWSVM.

Keywords. Methods: data analysis—methods: statistical—techniques: spectroscopic—astronomical data bases: miscellaneous—stars: fundamental parameters—stars: statistics

1. Introduction

With the development of telescopes, massive spectra have been obtained and how to automatically deal with these data has become a hotspot. In recent years, many classification methods are proposed by researchers. Two schemes for automated classification of stellar spectra, namely, χ^2-minimization and artificial neural network (ANN), are presented by Gulati and Gupta (1994). A fast and robust method of classifying a library of optical stellar spectra for O to M type stars is presented by Harinder *et al.* (1998). The method employs, as tools: (1) principal component analysis (PCA) for reducing the dimensionality of the data and (2) multilayer back propagation network (MBPN) based artificial neural network scheme to automate the process of classification. An artificial neural network (ANN) scheme has been employed that uses a supervised back-propagation algorithm to classify 2000 bright sources from the Calgary database of InfraRed Astronomical Satellite (IRAS) spectra in the 8–23 μm region (Gupta *et al.* 2004). An artificial neural network (ANN) technique is employed for classifying stars by using synthetic spectra in the ultraviolet (UV) region from 1250 to 3220 Å as the training set and International Ultraviolet Explorer (IUE) low-resolution spectra as the test set by Bora *et al.* (2008). Probabilistic Neural Network (PNN) is used for automatic classification of about 5000 SDSS spectra into 158 spectral type of a reference library ranging from O type to M type stars by Bazarghan *et al.* (2008).

Support vector machine (SVM) is a typical classification in data mining, machine learning and other related areas. A lot of attention is being paid to SVM by scientists and a lot of improved methods are being proposed. The proximal support vector machine is used to multi-task learning (Li *et al.* 2015). In order to solve the unbalanced classification problem, Near-Bayesian support vector machine (NBSVM) is proposed by Datta and Das (2015). The ramp loss nonparallel support vector machine (RNPSVM), a sparse and robust nonparallel hyperplane, is proposed by Liu *et al.* (2015). The nonparallel support vector machine (NPSVM), a novel nonparallel classifier, is proposed by Re Fiorentin *et al.* (2008). SVM is also quite popular in the research of astronomy, especially in the area of automatic spectra classification. SVM is used to classify spectra classification based on the dimension reduction method principal component analysis (PCA) (Re Fiorentin *et al.* 2008). SVM can also be used to separate different kinds of spectra, such as quasi-stellar objects, variable stars and non-variable stars (Kim *et al.* 2011). The method ISOMAP is used for dimension reduction and SVM is used to classify stellar spectra (Bu *et al.* 2014).

In the research of spectra classification with SVM, we have proposed a series of improved methods. Stellar spectral classification with locality preserving projections (LPP) and SVM is analysed by Liu (2016), in which LPP is used for dimension reduction and SVM is used to realize the stellar spectral classification. Meanwhile, another stellar spectral classification method based on Fisher criterion and manifold learning is proposed by Liu and Song (2015), in which a novel dimension reduction method modified discriminant analysis (MDA) based on Fisher criterion and manifold learning is proposed to reduce the dimensions and SVM is used to classify the stellar spectra.

Although SVM performs well in practice, it does not take the distribution of the classes into consideration, and its classification efficiencies are greatly influenced by the noise. In view of this, fuzzy minimum within-class support vector machine (FMWSVM) is proposed, in which the within-class scatter S_{W} in Fisher's discriminant analysis (FDA) is utilized to describe the distribution of the classes and the fuzzy membership function is introduced to describe the importance of the samples.

The rest of the paper is organized as follows: the background knowledge is introduced in section 2; FMWSVM is proposed in section 3; experiments are provided in section 4, and section 5 concludes our work.

2. Background knowledge

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 and N is the data size. Two different classes c_1 *and* c_2 with numbers N_1 and N_2 respectively and their label $y_i \in \{+1, -1\}$ are assumed.

2.1 *Review of SVM*

The optimization problem of SVM is to construct a hyperplane to separate two classes with a large margin (Bovolo *et al.* 2010). The hyperplane can be represented as $w^T x + b = 0$ and its corresponding margin is $2/\|w\|$, where w is the normal vector of the hyperplane. The optimization problem of SVM can be described as follows.

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

subject to

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

$$\xi_i \geq 0, \quad \text{for } i = 1, 2, \ldots, N, \tag{3}$$

where the constant C is the regularization parameter. The larger the C, the more the error term is emphasized. In order to allow existence of classification errors, the relaxation parameter ξ_i is introduced.

2.2 *Review of FDA*

Fisher's discriminant analysis is one of the typical dimension reduction methods. In FDA, two matrices namely the between-class scatter S_{B} and the within-class scatter S_{W} are defined as follows.

$$S_{\mathrm{B}} = N_1 (m - m_1)(m - m_1)^T + N_2 (m - m_2)(m - m_2)^T, \tag{4}$$

$$S_{\mathrm{W}} = \sum_{x \in c_1} (x - m_1)(x - m_1)^T + \sum_{x \in c_2} (x - m_2)(x - m_2)^T, \tag{5}$$

where m is the total mean of the dataset X, m_1 and m_2 are respectively the mean for the classes c_1 and c_2. S_{B} and S_{W} are respectively represented as the distance of each class center and each sample to its class center.

The optimization problem of FDA is to maximize the ratio of the between-class scatter to the within-class scatter, which can be described as follows:

$$J(W_{\mathrm{opt}}) = \max_{\mathrm{W}} \frac{W^T S_{\mathrm{B}} W}{W^T S_{\mathrm{W}} W}, \tag{6}$$

where W is the projection matrix, which satisfies the distances of the samples in different classes as far as possible and meanwhile, the distances of the data in the same class as close as possible. In summary, FDA can keep the distribution of the classes as invariant as possible.

2.3 *Review of fuzzy membership function*

Fuzzy technology is used to deal with the uncertainty problem. The possibility of a sample belonging to a class is characterized by the fuzzy membership of 0 to 1. The common fuzzy membership functions include the distance-based function and the density-based function.

(1) The distance-based fuzzy membership function can be defined as follows.

$$s(x_i) = 1 - \frac{\|x_i - \bar{x}\|}{R} + \delta,$$

where x_i is the sample, $\bar{x}$ is the class center, R is the radius of the class, which satisfies $R = \max_i \|x_i - \bar{x}\|$. The constant δ is small positive, which satisfies $s(x_i) > 0$.

(2) The density-based function can be defined as follows.

$$s_{i1} = \begin{cases} \frac{\delta + d_{i1}}{R_1}, & d_{i1} \leq T\varepsilon \\ \delta, & d_{i1} > T\varepsilon \end{cases}; \quad s_{i2} = \begin{cases} \frac{\delta + d_{i2}}{R_2}, & d_{i2} \leq T\varepsilon \\ \delta, & d_{i2} > T\varepsilon \end{cases},$$

where $R_1 = \max_i \|x_i - \bar{x}_1\|$ and $R_2 = \max_i \|x_i - \bar{x}_2\|$ are respectively the radius of each class. $\bar{x}_1$ and $\bar{x}_2$ are respectively the center of each class, $T = \|\bar{x}_1 - \bar{x}_2\|$ is the distance between each class center. $d_{i1} = \|x_i - \bar{x}_1\|$ and $d_{i2} = \|x_i - \bar{x}_2\|$ are respectively the distances between the samples in each class to their class center. The parameter ε holds $\varepsilon > 0$ and δ is small and positive, which satisfies $s_i > 0$.

3. FMWSVM

SVM has been widely used and it performs well in practice, while its classification efficiencies can not be greatly improved because of two limitations: (1) It does not take the class distribution into consideration and results to a non-robust solution; (2) Its classification accuracies are greatly influenced by the noise. In view of this, inspired by the maximization of the FDA and the SVM separability constraints, the FMWSVM has been introduced. FMWSVM takes into consideration both the distribution of the classes and the noise resistance and gives a robust solution. The optimization problem of FMWSVM is described as follows.

$$\min_{w,b,\xi} J(w, b, \xi_i) = w^T S_W w + C \sum_{i=1}^{N} s_i \xi_i, \tag{7}$$

subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \text{for } i = 1, 2, \ldots, N,$$
$$\xi_i \geq 0, \quad \text{for } i = 1, 2, \ldots, N,$$

where S_W is the within-class scatter in FDA. The definitions of w, C, ξ_i are the same with SVM. In equation (7), $w^T S_W w$ tries to find an optimized projection, in which S_W reflects the class distribution according to its definition. $C \sum_{i=1}^{N} s_i \xi_i$ denotes the soft margin (Tefas *et al.* 2001), which can improve the robustness of the proposed classifier FMWSVM. The fuzzy membership function s_i reflects the importance of the sample. As the value of s_i becomes larger, the more its importance. The noise is given a small s_i, if its influence to the classification efficiency is decreased.

When the data size is larger than its dimension, the within-class scatter S_W is nonsingular, if not, S_W is singular and thus its inverse can not be obtained. In this paper, the Singular Value Perturbation is added to the matrix S_W in order to keep it nonsingular.

The above optimization problem is reformulated to the Wolf dual problem based on Lagrangian theorem.

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T S_W x_j - \sum_{i=1}^{N} \alpha_i,$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0, 0 \leq \alpha_i \leq s_i C \quad (i = 1, 2, \ldots, N).$$

The corresponding decision function is similar to SVM, but the definition of w is different. It can be obtained by the above Lagrangian theorem.

$$w = \frac{1}{2} S_W^{-1} \sum_{i=1}^{N} \alpha_i y_i x_i.$$

The corresponding decision surface is

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

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 conditions (Fletcher 1987). The KKT conditions are necessary conditions that a solution to a general nonlinear-programming problem must satisfy, provided that the problem constraints satisfy a regularity condition called constraint qualification. If the problem is one in which the constraint set is convex and the maximizing (minimizing) objective function is concave (convex), the KKT conditions are sufficient. Applied to a linear-programming problem, the KKT conditions yield the complementary slackness conditions of the primal and dual problems.

For any $x_i \in S = \{x_i | \alpha_i > 0, i = 1, 2, \ldots, N\}$, 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 performances of our proposed method FMWSVM and the traditional SVM in this section. The datasets from Sloan Digital Sky Survey (SDSS), Data Release 8 are used in our experiment. 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) are $30 < \text{SNRs} < 40$, 3 subclasses of F-type spectra: F2-type, F5-type and F9-type, whose SNRs are $60 < \text{SNRs} < 70$, as listed in Tables 1–2. Table 1 shows the total number of K stars with $30 < \text{SNRs} < 40$, the number of K-type subclasses, including K1, K3, K5, K7, are respectively 5505, 6108, 5151, 2689, which corresponds to the number of spectra. Table 2 shows the total number of F stars with $60 < \text{SNRs} < 70$, the number of F-type subclasses, including F2, F5, F9, are respectively 810, 4136 and 7588. 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 distance-based fuzzy membership function is used in our experiment.

The main process of our experiment is as follows. Firstly, the experimental dataset is divided into two parts: the training dataset and the test dataset. Secondly, the dimension reduction method PCA (Deeming 1964), a dimensionality-reduction technique that is often used to transform a high-dimensional dataset into a smaller-dimensional subspace prior to running a machine learning algorithm on the data, is used to preprocess the stellar spectra. Thirdly, the experimental datasets are mapped to the lower dimension by PCA. Fourthly, the decision surface of FMWSVM and SVM are respectively trained on the training dataset. Finally, FMWSVM and SVM are used for classification on the test dataset and we can obtain their classification accuracies.

Table 1. The total number of K stars with $30 < \text{SNRs} < 40$.

Stellar subclass type	K1	K3	K5	K7
Number	5505	6108	5151	2689

Table 2. The total number of F stars with $60 < \text{SNRs} < 70$.

Stellar subclass type	F2	F5	F9
Number	810	4136	7588

Table 3. The comparative experimental results on the K-type dataset.

Training size	Test size	SVM	FMWSVM
30% (5836)	70% (13617)	0.6500	0.7680
40% (7781)	60% (11672)	0.6820	0.7575
50% (9727)	50% (9726)	0.8066	0.8870
60% (11672)	40% (7781)	0.8782	0.9302
70% (13617)	30% (5836)	0.9304	0.9698
Average classification accuracy		0.7894	0.8625

Table 4. The comparative experimental results on the F-type dataset.

Training size	Test size	SVM	FMWSVM
30% (3760)	70% (8774)	0.5597	0.5898
40% (5014)	60% (7520)	0.6396	0.7600
50% (6267)	50% (6267)	0.6769	0.7997
60% (7520)	40% (5014)	0.7459	0.7908
70% (8774)	30% (3760)	0.7918	0.8886
Average classification accuracy		0.6828	0.7658

The training dataset consists of 30%, 40%, 50%, 60%, 70% of K-, F- type spectra respectively and the remaining are used for testing in the following experiments. The reduced dimension of each spectra is 5 by PCA. The comparative experimental results are shown in Tables 3–4 and Fig. 1.

It can be seen from Tables 3-4 and Fig. 1 that the performances of SVM and FMWSVM are well and in view of classification accuracies. Both of them met our requirements for stellar spectra classification. The accuracies show an increase tendency with the size of the training dataset rise. Compared with SVM, FMWSVM performs better on the different-size training datasets. The above conclusion is also tenable in the view of average classification accuracies. Based on the above analysis, it can be concluded that the classification performance of FMWSVM is better than SVM.

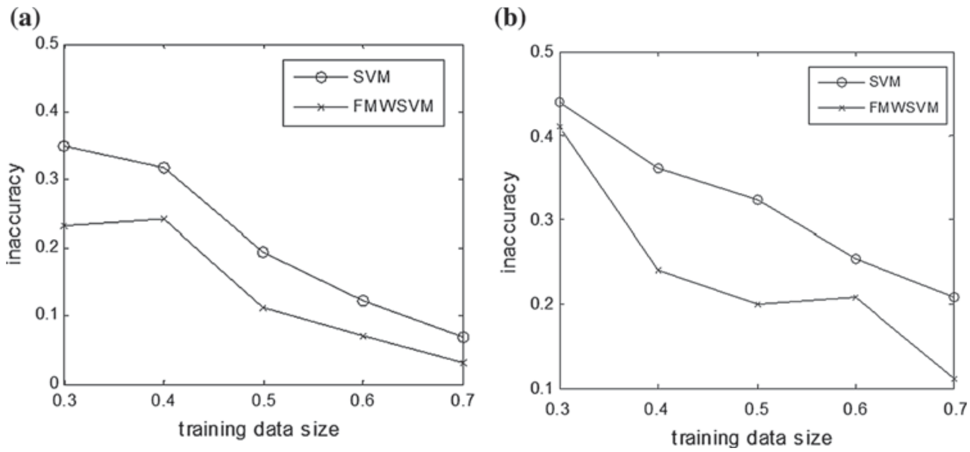

Figure 1. The relationships between the inaccuracy and the training data size on the K-type and F-type datasets. **(a)** The relationship between the inaccuracy and the training data size on the K-type dataset. **(b)** The relationship between the inaccuracy and the training data size on the F-type dataset

5. Conclusions

SVM is a typical classification method and widely used in astronomy, especially in automatic spectral classification. Although it performs well, in practice, its classification accuracies can not be greatly improved because it does not take the class distribution into consideration and it is sensitive to the noise. In view of this, inspired by the maximization of the FDA and the SVM separability constraints, FMWSVM is proposed in this paper. In FMWSVM, the distribution of the classes is reflected by the within-class scatter in FDA and the influence of noise to the classification efficiency can be decreased by introducing the fuzzy membership function. The comparative experiments on the SDSS datasets show that the proposed classifier FMWSVM performs better than SVM.

References

Bazarghan, M., Gupta, R. 2008, *Astrophys. Space Sci.*, **315**, 201.

Bora, A., Gupta, R., Harinder, P. S. *et al.* 2008, *MNRAS*, **384(2)**, 827.

Bovolo, F., Bruzzone, L., Carlin, L. 2010, *IEEE TIP*, **19(11)**, 2983.

Bu, Y. D., Chen, F. Q., Pan, J. C. 2014, *New Astron.*, **28**, 35.

Datta, S., Das, S. 2015, *Neural Netw.*, 70, 39.

Deeming, T. J. 1964, *MNRAS*, **127**, 493.

Fletcher, R. 1987, *Practical Methods of Optimization*, 2nd ed. (New York: Wiley).

Gulati, R. K., Gupta, R.1994, *ApJ*, **426(1)**, 340.

Gupta, R., Harinder, P. S., Volk K. *et al.* 2004, *ApJS*, **152(2)**, 201.

Harinder, P. S., Gulati, R. K., Gupta, R. 1998, *MNRAS*, **295(2)**, 312.

Kim, D. W., Protopapas, P., Byun, Y. I. *et al.* 2011, *APJ*, **735(2)**, 68.

Li, H., Chung, F. L., Wang, S. T. 2015, *Appl. Soft Comput.*, **36**, 228.

Liu, D. L., Shi, Y., Tian, Y. J. 2015, *Knowl. Based Syst.*, **85**, 224.

Liu, Z. B. 2016, *J. Astrophys. Astr.*, **37**(2), 1.

Liu, Z. B., Song, L. P. 2015, *PASP*, **127**(954), 789.

Re Fiorentin, P., Bailer-Jones C. A., Beers, T. C. *et al.* 2008, *Am. Inst. Phys. Conf. Ser.*, **1082**, 76.

Tefas, A., Koreopoulos, C., Pitas, I. 2001, *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(7), 735.

Tian, Y. J., Qi, Z. Q., Ju, X. C. *et al.* 2014, *IEEE Trans. Cybern.*, **44**(7), 1067.