ICIC Express Letters
Volume **11**, Number **3**, March **2017**
 ISSN 1881-803X
pp. **599–604**

FUZZY SEMI-SUPERVISED SUPPORT VECTOR MACHINE BASED ON FCS CLUSTERING ALGORITHM

Qing Wu[1,2,*], Yi Bai[1] and Leyou Zhang[3]

[1]School of Automation
Xi'an University of Posts and Telecommunications
No. 563, Chang'an South Road, Yanta District, Xi'an 710061, P. R. China

[2]State Key Laboratory for Strength and Vibration of Mechanical Structures
Xi'an Jiaotong University
No. 28, Xianning West Road, Xi'an 710049, P. R. China
*Corresponding author: xiyouwuq@126.com

[3]School of Mathematics
Xidian University
No. 2, South Taibai Road, Xi'an 710071, P. R. China
lyzhang@mail.xidian.edu.cn

Received August 2016; accepted November 2016

Abstract. *In order to utilize a large number of unlabeled data in semi-supervised support vector machine, fuzzy semi-supervised support vector machine based on a fuzzy compactness and separation (FCS) clustering algorithm (FS^{3}VM-FCS) is proposed. Compared with fuzzy c-means (FCM) clustering algorithm, FCS clustering algorithm proposed by Wu, Yu and Yang achieves a large between-cluster variation besides a small within-cluster variation. Fuzzy membership acquired by FCS clustering algorithm is utilized to label unlabeled data and solve the problem of noise points. The experimental results on artificial and real datasets show that the proposed FS^{3}VM-FCS has a good classification performance.*
Keywords: FCS, Clustering, Semi-supervised support vector machine, Fuzzy membership

1. **Introduction.** Support vector machine (SVM), which is an effective tool to solve some problems of data mining by using optimization method, has a good generalization ability [1]. SVM has drawn much attention in recent years [2-4]. Standard SVM belongs to supervised learning and it overcomes some traditional problems, such as the curse of dimensionality and over-fitting problem. Although SVM has the effectiveness of solving practical problems, it needs a lot of labeled data to train a classifier. However, it is difficult to obtain labeled data. How to effectively utilize a large number of unlabeled data and a small number of labeled data to obtain a classifier has become an important problem. Semi-supervised learning is a situation in which in your training data some of the samples are not labeled. Therefore, semi-supervised support vector machine (S^{3}VM) was proposed [5] based on the idea of semi-supervised learning. There are many researches about the applications of S^{3}VM recently, such as its applications in a brain computer interface (BCI) system [6], physical education effect evaluation [7], natural hazards forecasting [8], and classifying the bugs [9].

Benbrahim and Bramer proposed a fuzzy semi-supervised support vector machine [10] which applies fuzzy c-means (FCM) clustering algorithm to S^{3}VM. However, FCM clustering algorithm is based on a within-cluster scatter matrix with a compactness measure [11]. Compared with FCM clustering algorithm, FCS clustering algorithm proposed by Wu, Yu and Yang is based on a between-cluster scatter matrix with a separation measure and a within-cluster scatter matrix with a compactness measure. FCS clustering

algorithm attempts to minimize the compactness measure and simultaneously maximize the separation measure [11]. In this paper, FS^{3}VM based on FCS clustering algorithm (FS^{3}VM-FCS) is proposed. FS^{3}VM-FCS can be described as two steps. Firstly, a large number of unlabeled data will be labeled by fuzzy membership obtained by FCS clustering algorithm. Every newly labeled sample has its own fuzzy membership value solving the problem of noise points. Then FS^{3}VM-FCS trains a new classifier using a large number of newly labeled data and a small number of labeled data.

The rest of the paper is organized as follows. FCS clustering algorithm will be described in Section 2. FS^{3}VM-FCS is presented in Section 3. Section 4 shows the experimental results. At last, the brief conclusions of this paper will be given in Section 5.

2. **Fuzzy Compactness and Separation Clustering Algorithm.** Machine learning methods mainly include supervised learning and unsupervised learning. As a fairly common method in classification problems, supervised learning utilizes labeled data to train an optimal model. However, unlabeled data can be obtained relatively easily from the world. So the application of unsupervised learning is widely researched. Clustering algorithm is an unsupervised learning method. Traditional clustering algorithms belong to hard clustering, such as hard c-means (HCM) clustering algorithm [12]. HCM is an either-or clustering algorithm. That is to say a sample belongs to either one class or the other. In fact, every sample does not exactly belong to one of the two classes, so every sample should be assigned a fuzzy membership value. Therefore, Dunn proposed the FCM clustering algorithm [13] that utilizes fuzzy membership to label unlabeled data. Every sample has its fuzzy membership value between 0 and 1.

Hybrid dataset $X = (x_1, x_2 \ldots, x_n)$, FCM clustering algorithm utilizes Euclidean distance as a measure of similarity and X has c classes. FCM clustering adopts the sum-of-squared-error criterion and then its objective function can be defined as follows

$$J_{FCM}(U, V) = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} \left\| x_i - v_j \right\|^2, \tag{1}$$

where v_j is the clustering center. μ_{ij} is the fuzzy membership value of x_i belonging to class j. $\sum_{j=1}^{c} \mu_{ij} = 1$, $j = 1, 2, \ldots, c$.

Because the objective function of FCM clustering algorithm is based on a within-cluster scatter matrix, one sample point is close to its similar sample points. In fact, between-cluster scatter matrix and within-cluster scatter matrix are both essential to the objective function of clustering algorithm. By revising the cluster validity index $FS(c) = trace(S_{fw}) - trace(S_{fb})$ [14], the objective function of FCS clustering algorithm is defined as follows

$$J_{FCS}(U, V) = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} \left\| x_i - v_j \right\|^2 - \sum_{j=1}^{c} \sum_{i=1}^{n} \eta_j \mu_{ij}^{m} \left\| v_j - \bar{x} \right\|^2, \tag{2}$$

where $\bar{x}$ is sample mean, $\sum_{j=1}^{c} \mu_{ij} = 1$, $m > 1$ and $\eta_j \geq 0$. If $\eta_j = 0$, then $J_{FCS} = J_{FCM}$, else if $\eta_j = 1$, then $J_{FCS} = FS(c)$. The first part of $J_{FCS}(U, V)$ is based on a fuzzy within-cluster scatter matrix, and the second part of $J_{FCS}(U, V)$ is based on a fuzzy between-cluster scatter matrix. The Lagrange function is given by

$$L_{FCS} = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} \left\| x_i - v_j \right\|^2 - \sum_{j=1}^{c} \sum_{i=1}^{n} \eta_j \mu_{ij}^{m} \left\| v_j - \overline{x} \right\|^2 + \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{c} \mu_{ij} - 1 \right). \tag{3}$$

By taking derivatives with respect to v_j and μ_{ij}, we can obtain

$$v_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m x_i - \sum_{i=1}^{n} \eta_j \mu_{ij}^m \overline{x}}{\sum_{i=1}^{n} \mu_{ij}^m - \sum_{i=1}^{n} \eta_j \mu_{ij}^m}, \tag{4}$$

$$\mu_{ij} = \frac{\left(\|x_i - v_j\|^2 - \eta_j \|v_j - \overline{x}\|^2\right)^{-\frac{1}{m-1}}}{\sum_{k=1}^{c} \left(\|x_i - v_k\|^2 - \eta_k \|v_k - \overline{x}\|^2\right)^{-\frac{1}{m-1}}}, \tag{5}$$

where $\mu_{ij} \in [0, 1]$. If $\mu_{ij} < 0$, [11] will update μ_{ij} as follows

$$\text{if } \|x_i - v_j\|^2 \leq \eta_j \|v_j - \overline{x}\|^2, \text{ then } \mu_{ij} = 1, \text{ and if } j' \neq j, \text{ then } \mu_{ij'} = 0. \tag{6}$$

Parameter η_j is decided by (7), the performance of FCS clustering depends on a suitable parameter β.

$$\eta_j = \frac{\frac{\beta}{4} \min_{j' \neq j} \|v_j - v_{j'}\|^2}{\max_k \|v_k - \overline{x}\|}, \quad 0 \leq \beta \leq 1. \tag{7}$$

The initial clustering center of FCS clustering is calculated as follows

$$v_j = \frac{1}{n_j} \sum_{l=1}^{n_j} x_l, \quad j = 1, 2, \ldots, c, \tag{8}$$

where n_j is the amount of samples of j class and x_l is one of the samples of j class ($l = 1, \ldots, n_j$).

3. **Fuzzy Semi-Supervised Support Vector Machine Based on FCS Clustering Algorithm.** In machine learning, SVMs are supervised learning models with associated learning algorithms that analyze data used for classification. When a lot of data are not labeled, supervised learning is not possible. Therefore, a semi-supervised learning approach is required. How to use a small amount of labeled data and a large number of unlabeled data has become an important research problem, and then S^3VM was proposed. Lin and Wang proposed the fuzzy support vector machines (FSVM) by introducing the concept of fuzzy membership into SVM [15,16]. Every sample has its own fuzzy membership value that can measure the contribution of samples to the classifier. In order to reduce the influence of noise points to the classifier, a small fuzzy membership value needs to be given to each noise point. Labeled by fuzzy membership acquired by FCS, a large number of newly labeled data have fuzzy membership values. FS^3VM-FCS is proposed in this paper.

Training dataset includes two parts: labeled dataset $(x_1, y_1), \ldots, (x_L, y_L) \in R^n \times \{+1, -1\}$ and unlabeled dataset $X_U = (x_{L+1}, \ldots, x_{L+U}) \in R^n$. FS^3VM-FCS can be described as follows

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 + C\left[\sum_{l=1}^{L} \xi_l + \sum_{u=L+1}^{L+U} \mu_u \eta_u\right] \\ \text{s.t.} \quad & y_l(\omega \cdot \phi(x_l) + b) + \xi_l \geq 1, \quad l = 1, 2, \ldots, L \\ & y_u(\omega \cdot \phi(x_u) + b) + \eta_u \geq 1, \quad u = L+1, L+2, \ldots, L+U \\ & \xi_l \geq 0, \quad l = 1, 2, \ldots, L \\ & \eta_u \geq 0, \quad u = L+1, L+2, \ldots, L+U, \end{aligned} \tag{9}$$

where penalty parameter C is constant and $C > 0$, ξ_l is the slack parameter and $\mu_u \eta_u$ is the slack parameter with weight. x_u will not be very important when μ_u is very small, so

its misclassification has a little influence on the classifier. y_u and μ_u can be obtained by the following equations

$$y_i = \arg \max_{j=1,\ldots,c} \mu_{ij}, \quad \forall i \in (L+1, \ldots, L+U), \tag{10}$$

$$\mu_i = \max_{j=1,\ldots,c} \mu_{ij}, \quad \forall i \in (L+1, \ldots, L+U). \tag{11}$$

We apply Lagrange multiplier and kernel function to solving the problem (9). Thus, the quadratic optimization problem (9) can be transformed as follows

$$\begin{aligned} \min \quad & L = -\sum_{l=1}^{L} \alpha_l - \sum_{u=L+1}^{L+U} \beta_u + \frac{1}{2} \left(\sum_{l=1}^{L} \sum_{k=1}^{L} k(x_l, x_k) y_l y_k \alpha_l \alpha_k \right. \\ & \left. + \sum_{l=1}^{L} \sum_{u=L+1}^{L+U} k(x_l, x_u) y_l y_u \alpha_l \beta_u + \sum_{u=L+1}^{L+U} \sum_{h=L+1}^{L+U} k(x_u, x_h) y_u y_h \beta_u \beta_h \right) \\ \text{s.t.} \quad & 0 \le \alpha_l \le C, \ l = 1, 2, \ldots, L \\ & 0 \le \beta_u \le \mu_u C, \ u = L+1, L+2, \ldots, L+U \\ & \sum_{l=1}^{L} y_l \alpha_l + \sum_{u=L+1}^{L+U} y_u \beta_u = 0. \end{aligned} \tag{12}$$

The steps of FS^{3}VM-FCS are described as follows:

Step 1: Use the FCS clustering algorithm to calculate fuzzy membership matrix U, then label the unlabeled data by (10) and calculate fuzzy membership values of newly labeled data by (11). FCS clustering algorithm is described as follows:

Input: Unlabeled dataset X_U, initial clustering center $v = (v_1, \ldots v_c)$, the number of clustering c, parameters β and $\varepsilon > 0$.
Output: Fuzzy membership matrix U.
1st: Calculate fuzzy membership matrix by (5);
2nd: Revise fuzzy membership matrix U by (6) and (7);
3rd: Update clustering center matrix v' by (4);
4th: If $\|v' - v\| < \varepsilon$, stop. Else return to the 1st and $v = v'$;

Step 2: Use sample dataset $\{(x_1, y_1), \ldots, (x_L, y_L)\}$, $\{(x_{L+1}, y_{L+1}), \ldots, (x_{L+U}, y_{L+U})\}$ and $\mu = (\mu_{L+1}, \ldots, \mu_{L+U})$ to train a classifier;
Step 3: Use test dataset to test accuracy of the classifier.

4. **Experiments.** The experiment is conducted on one artificial dataset G50c and two publicly available benchmark datasets from the UCI Repository [17] Diabetes and Thyroid. All experiments are run on a personal computer with 2GB memory. The codes of models are written in Matlab language. We compare FS^{3}VM-FCS with FS^{3}VM based on FCM clustering algorithm (FS^{3}VM-FCM).

In the experiment, three datasets have two classes respectively. Table 1 gives features of three world datasets. The numerical results of FS^{3}VM-FCS and FS^{3}VM-FCM for the three datasets are also included in Table 2. The experimental results in Table 2 show that the proposed FS^{3}VM-FCS has higher testing accuracy than FS^{3}VM-FCM.

Table 1. Features descriptions of three datasets

Dataset	Number of the labeled data	Number of the unlabeled data	Number of the test data	Class
G50c	50	300	200	2
Diabetes	50	418	300	2
Thyroid	20	120	75	2

TABLE 2. The performance comparisons of two algorithms

Dataset	Algorithm	Test accuracy
G50c	FS^{3}VM-FCM	95.00%
	FS^{3}VM-FCS	95.50%
Diabetes	FS^{3}VM-FCM	71.34%
	FS^{3}VM-FCS	72.34%
Thyroid	FS^{3}VM-FCM	84.00%
	FS^{3}VM-FCS	85.34%

5. **Conclusions.** This paper proposes FS^{3}VM-FCS. FCS clustering algorithm attempts to minimize the fuzzy within-cluster variation and maximize the fuzzy between-cluster variation. A large number of unlabeled data are labeled by fuzzy membership obtained by FCS clustering algorithm. Then FS^{3}VM-FCS trains the classifier by utilizing a small number of labeled data and a large number of newly labeled data. Compared with FS^{3}VM-FCM, FS^{3}VM-FCS has much better classification performance.

In this paper, we utilize FCS clustering algorithm to assign a fuzzy membership value to each unlabeled training sample of S^{3}VM. In the future, we will pursue to select a proper fuzzy membership function to the problem of noise points of S^{3}VM.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grants (61100165, 61100231, 61472307, 61562001), Natural Science Basic Research Plan in Shaanxi Province of China (Program 2014JM8313, 2016JM6004), Academics Propulsion Technology Transfer Projects of the Xi'an Science and Technology Bureau (CXY1516(6)), New Star Team of Xi'an University of Posts and Telecommunications and Open Fund of State Key Laboratory for Strength and Vibration of Mechanical Structures (SV2015-KF-04).

REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
[2] S. Lin and B. Xue, The application of improved SVM for data analysis in tourism economy, *The 7th International Conference on Intelligent Computation Technology and Automation*, Changsha, China, pp.769-772, 2014.
[3] Y. Leng, C. Sun, X. Xu et al., Employing unlabeled data to improve the classification performance of SVM and its application in audio event classification, *Knowledge-Based Systems*, vol.98, pp.117-129, 2016.
[4] Y. Dai, J. Tian, H. Rong and T. Zhao, Hybrid safety analysis method based on SVM and RST: An application to carrier landing of aircraft, *Safety Science*, vol.80, pp.56-65, 2015.
[5] K. P. Bennett and A. Demiriz, Semi-supervised support vector machines, *Proc. of Neural Information Processing Systems*, Denver, vol.12, 1998.
[6] Y. Q. Li, H. Q. Li, C. T. Guan and Z. Y. Chin, A self-training semi-supervised support vector machines algorithm and its applications in brain computer interface, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol.1, pp.I385-I388, 2007.
[7] L. Peng and X. Yang, Explore semi-supervised support vector machine algorithm for the application of physical education effect evaluation, *International Journal of Advancements in Computing Technology*, vol.4, no.9, pp.266-271, 2012.
[8] A. Pozdnoukhov, R. S. Purves and M. Kanevski, Semi-supervised support vector machine for natural hazards forecasting. Case study: Snow avalanches, *Proc. of iEMSs 4th Biennial Meeting – Int. Congress on Environmental Modelling and Software: Integrating Sciences and Information Technology for Environmental Assessment and Decision Making*, vol.1, pp.328-335, 2008.
[9] A. Nigam, B. Nigam, C. Bhaisare and N. Arya, Classifying the bugs using multi-class semi supervised support vector machine, *Proc. of International Conference on Pattern Recognition, Informatics and Medical Engineering*, Salem, Tamilnadu, pp.393-397, 2012.

[10] H. Benbrahim and M. Bramer, A fuzzy semi-supervised support vector machines approach to hypertext categorization, *The International Federation for Information Processing*, vol.276, pp.97-106, 2008.
[11] K. Wu, J. Yu and M. Yang, A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality test, *Pattern Recognition Letters*, vol.26, no.5, pp.639-652, 2005.
[12] E. W. Forgy, Cluster analyses of multivariate data: Efficiency versus interpretability of classifications, *Biometrics*, vol.21, pp.768-769, 1965.
[13] J. C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics*, vol.3, pp.32-37, 1973.
[14] X. L. Xie and G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.13, no.8, pp.841-847, 1991.
[15] C. F. Lin and S. D. Wang, Fuzzy support vector machines, *IEEE Trans. Neural Networks*, vol.13, no.2, pp.464-471, 2002.
[16] C. F. Lin and S. D. Wang, *Fuzzy Support Vector Machines with Automatic Membership Setting, Studies in Fuzziness and Soft Computing*, Springer Berlin Heidelberg, 2005.
[17] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, http://www.ics.uci.edu/_mlearn/MLRepository.html, 1998.