FUZZY UNIFIED SUPPORT VECTOR MACHINE FOR BRAIN-COMPUTER INTERFACE

LE MINH TRUNG*, PHUONG DUONG**,
MINH NGUYEN**, THIEN PHAM**, and MI DINH**

ABSTRACT

Support Vector Machine (SVM) and Support Vector Data Description (SVDD) are two well-known kernel-based methods for classification problems. In this paper, we introduce Fuzzy Unified Support Vector Machine (FUSVM) which is a fuzzy model, and unites SVM and SVDD. The experiment on the data set III of BCI competition II shows performance of the proposed method.

Keywords: Support Vector Machine, Support Vector Data Description, Kernel Method, Fuzzy Classifier, Brain-Computer Interface.

TÓM TẮT

Mô hình Fuzzy của Support Vector Machine Hợp nhất

Support Vector Machine (SVM) và Support Vector Data Description (SVDD) là hai phương pháp kernel nổi tiếng dùng cho bài toán phân loại. Trong bài báo này, chúng tôi đề xuất Fuzzy Unified Support Vector Machine (FUSVM), một mô hình fuzzy hợp nhất SVM và SVDD. Thực nghiệm được tiến hành trên data set III của cuộc thi BCI II chứng tỏ tính ưu việt của phương pháp đề xuất.

Từ khóa: Support Vector Machine, Support Vector Data Description, phương pháp Kernel, máy phân loại Fuzzy, Brain-Computer Interface.

1. Introduction

Given a binary training data set including both positive and negative data points $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ where labels $y_1, y_2, \ldots, y_n \in \{-1,1\}$. Support Vector Machine (SVM) [2] aims at constructing optimal hyperplane so that margin, distance from the closest point to the optimal hyperplane, is maximized. SVM is usually used to cope with balanced data sets. On the other hand, Support Vector Data Description (SVDD) [20] targets to building up optimal hypersphere which can include all data points of positive class and exclude all data points of negative class. SVDD is often utilized to deal with imbalanced data sets. However, in practice sometimes it is hard to assert whether a given binary data set is balanced or imbalanced. Hence, it is really necessary to have a model that can learn both the balanced and imbalanced data sets.

* Dr., Faculty of Information Technology, HCMC University of Education
** Students, Faculty of Information Technology, HCMC University of Education

SVM and SVDD are two state-of-the-art classifiers for two-class and one-class classification problems. However, the original SVM and SVDD regard all data points identically regardless of their natures and their contributions to decision boundary construction. To reduce the impacts of the noises and the less confident data, the framework for Fuzzy Support Vector Machine was proposed in [16]. Each data point was associated with a membership to specify its extent in the construction of decision boundary. The model favored the data points with high membership and disfavored the less confident data points. The framework proposed in [16] was extended by other researchers. Density-induced information was incorporated to data points as memberships to promote data points with high influence and reduce the effects of less-imposed data points [15]. Some authors linked to fuzzy theory and used the approach of fuzzy theory [10] to obtain fuzzy membership in two-class case [4], and [7].

Brain-Computer Interface (BCI) is an emerging research field attracting a lot of research effort from researchers around the world. Its aim is to build a new communication channel that allows a person to send commands to an electronic device using his/her brain activities [1]. The performance of a BCI system depends on data pre-processing, feature extraction and classification methods used to build the classifier in that BCI system. Currently, numerous pre-processing, feature extraction and classification methods have been proposed and explored for BCI systems. For data preprocessing and feature extraction, the following features have been applied: raw electroencephalograph (EEG) signals [9], band powers [8], power spectral density values [19], auto regressive parameters [5] and wavelet features [3]. For classification, perceptron and multi-layer perceptron [5], various SVMs [5], [11]and linear discriminant analysis [5] methods have been applied.

In this paper, we introduce the proposed model Fuzzy Unified Support Vector Machine (FUSVM) to answer two aforementioned questions: how to unify SVM and SVDD to have a model which can deal with all kinds of data set, how to employ fuzziness to reduce the effects of outliers and less-imposed data points. FUSVM is based on the previous works: Unified Support Vector Machine [12], [14] for uniting SVM and SVDD, and the methods for evaluating memberships of data points proposed in [10], [15], [17], [6]. The experiment which was established on BCI data sets shows the performance of the proposed model.

2. Unified Support Vector Machine (USVM)

Unified Support Vector Machine (USVM) [12],[14] is the model which unites SVM and SVDD. The curving degree parameter k governs the shape of decision boundary. It has been proven that when $k = 1$, USVM coincides SVM and when k approaches $+\infty$,USVM also approaches SVDD. Besides, USVM can provide the intermediate models by means of varying k . It is fair to claim that USVM can learn the real-world data sets better than SVM and SVDD.

$$\min_{a,b,\xi}\left(\|b-a\|^2 + C\sum_{i=1}^{n}\xi_i\right) \quad \mathbf{(1)}$$

subject to

$$\left(\|\phi(x_i)-a\|^2 - k\|\phi(x_i)-b\|^2\right)y_i \geq 1-\xi_i$$

$$\xi_i \geq \mathbf{0}, i = 1, \ldots, n \qquad (2)$$

where a, b are coordinates of two optimal points A, B located in feature space, and k is curving degree parameter.

3. Fuzzy Unified Support Vector Machine (FUSVM)

3.1. Formulation

To control the effects of data points, we associate each data point x_i with a membership λ_i. The optimization problem of FUSVM is as follows:

$$\min_{a,b,\xi}\left(\|b-a\|^2 + C\sum_{i=1}^{n}\lambda_i\xi_i\right) \quad \mathbf{(3)}$$

subject to

$$\left(\|\phi(x_i)-a\|^2 - k\|\phi(x_i)-b\|^2\right)y_i \geq 1-\xi_i$$

$$\xi_i \geq \mathbf{0}, i = 1, \ldots, n \qquad (4)$$

Since we are trying to minimize $\sum_{i=\mathbf{0}}^{n}\xi_i$, if λ_i is big then ξ_i, error at data point x_i, must be small and vice versa. Naturally, we can employ membership λ_i to control the importance degree of data point x_i.

3.2. Solution

To derive the new optimization problem, we refer to *Karush-Kuhn-Tucker (KKT)* theorem. The Lagrange function is of the following form:

$$L(a,b,\xi,\alpha,\beta) = \|b-a\|^2 + C\sum_{i=1}^{n}\lambda_i\xi_i - \sum_{i=1}^{n}\alpha_i[(\|\phi(x_i)-a\|^2 - k\|\phi(x_i)-b\|^2)y_i - 1 + \xi_i] - \sum_{i=1}^{n}\beta_i\xi_i \qquad (5)$$

Setting the derivatives to 0, we gain:

$$\frac{\delta L}{\delta a} = 0 \;\rightarrow a - b = \sum_{i=1}^{n}\alpha_i y_i\left(a-\phi(x_i)\right) \qquad (6)$$

$$\frac{\delta L}{\delta b} = 0 \;\rightarrow b - a = k\sum_{i=1}^{n}\alpha_i y_i\left(\phi(x_i)-b\right) \qquad (7)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i + \beta_i = \lambda_i C, i = 1, \cdots, n \tag{8}$$

Adding the product of Equation (6) to k and Equation (7), we obtain:

$$(k-1)(a-b) = k\sum_{i=1}^{n} \alpha_i y_i (a-b) \tag{9}$$

Since $a - b \neq 0$, from Equation (9), we have:

$$\sum_{i=1}^{n} \alpha_i y_i = \frac{k-1}{k} \tag{10}$$

Substituting Equation (10) to Equation (6), we have:

$$a - kb = -k\sum_{i=1}^{n} y_i \alpha_i \phi(x_i) \tag{11}$$

Substituting back to Equation (5), we achieve the dual form:

$$L = \|b-a\|^2 + C\sum_{i=1}^{n} \lambda_i \xi_i - \sum_{i=1}^{n} \alpha_i [(\|\phi(x_i) - a\|^2 - k\|\phi(x_i) - b\|^2) y_i - 1 + \xi_i]$$
$$- \sum_{i=1}^{n} \beta_i \xi_i = \|b-a\|^2 - \frac{2}{k}\|a-kb\|^2 - \frac{k}{k-1}\left(\|a\|^2 - k\|b\|^2\right)$$
$$+ \sum_{i=1}^{n} \alpha_i - (k-1)\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_i)$$
$$= -\frac{1}{k}\|a-kb\|^2 + \sum_{i=1}^{n} \alpha_i - (k-1)\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_i)$$
$$= -k\left\|\sum_{i=1}^{n} y_i \alpha_i \phi(x_i)\right\|^2 + \sum_{i=1}^{n}\left((k-1)\mathbf{y}_i K(\mathbf{x}_i, \mathbf{x}_j) + 1\right)\boldsymbol{\alpha}_i \tag{12}$$

We end up with the following optimization problem:

$$\min_{\alpha}\left(k\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{y}_i \mathbf{y}_i K(\mathbf{x}_i, \mathbf{x}_j)\boldsymbol{\alpha}_i \boldsymbol{\alpha}_j - \sum_{i=1}^{n}\left((k-1)\mathbf{y}_i K(\mathbf{x}_i, \mathbf{x}_j) + 1\right)\boldsymbol{\alpha}_i\right) \tag{13}$$

subject to

$$\sum_{i=1}^{n} \mathbf{y}_i \boldsymbol{\alpha}_i = \frac{k-1}{k}; 0 \leq \boldsymbol{\alpha}_i \leq \boldsymbol{\lambda}_i C, i = 1, \ldots, n \tag{14}$$

For classifying an unknown data point x , the following decision function is used:

$$f(x) = sign(\|\phi(x) - a\|^2 - k\|\phi(x) - b\|^2)$$

$$= sign\left((1-k)K(x,x) + 2k\sum_{j=1}^{n}\alpha_j \mathbf{y}_j K(x,x_j) + d \right) \qquad (15)$$

where $d = \|a\|^2 - k\|b\|^2$

In practice, we calculate $d = \|a\|^2 - k\|b\|^2$ as follows:

$$\|a\|^2 - k\|b\|^2 = \frac{1}{|I|}\sum_{i\in I}\left(y_i + (k-1)K(x_i,x_i) - 2k\sum_{j=1}^{n}\alpha_j \mathbf{y}_j K(x_i,x_j) \right) \qquad (16)$$

where I is the set of indices of all support vectors, i.e., $I = \{i: 0 < \alpha_i < \lambda_i C; 1 \le i \le n\}$.

Similar to USVM, the decision boundary of FUSVM is either hyperplane or hypersphere. It coincides Fuzzy Support Vector Machine (FSVM) [16] when $k = 1$ and approaches Fuzzy Support Vector Data Description [13] when k approaches $+\infty$.

3.3. *Evaluation of Memberships*

To evaluate memberships associated with data points, we uses 4 methods which were proposed in [13], [15], and [10].

METHOD 1 FOR EVALUATING MEMBERSHIPS [13]

Do clustering algorithm Fuzzy C-Means ininput space.

Discover clusters that contain both positive and negative data. Denote this set by MIXEDCLUS.

For each clus $\in$ MIXEDCLUS and p $\in$ clus

Set $\lambda_p = 1$

EndFor

For each *clus* $\notin$ MIXEDCLUS and p $\in$ clus

Find out the cluster whose center is closest to p.

Denote this closest cluster by clus[p].

Set $\lambda_p = Membership[p]$ where Membership[p] is membership of p with respect to cluster clus[p].

EndFor

METHOD 2 FOR EVALUATING MEMBERSHIPS [15]

For $i = 1$ to n do

Set $\lambda_i = e^{\omega \times \frac{K}{d(x_i, x_i^k)}}$

EndFor

where $d(x_i, x_i^k)$ is distance from x_i to its k-nearest-neighbor, $K = \frac{1}{t}\sum_{i=1}^{t} d(x_i, x_i^k)$ with t is number of data points in a target class, and ω is a weighting factor.

METHOD 3 FOR EVALUATING MEMBERSHIPS [15]

For $i = 1$ **to** n **do**

Set $\lambda_i = e^{\omega \times \frac{Par(x_i)}{J}}$

EndFor

where $Par(x_i) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\sqrt{(2\pi)^d s}} e^{-\frac{1}{2\pi} d(x_i, x_j)}$ with d is dimensionality of input space and s is smoothing parameter of Parzen-window density, and $J = \frac{1}{n}\sum_{i=1}^{n} Par(x_i)$.

METHOD 4 FOR EVALUATING MEMBERSHIPS [10]

For $i = 1$ **to** n **do**

If $y_i = 1$ **then**

Set $\lambda_i = 0.5 + \frac{e^{\frac{f[d_{-1}(\phi(x_i)) - d_1(\phi(x_i))]}{d}} - e^{-f}}{2(e^f - e^{-f})}$

Else

Set $\lambda_i = 0.5 + \frac{e^{\frac{f[d_1(\phi(x_i)) - d_{-1}(\phi(x_i))]}{d}} - e^{-f}}{2(e^f - e^{-f})}$

EndIf

EndFor

where $d_1(\phi(x_i))$ is distance between $\phi(x_i)$ and mean of positive class in feature space, and $d_{-1}(\phi(x_i))$ is distance between $\phi(x_i)$ and mean of negative class in feature space, and f is a constant which controls the rate at which memberships decrease toward 0.5.

4. Experiment

4.1. Data set

The chosen data set was the well-known data set III provided by Department of Medical Informatics, Institute of Biomedical Engineering, and Graz University of Technology for motor imagery classification problem in BCI Competition II [18]. In data collection stage, a female normal subject was asked to sit in a relaxing chair with armrests and tried to control a feedback bar by means of imagery left or right hand movements. The sequences of left or right orders are random. The experiment consisted of 7 runs with 40 trials in each run. There were 280 trials in total and each of them lasted 9seconds of which the first 3 seconds are used for preparation. Collected data was equally divided into two sets for training and testing. The data was recorded in three EEG channels which were C3, Cz and C4, sampled at 128 Hz, and filtered between 0.5 Hz and 30Hz. Most of current algorithms only applied to the channels C3 andC4, and ignored the channel Cz. They argued that from brain theory, signals from channel Cz provide very little meaning to motor imagery problem. We truncated the first 3 seconds of each trial and used the rest for further processing. All trials are pre-processed by subtracting the ensemble mean of all trials. For each trial we extracted Combined Short-Window Bivariate Autoregressive Feature (CSWBVAR) parameters with window size of 512 data points corresponding to 1s-segment and moving window step of 75% of the window size. We did not try experiments with segment's size greater than1s due to keeping signal approximately stationary and being comfortable with nature of brain signal.

4.2. Parameter Settings

The popular RBF Kernel $K(x,x') = e^{-\gamma\|x-x'\|^2}$ was applied whereas the parameter γ was varied in grid $\{2^i : i = 2j+1, j = -8, \dots, 2\}$. The trade-off parameter C was selected in grid $\{2^i : i = 2j+1, j = -8, \dots, 2\}$. The curving degree k was searched in grid $\{0.6 + 0.1i : i = 0, \dots, 9\}$. The number of clusters for FUSVM 1 was set to 15. The parameters w; s; f and k were set to 0.01; 0.4; 0.1 and 0.4 respectively. By the way, five folds cross-validation was employed in experiment.

We applied 4 methods for evaluating memberships to framework of FUSVM to form FUSVM1, FUSVM2, FUSVM3, and FUSVM4. The experimental results as seen in Figure 1 show that variations of FUSVM outperform others and USVM surpasses SVM in terms of classification accuracy.

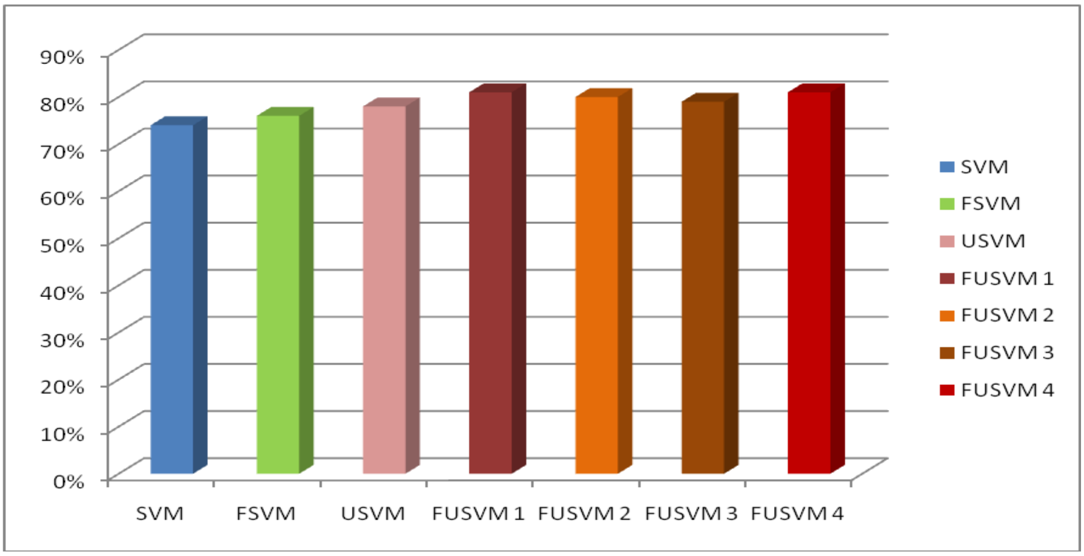

Figure 1. *The experimental results on the data set.*

5. Conclusion

We propose Fuzzy Unified Support Vector Machine (FUSVM) for two main purposes: unite Fuzzy Support Vector Machine (FSVM) and Fuzzy Support Vector Data Description (FSVDD), and apply fuzziness to reduce effects of outliers and less-imposed data points. It has been proven that when curving degree parameter $k = 1$, FUSVM is FSVM and when k approaches $+\infty$, FUSVM also approaches FSVDD. Besides, by varying k in other domains FUSVM provides other intermediate models which are different to FSVM and FSVDD. Therefore, it stays on reason to claim that FUSVM is able to learn the real-world data sets better than FSVM and FSVDD. In the paper, four popular methods for evaluating memberships were investigated. Experiment established on brain data set shows that FUSVM outperforms SVM, FSVM.

REFERENCES

1. Babiloni F., Cichocki A., and Gao S. (2007), "Brain computer interfaces: Towards practical implementations and potential applications", *Intelligence Neuroscience*, pp. 1687–5265.
2. Boser B., Guyon I., and Vapnik V. (1992), "A training algorithm for optimal margin classifiers", *In Proceedings of the5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152.
3. Bostanov V. (2004), "Bci competition 2003 data sets ib and iib: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-valuescalogram", *IEEE Trans. Biomed. Eng*, 51, pp. 1057–1061.
4. Chen J.-H. and Chen C.-S. (2002), "Fuzzy kernel perceptron", *IEEE Transaction on Neural Networks*, 13(6), pp. 1464–1473.
5. Garrett D., Peterson D. A., Anderson C. W., and Thaut M. H. (2003), "Comparison of linear, nonlinear, and feature selection methods for eeg signal classification", *IEEE Trans. Neural System and Rehabilitation Eng.*, 11, pp. 141–144.
6. Hoang T., Nguyen P., Le T., Tran D., and Sharma D. (2010), "Enhancing performance of svm-based brain computer interface systems", *Special ICONIP issue of the Australian Journal of Intelligent Information Systems*.
7. Jayadeva R., Khemchandani R., and Chandrab S. (2004), "Fast androbust learning through fuzzy linear proximal support vector machines", *Neurocomputing*, 61, pp. 401–411.
8. Kalcher J., Flotzinger D., Neuper C., Glly S., and Pfurtscheller G. (1996), "Graz brain-computer interface ii: towards communication between humans and computers based on online classification of three different eeg patterns", *Med. Bio. Eng. Computing*, 34(5), pp. 382–388.
9. Kaper M., Meinicke P., Grossekathoefer U., Lingner T., and Ritter H. (2004), "BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm", *IEEE Transactions on Biomedical Engineering*, 51(6), pp. 1073 –1076.
10. Keller J. M. and Hunt D. J (1985), "Incorporating fuzzy membership functions into the perceptron algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 693 –699.
11. Kuncheva L. I. and Rodriguez J. J. (2010), "Classifier ensembles for fmri data analysis: an experiment, magnetic resonance imaging", *In ISSN 0730-725X*.
12. Le T., Tran D., Hoang T., Ma W., and Sharma D. (2011), "Generalised support vector machine for brain-computer interface", *In ICONIP 2011*, volume 7062, pp. 537–544.
13. Le T., Tran D., Ma W., and Sharma D. (2010), "A new fuzzy membership computation method for fuzzy support vector machines", *In 2010 Third International Conference on Communications and Electronics (ICCE)*.

14. Le T., Tran D., Ma W., and Sharma D. (2012), "A unified model for support vector machine and support vector data description", *In International Joint Conference on Neural Networks (IJCNN 2012)*.
15. Lee K.-Y., Kim D.-W., Lee K. H., and Lee D. (2007), "Density induced support vector data description", *IEEE Transactions on Neural Networks*, 18(1), pp. 284 – 289.
16. Lin C.-F. and Wang S.-D. (2002), "Fuzzy support vector machines", *IEEE Transactions on Neural Networks*,13(2), pp. 464 – 471.
17. Nguyen P., Le T., Tran D., Huang X., and Sharma D. (2010), "Fuzzy support vector machines for age and gender classification", *In INTERSPEECH*, pp. 2806–2809.
18. Pfurtscheller G. and Schlogl A., "Data set III in BCIcompetition II, http://www.bbci.de/competition/ii/".
19. Solhjoo S. and Moradi M. H. (2004), "Mental task recognition: A comparison between some of classification methods", *In BIOSIGNAL 2004 International EURASIP Conference*.
20. Tax D. M. J. and Duin R. P. W. (2004), "Support vector data description", *Journal of Machine Learning Research*, 54(1), pp. 45–66.

(*Received: 13/10/2013; Revised: 18/11/2013; Accepted: 13/12/2013)*