STUDIA INFORMATICA 2009
Volume 30 Number 2A (83)

Jerzy MARTYNA
Uniwersytet Jagielloński, Instytut Informatyki

FUZZY SUPPORT VECTOR MACHINES BASED ON DENSITY ESTIMATION WITH GAUSSIAN MIXTURE FOR MULTICLASS PROBLEMS

Summary. In this paper, we introduce new Fuzzy Support Vector Machines (FSVMs) for a multiclass classification. The suggested Fuzzy Support Vector Machines include the data distribution with the density estimated in a set of functions defined as Gaussian mixture. The proposed method gives more appropriate boundaries than the classical FSVM method. We demonstrate some examples which confirm our approach.

Keywords: Fuzzy Support Vector Machine, density, multiclass problems, membership functions

ROZMYTA METODA SVM OPARTA NA ESTYMACJI GĘSTOŚCI Z MIESZANKĄ GAUSSOWSKĄ DLA ROZWIĄZYWANIA PROBLEMÓW WIELOKLASOWYCH

Streszczenie. W pracy przedstawiono matematyczny model, jakim jest Fuzzy Support Vector Machine (FSVM), czyli rozmyta maszyna wektorów podpierających. Wprowadzono w nim estymację gęstości opartą na zbiorze funkcji definiowanych jako mieszanka funkcji gaussowskich. Zaproponowana metoda dostarcza lepszych ograniczeń niż dotychczas stosowany model FSVM. Demonstrujemy kilka przykładów, które potwierdzają opisywane podejście.

Słowa kluczowe: rozmyta maszyna wektorów podpierających, gęstość, problemy wieloklasowe, funkcje przynależności

1. Introduction

Support Vector Machines (SVMs) [10, 8] have been used in many applications for classification and regression [9, 4, 6, 7]. The SVM method is mainly used for classification of two classes. It is caused by the existence of some unclassifiable regions which appear in the multiclass problems.

In order to avoid this problem the Fuzzy Support Vector Machines (FSVMs) were proposed [5, 1, 3]. In these papers, the fuzzy memberships are assigned according to the distance between the patterns. Nevertheless, so treated FSVMs do not take the distribution of the data. Therefore, given FSVMs cannot well adjust the decision boundaries for regions of data sets.

The main goal of this paper is to introduce a new method of a multiclass classification in which an unclassifiable region can be resolved. In the proposed FSVMs decision boundaries are used which consider not only the optimal class separating the hyperplanes in the SVM, but also the density of the distribution of the patterns. As one of the best approximations of the density estimation we used a set of functions defined in [0,1) as Gaussian mixtures. As a result, the multiclass problem can be better solved for data which are generally distributed.

The structure of the paper is as follows: section 2 explains the FSVMs method. The next section presents our solution based on an approximation of a density with Gaussian mixtures. In section 4, we give our proposed algorithm. In section 5, we present some numeric results of focusing on the advantages and disadvantages of our approach. Finally, in section 6, we give our conclusion and propose some future research.

2. Fuzzy Support Vector Machines

The Fuzzy Support Vector Machines were introduced by T. Inoue, S. Abe, T. Daisuke in the papers [5, 1, 3].

Let $[x_i, y_i]_{i=1}^{n}$ be training data where $x_i \in \Re^n$ is the input and $y_i \in \{-1,1\}$ is the output. The optimal separating hyperplane defined as $D(x) = w \cdot x + b$ is the decision function. It can be obtained by solving the following problem:

minimize:

$$\frac{1}{2}\|w\|^2 \tag{1}$$

subject to $y_i (w^T x_i + b) \geq 1,\ i = 1, \dots, n$.

The above Eq. (1) can be formulated in a simple manner, namely minimize:

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} y_i y_j \alpha_i \alpha_j x_j x_i \quad (2)$$

subject to $\sum_{i=1}^{n} y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots n,$

where α is a Lagrange multiplier. The optimal weight vector w^* and bias b^* can be obtained as follows:

$$\begin{cases} w^* = \sum_{i=1}^{n} y_i \alpha_i x_i \\ b^* = -\frac{1}{2}[\max_{y_i=-1}(w^* x_i) + \min_{y_i=1}(w^* x_i)] \end{cases} \quad (3)$$

The decision function $D(x)$ can be calculated from the above results, namely when $D(x) > 0$, pattern x is classified as class 1. Otherwise, it is classified as class 2.

The multiclass problem is achieved by defining the decision function for the pair class i and j as follows:

$$D_{ij}(x) = w_{ij}^T x + b_{ij} \quad (4)$$

where $D_{ij}(x) = -D_{ji}(x)$. For training data x_i, $i = 1, \ldots, n$ we have

$$D_i = \sum_{i=1, j \neq i}^{k} sign(D_{ij}(x)), \quad (5)$$

where $sign(.) = 1$ for $(.) > 0$ and zero in otherwise. The value of x is categorized by

$$\arg \max_{i=1,\ldots,k} D_i(x) \quad (6)$$

3. Fuzzy Support Vector Machines based on the density with Gaussian

Mixtures for multiclass problems In this section, we present Fuzzy Support Vector Machines based on the density with Gaussian mixtures for multiclass problems.

We resolve the problem of multiclass regions in the above presented FSVMs classification with the use of an estimator:

$$\widehat{f}_M(x) = \frac{1}{M} \sum_{i=1}^{M} K_M(x, x_i) \quad (7)$$

where $x_1, x_2 \ldots, x_M$ are the empirical data obtained from the observation of a n-dimensional random variable x with the probability density function $\widehat{f}$, K_M is the kernel function. As the kernel function we can use the so called Parzen kernel [10] given as follows:

$$K_M(x,u) = h_M^{-1} K(\frac{x-u}{h_M}) \tag{8}$$

where h_M is a function of the length of training data

$$\lim_{M\to\infty} h_M = 0 \quad \text{and} \quad \lim_{M\to\infty} Mh_M^h = \infty \tag{9}$$

We can provide [2] that

$$E[\widehat{f}_M(x) - f_M(x)]^2 \xrightarrow{M} 0 \tag{10}$$

in the absolutely continuous points of f. Function K can be given in the form

$$K(x) = \prod_{i=1}^{M} H(x^{(i)}) \tag{11}$$

Assuming that function H is a Gaussian form type, we have

$$\widehat{f}_M(x) = \frac{1}{(2\pi)^{\frac{n}{2}} nh_M^n} \sum_{i=1}^{M} \exp\left(-\frac{(x-x_i^T)(x-x_i)}{2h} \right) \tag{12}$$

With the help of the above-given equation, we can redefine the decision boundary $D_{ij}(x)$ in Eq. (5), namely

$$D_{ij}(x) = \gamma(w_{ij}^T + b_{ij}) + (1-\gamma)(\widehat{f}_{M_i}(x) - \widehat{f}_{M_{ji}}(x)) \tag{13}$$

where i and j denote the class pair, γ is a parameter that indicates the weight between the FSVMs and the approximation of density with Gaussian mixture.

The membership function of x for a given class i is defined with the help of the minimum operator, namely

$$m_i(x) = \min_{j=1,\ldots,n} m_{ij}(x) \tag{14}$$

4. Experimental Results

Based on the FSVM concept, this paper has made use of a script with this method included to Oracle Data Mining Software. The script are prepared using kernel-dependent formula such as the ones given for polynomial kernel with degree 2 or Gaussian mixture as kernel.

We use our method for classification in high dimensional data sets such as *glass*, and *cereals*. For instance, Fig. 1 shows a snapshot of the system for the classification tree obtained for the *cereals* data set with the help of FSVM with polynomial kernel. Table 1 shows the classification results for these data sets.

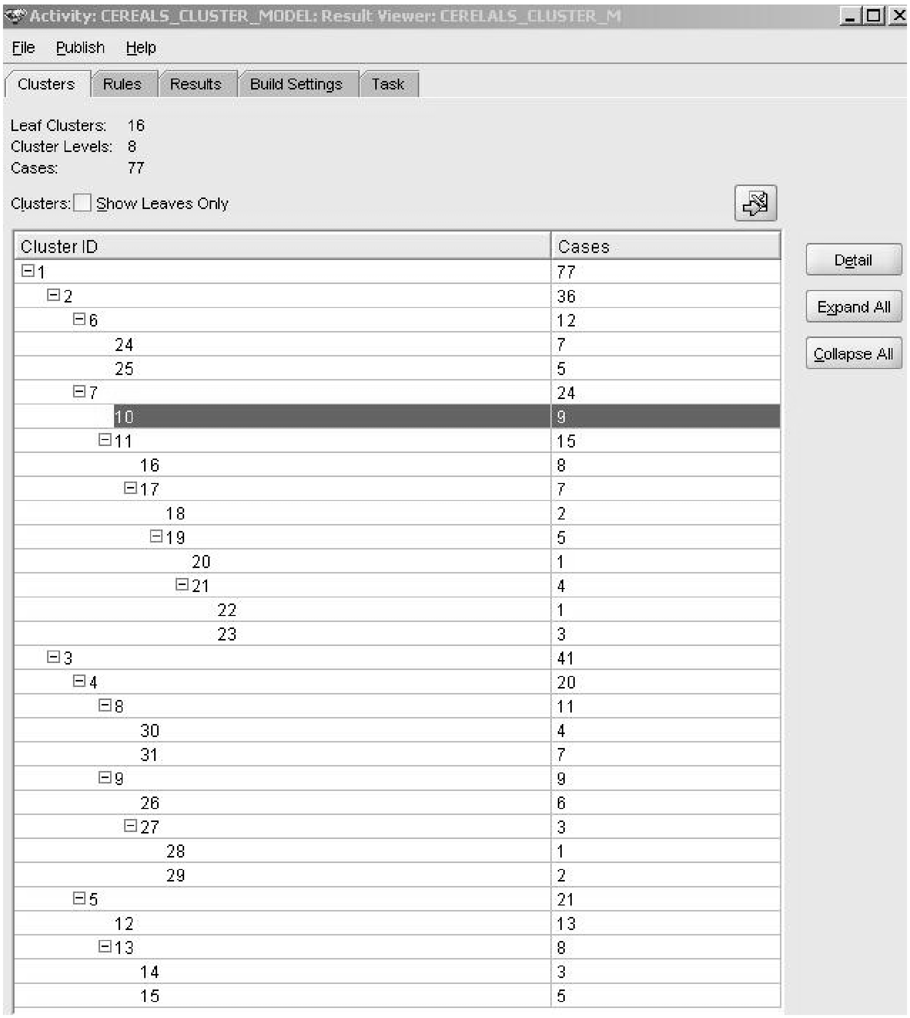

Fig. 1. The classification tree obtained for the *cereals* data set with the help of FSVM method
Rys. 1. Drzewo klasyfikacji uzyskane dla zbioru danych *cereals* przy użyciem metody FSVM

Table 1

Classification results for *glass* and *cereals* data sets

Data set	Class	Feature
glass	6	16
cereals	16	77

Recognition rate results for these data sets are shown in Fig. 2. We applied the first 60% of data records for training and we used the remaining 40% patterns for testing. Parameter γ was studied in relation to the number of classes. By varying parameter we have observed that for going to zero, we obtained the same result as the result obtained with the help of the FSVM with polynomial kernel. If parameter γ increases to value 1, we have obtain the most recognition of the pattern. These results indicate that the decision boundaries are thus not uniformly distributed as in the FSVMs method.

Figures 3 and 4 illustrate the ROC curve of two classifications for the *glass* data set that are obtained with the FSVM method based on the polynomial and on the Gaussian mixture as kernels, respectively. As we can see, the model of classification from Fig. 4 has better true positives than model of classification from Fig. 3.

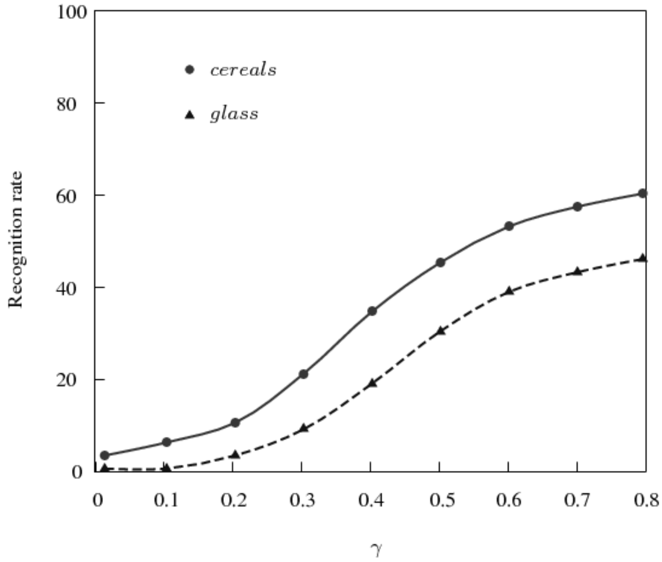

Fig. 2. Recognition rate for the *cereals* and *glass* data sets
Rys. 2. Intensywność rozpoznawania dla zbiorów danych *cereals* oraz *glas*

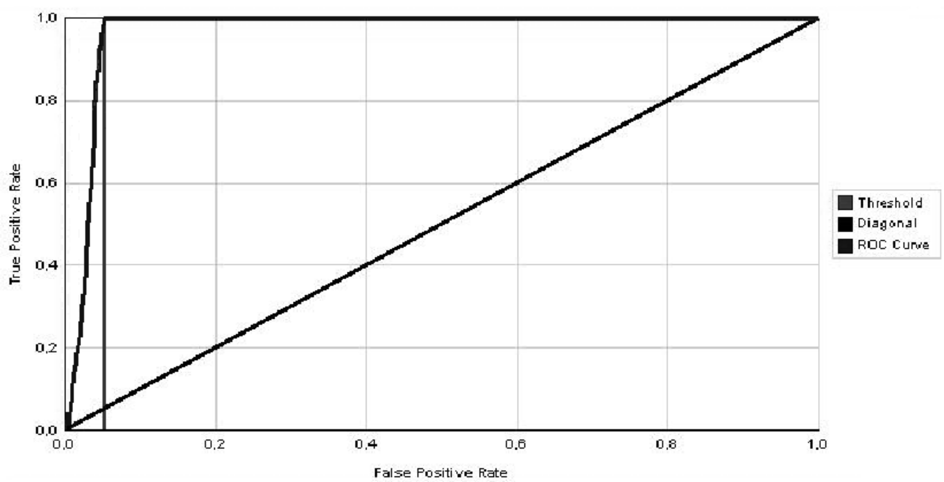

Fig. 3. The ROC curve of classification for the *glass* data set obtained with the FSVM method based on the polynomial kernel
Rys. 3. Krzywa ROC klasyfikacji zbioru danych *glass* uzyskana metodą FSVM z jądrem wielomianowym

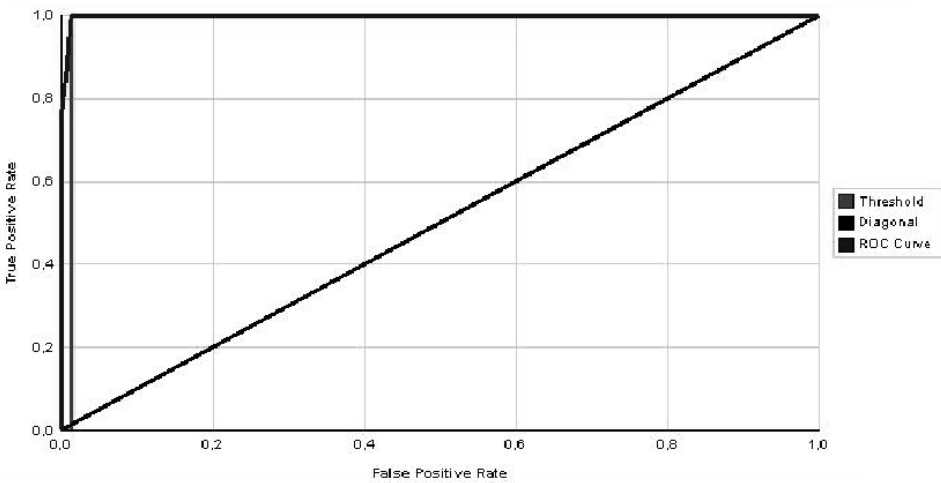

Fig. 4. The ROC curve of classification for the *glass* data set obtained with the FSVM method based on Gaussian functions mixture kernel

Rys. 4. Krzywa ROC klasyfikacji zbioru danych *glass* uzyskana metodą FSVM z jądrem w postaci mieszanki funkcji Gaussa

5. Conclusion

The paper proposed a new method for a multiclass classification with the help of Fuzzy Support Vector Machines based on Gaussian density functions. Our method can improve the problem of unclassifiable regions, which is typical in FSVM. Our FSVM method with density based on Gaussian functions allows us to overcome these difficulties. Moreover, selecting appropriate parameters can provide adequate accuracy of classification.

In future, we will investigate percentage errors in classification with the help of FSVM based on Gaussian density function and compare the obtained results with other fuzzy classification methods.

BIBLIOGRAPHY

1. Abe S., Inoue T.: Fuzzy Support Vector Machines for Multiclass Problems. Neural Networks, Vol. 16, 2003, p. 785÷792.
2. Cacoullos T.: Estimation of Multivariate Density. Ann. Inst. Statist. Math., Vol. 18, 1965, p. 179÷189.
3. Daisuke T., Shingo A.: Fuzzy Least Squares Vector Machines for Multiclass Problems. Neural Networks, Vol. 16, 2003, p. 785÷792.
4. Drucker H. D., Wu D., Vapnik V. N.: Support Vector for Spam Categorization. Trans. on Neural Networks, Vol. 10, No. 5, 1999, p. 1048÷1054.

5. Inoue T., Abe S.: Fuzzy Support Vector Machines for Pattern Classification. In: Proc. of the Int. Joint Conf. on Neural Networks, 2000, p. 1449÷1454.
6. Joachims T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In: Proc. of the European Conference on Machine Learning. Springer-Verlag, 1998, p. 137÷142.
7. Müller K. R., Smola A. J., Rätsch G., Schölkopf B., Kohlmorgen J., Vapnik N. V.: Predicting Time Series with Support Vector Machines. In: Proc. Int. Conf. on Artificial Neural Networks. ICANN-97, 1997, p. 999÷1004.
8. Nello C., John S.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.
9. Vapnik V. N., Golowich S. E., Smola A.: Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In: M. Mozer, M. I. Jordan, T. Petsche (eds.): Advances in Neural Information Processing System 9. Morgan Kaufmann, 1997, p. 281÷287.
10. Vapnik V. N.: Statistical Learning Theory. John Wiley and Sons, 1998.

Recenzent: Dr inż. Jerzy Respondek

Wpłynęło do Redakcji 5 marca 2009 r.

Omówienie

W pracy przedstawiono rozmytą maszynę wektorów wspierających (FSVM), w której zastosowano estymację gęstości opartą na zbiorze funkcji gaussowskich. Rozwiązanie to pozwala nie tylko na optymalną separację klas, jak to miało miejsce w dotychczas stosowanej metodzie FSVM, lecz także na lepszą aproksymację gęstości we wzorcach uczących. W rezultacie uzyskano dokładniejsze ograniczenia przy rozwiązywaniu problemów wieloklasowych.

Address

Jerzy MARTYNA: Uniwersytet Jagielloński, Instytut Informatyki, ul. Łojasiewicza 6,
30-348 Kraków, Polska, martyna@softlab.ii.uj.edu.pl.