Fuzzy Support Vector Machines for Age and Gender Classification

Phuoc Nguyen, Trung Le, Dat Tran, Xu Huang, Dharmendra Sharma

Faculty of Information Sciences and Engineering, University of Canberra, Australia

{phuoc.nguyen, trung.le, dat.tran, xu.huang, dharmendra.sharma@canberra.edu.au}

Abstract

Support vector machine (SVM) has been proven as a powerful tool for solving age and gender classification problems. However, SVM is sensitive to noise and outliers. In this paper we propose a new fuzzy SVM based on an assumption that training data points should not be treated equally to avoid the problem of sensitivity to noise and outliers. This can be achieved by assigning a fuzzy membership as a weight to each training data point. A method to calculate fuzzy memberships is also presented. Experiments performed on the aGender corpus for INTERSPEECH 2010 Paralinguistic Challenge show that the proposed fuzzy SVM can improve age and gender classification accuracy.

Index Terms: Fuzzy Support Vector Machine, Age Classification, Gender Classification, Paralinguistic Challenge

10.21437/Interspeech.2010-742

1. Introduction

Current speech recognition systems are highly speaker dependent. Parametric representations and their probability distributions suitable for a certain speaker may not be suitable for other speakers [1]. For example, the speech recognition performance for female speakers is almost worse than that for male speakers [2]. To improve the performance of speaker-independent speech recognition systems, gender classification and separate female and male speech models should be included [3]. Gender classification is also useful in speech emotion classification, face recognition, video summarization and human-robot interaction.

Age classification is concerned as a useful tool in different applications, for example, offering different advertisements to children and adults in the waiting queue and issuing different permission levels for different aging groups [4]. Statistical information on the age distribution of a caller group might also be an application [5].

Age and gender classification systems were developed only recently [6] [7]. It is hard to compare the quality of these systems as they vary considerably regarding the number and distribution of speaker age and speech material. Metze *et al.* [4] presented a detailed comparison of age and gender classification methods including Bayes method using sequential floating forward selection for feature selection, probabilistic Neural Networks, support vector machine (SVM) [8], K-nearest neighbour and Gaussian mixture model (GMM). Bocklet *et al.* presented age and gender recognition for telephone applications based on GMM supervectors and SVM [5]. The SVM systems outperformed all of these approaches for the same domain corpus.

Although SVM has been proven as a powerful tool for solving age and gender classification problems, SVM is very sensitive to noisy data or outliers. In our viewpoint, this sensitivity could be from the fact that SVM considers all data points with the same importance in classification problems. In order to reduce the sensitivity of less important data, we assume that training data points should not be treated equally. Therefore it is very important to distinguish the meaningful training data points from outliers or noisy data. This can be achieved by assigning a fuzzy membership value as a weight to each training data point [9]. Lee *et al.* [10] introduced a similarity measure function to compute those weights for pattern recognition. However, they had to assume that outliers should be somewhat separate from the normal data. Recently Hsu *et al.* [11] have investigated the effect of the trade-off parameter C to the model of conventional two-class SVM and introduced a triangular membership function to set higher grades to the data points in overlapping regions which contain data points of different classes. However these fuzzy SVMs are complicated and their fuzzy membership functions consist of free parameters whose optimum values are to be determined empirically. We have proposed a new approach to calculation of fuzzy memberships [12] and in this paper we apply this approach to age and gender classification problem. We propose a new method to compute fuzzy memberships, which can apply to fuzzy SVMs for one-class classification as well. Our viewpoint is that data points in overlapping regions are more important than others. We propose to use a fuzzy clustering technique to determine clusters in these regions. Data points in these clusters will have the highest fuzzy membership value. Fuzzy memberships for other data points are determined by their closest cluster accordingly, therefore their fuzzy membership values will be lower. Experiments performed on the aGender corpora for INTERSPEECH 2010 Paralinguistic Challenge [13] show that the proposed fuzzy SVMs can provide promising age and gender classification rates.

The paper is organised as follows. Section 2 reviews current SVM. Section 3 presents our proposed fuzzy SVMs and explains how to compute fuzzy memberships in Section 4. Section 5 presents experimental results, and finally, we conclude in Section 6.

2. Support Vector Machine

Let $\{x_i, y_i\}$, $x_i \in R^d$, $i = 1, \ldots, n$ be the training data set with two labels $y_i \in \{-1, +1\}$. SVM will find the optimal hyperplane [8]

$$f(x) = w^T \phi(x) + b \tag{1}$$

to separate the training data by solving the following optimisation problem:

$$\min_{w,b} \left(\frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \xi_i \right) \tag{2}$$

subject to

$$\begin{aligned} &y_i[w^T \phi(x_i) + b] \geq 1 - \xi_i \qquad i = 1, \ldots, n \\ &\xi_i \geq 0, \quad i = 1, \ldots, n \end{aligned} \tag{3}$$

where w is the normal vector of the hyperplane, C and b are real numbers, ξ_i, $i = 1, \ldots, n$ are non-negative slack variables, and $\phi(.)$ is a kernel function.

The optimisation problem will guarantee to maximise the hyperplane margin while minimising the cost of error. For an error to occur, the corresponding ξ_i must exceed unity, so $\sum_i \xi_i$ is an upperbound on the number of training errors. Hence an extra cost $C \sum_i \xi_i$ for errors is added to the function (1) where C is a parameter chosen by a user.

The decision function is $f(x) = sign(w^T \phi(x) + b)$. The unknown data point x belongs to positive class if $f(x) = +1$ or negative class if $f(x) = -1$.

3. Fuzzy Support Vector Machine

Fuzzy SVM is modeled as follows

$$\min_{w,b} \Big(\frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \lambda_i \xi_i \Big) \tag{4}$$

subject to

$$\begin{aligned} & y_i [w^T \phi(x_i) + b] \geq 1 - \xi_i \qquad i = 1, \ldots, n \\ & \xi_i \geq 0, \quad i = 1, \ldots, n \end{aligned} \tag{5}$$

where weights $\lambda_i \in [0, 1]$, $i = 1, \ldots, n$ are regarded as fuzzy memberships. This approach assumes that training data points should not be treated equally to avoid the problem of sensitivity to noise and outliers. The corresponding dual form is as follows

$$\min_{\alpha} \Big(\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{n} \alpha_i \Big) \tag{6}$$

subject to

$$\begin{aligned} & 0 \leq \alpha_i \leq \lambda_i C, \quad i = 1, \ldots, n \\ & \sum_{i=1}^{n} y_i \alpha_i = 0, \quad i = 1, \ldots, n \end{aligned} \tag{7}$$

The same decision function is used: $f(x) = sign(w^T \phi(x) + b)$. The unknown data point x belongs to positive class if $f(x) = +1$ or negative class if $f(x) = -1$.

4. Calculating Fuzzy Memberships

We propose a simple yet efficient method to determine fuzzy memberships. The positive and negative data points are normally overlapped and the task of fuzzy SVM is to construct a hyperplane in feature space to separate positive data from negative data. Hence we assume that the data points in the overlapping regions are important and they should have the highest fuzzy membership value. Other data points are less important and should have lower fuzzy membership values.

4.1. Fuzzy Clustering Membership

We use fuzzy clustering techniques to determine clusters in the overlapping regions. These clusters contain both positive and negative data points. Fuzzy memberships of these data points are set to 1 and fuzzy memberships of other data points are determined by their closest cluster accordingly. Although clustering is performed in the input space, according to most of current kernel functions, relative distances between data points are preserved so we can apply the clustering results obtained in the input space to the feature space. The following algorithm is used to determine fuzzy memberships for all data points:

Fuzzy Membership Calculation Algorithm

1. Select a clustering algorithm
2. Perform clustering on the training data set
3. Determine a subset containing clusters that contain both positive and negative data. Denote this subset as MIXEDCLUS.
4. For each data point $x \in$ MIXEDCLUS, set its fuzzy membership to 1
5. For each data point $x \notin$ MIXEDCLUS, do the following
   - Find nearest cluster to x
   - Calculate fuzzy membership of x to this cluster

4.2. The Role of Fuzzy Memberships

The term $\sum_i \lambda_i \xi_i$ is regarded as a weighted sum of empirical errors to be minimised in fuzzy SVMs. If a misclassified point x_i is not in a mixed cluster, its fuzzy membership λ_i is small and hence its error ξ_i can be large as long as $\lambda_i \xi_i$ is still minimised. On the other hand, if it is in a mixed cluster, its fuzzy membership is 1 and hence its error ξ_i must be small such that $\lambda_i \xi_i$ remains minimised. This means that the decision boundary tends to move to overlapping region to reduce empirical errors in this region.

5. Experimental Results

The $aGender$ corpus for the Age and Gender Sub-Challenges [14] was used to evaluate the proposed fuzzy SVM method. There are 7 groups in this corpus including 4 age groups which are child (C), youth (Y), adult (A), and senior (S), 2 gender groups which are female (f) and male (m), and 1 group of children (x) without gender discrimination. The number of features is 450 as detailed in [13]. Theses features were extracted using the open-source Emotion and Affect Recognition toolkits feature extracting backend openSMILE [15].

Speech data in the $aGender$ corpus were divided into 3 subsets which were labeled as $Train$, $Develop$, and $Test$. There were 7 classes including Cx (class 1), Yf (class 2), Ym (class 3), Af (class 4), Am (class 5), Sf (class 6) and Sm (class 7) used in our experiments. The $Train$ set was used to train SVM and fuzzy SVM, and the $Develop$ set was used for testing. From the classification results for 7 classes, we combined them to have the classification results for the 4 age groups and for the 3 gender groups. We also used the combined $Train$ and $Develop$ set for training and the $Test$ set for testing. This experiment was designed for the Challenge competition.

Experiments were performed using LIBSVM tool [18]. All feature values were scaled to range $[-1, 1]$ in order to avoid domination of some dimension to general performance of classifiers. We performed preliminary experiments to compare RBF kernel with linear kernel in SVM for gender classification (3 classes x, f and m). Different values of parameters $C = 2^{-2}, 2^{-1}, \ldots, 2^{12}$ and $\gamma = 2^{-12}, 2^{-11}, \ldots, 2^{2}$ were applied to search for the best accuracy. The $Train$ set was used for training and the $Develop$ set for testing. Results are shown in Figure 1 and we found the best values $C = 2^{7}, \ldots, 2^{12}$ and $\gamma = 2^{-11}, \ldots, 2^{-9}$, for example unweighted accuracy = 77.67% with $C = 2^{11}$ and $\gamma = 2^{-11}$. This result is higher than that (76.99%) for linear kernel reported in [13].

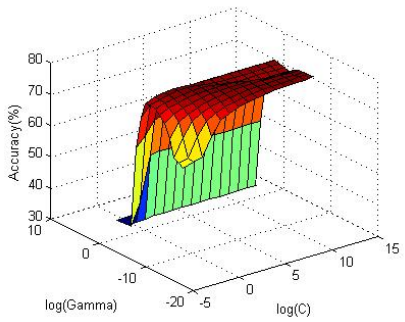

Figure 1: Classification accuracy versus C and γ for SVM using an RBF kernel.

5.1. Experiment 1: Equal membership for feature vectors in the same class

In this experiment, feature vectors in the same class were assigned the same membership. We set $\lambda_i = 0.75$ to all feature vectors in class 1 (Cx) and $\lambda_i = 0.5$ to all feature vectors in other classes. The reason is that feature vectors in class 1 were highly misclassified so they need to have a larger membership value. SVM parameters $C = 2^{10}$ and $\gamma = 2^{-11}$. From this confusion matrix in Table 1 for 7 classes, we can calculate classification rates for age and gender groups and list these results in Table 2. Confusion matrices for gender only and age only are not listed here. Table 3 compares the baseline method and the proposed method. We can see that the fuzzy SVM achieved higher results for both age and gender classification compared to the baseline results presented in [13].

Table 1: Confusion matrix for fuzzy SVM where the same membership was assigned to feature vectors in the same class.

	Cx	Yf	Ym	Af	Am	Sf	Sm
Cx	**1262**	494	96	228	88	166	62
Yf	469	**1412**	13	503	14	283	28
Ym	31	13	**917**	47	568	105	489
Af	161	783	41	**1315**	28	1006	27
Am	8	6	737	100	**785**	61	815
Sf	176	460	70	1128	59	**1554**	114
Sm	11	2	531	68	1021	107	**2086**

Table 2: Classification results where unweighted accuracy (% UA) was calculated.

Age	% UA
C	52.67
Y	51.58
A	37.94
S	52.27
Average	**48.61**

Gender	% UA
x	52.67
f	87.56
m	93.43
Average	**77.89**

Table 3: Classification results for the baseline [13] and fuzzy SVM 1 methods. The results are unweighted accuracy (% UA)

Sub-Ch	Task	Baseline	FSVM1
	{1,...,7}	44.24	45.04
Age	{1,...,7} → {C, Y, A, S}	47.11	48.61
	{C, Y, A, S}	46.22	47.61
Gender	{1,...,7} → $\{x, f, m\}$	77.28	77.89
	$\{x, f, m\}$	76.99	78.58

5.2. Experiment 2: Different memberships for all feature vectors

In this experiment, feature vectors were assigned different membership values using the algorithm in Section (4.1). We use fuzzy c-means clustering where number of clusters was approximated as $\sqrt{(n_1 + n_2)/2}$, n_1 and n_2 are number of vectors of two classes in one pair of pairwise SVM. The best SVM parameters were $C = 2^6$ and $\gamma = 2^{-9}$.

Table 4: Confusion matrix for fuzzy SVM where different memberships were assigned to feature vectors.

	Cx	Yf	Ym	Af	Am	Sf	Sm
Cx	**1261**	500	96	216	79	176	68
Yf	473	**1411**	12	500	12	287	27
Ym	33	12	**952**	52	470	107	544
Af	171	778	37	**1290**	24	1036	25
Am	10	5	770	99	**684**	70	874
Sf	180	452	68	1070	47	**1624**	120
Sm	17	2	560	68	891	115	**2173**

Table 5: Classification results obtained where unweighted accuracy (% UA) was calculated.

Age	% UA
C	52.63
Y	52.28
A	35.71
S	54.58
Average	**48.80**

Gender	% UA
x	52.63
f	87.60
m	93.07
Average	**77.76**

Table 4 presents the confusion matrix which was used to calculate classification rates for age and gender groups listed in Table 5. Confusion matrices for gender only and age only are not listed here. Table 6 compares the baseline result and experiment 2 result. The fuzzy SVM in the second experiment achieved the highest results for both age and gender classification compared to the baseline and the experiment 1 results.

6. Conclusion

We have presented a fuzzy approach to support vector machine and applied this method to age and gender classification. Experimental results on the $aGender$ corpus have shown that the

Table 6: Classification results for the baseline [13] and fuzzy SVM 2 methods. The results are unweighted accuracy (% UA)

Sub-Ch	Task	Baseline	FSVM2
	Train vs Develop		
	$\{1,\ldots,7\}$	44.24	45.19
Age	$\{1,\ldots,7\} \rightarrow \{\text{C, Y, A, S}\}$	47.11	48.80
	$\{\text{C, Y, A, S}\}$	46.22	47.68
Gender	$\{1,\ldots,7\} \rightarrow \{x, f, m\}$	77.28	77.76
	$\{x, f, m\}$	76.99	78.99
	Train + Develop vs Test		
Age	$\{1,\ldots,7\} \rightarrow \{\text{C, Y, A, S}\}$	48.83	49.06
Gender	$\{1,\ldots,7\} \rightarrow \{x, f, m\}$	81.21	81.65

proposed fuzzy support vector machine could improve the classification accuracy for both age and gender classification compraing with the baseline results for the INTERSPEECH 2010 Paralinguistic Challenge.

7. References

[1] Mitchell, A.G. and Delbridge, A., "The Pronunciation of English in Australia", A. & R., Syd, 1119, 1965

[2] Vergin, R., Farhat, A. and O'Shaughnessy, D.,"Robust Gender-Dependent Acoustic-Phonetic Modelling in Continuous Speech Recognition Based on a New Automatic Male-Female Classification", in Proceedings of International Conferences on Spoken Language Processing (ICSLP), 2, 1081-1084, 1996.

[3] Huang, X.D., "Improved acoustic modeling for the SPHINX speech recognition system", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 345-348, Toronto, Canada, 1991.

[4] Metze, F. Ajmera, J. Englert, R. Bub, U. Burkhardt, F. Stegmann, J. Muller, C. Huber, R. Andrassy, B. Bauer, J.G. Littel, B., "Comparison of Four Approaches to Age and Gender Recognition for Telephone Applications", Proc ICASSP 2007, 4, 1089-1092, 2007.

[5] Bocklet, T., Maier, A., Bauer, J.G., Burkhardt, F. and Noth, E., "Age and gender recognition for telephone applications based on GMM supervectors and support vector machines," Proc. ICASSP, 7, 7.2, 2008.

[6] Muller, C., Wittig, F. and Baus, J., "Exploiting speech for recognizing elderly users to respond to their special needs," in Proc. Eurospeech 2003, Geneva; Switzerland, Sept. 2003.

[7] Minematsu, N., Sekiguchi, M. and Hirose, K., "Automatic estimation of ones age with his/ her speech based upon acoustic modeling techniques of speakers," in Proc. ICASSP 2002, Orlando, FL; USA, May 2002.

[8] Vapnik, V., "The nature of statistic learning theory", New York: Springer Verlag, 1995.

[9] Lin, C.-F. and Wang, S.-D., "Fuzzy Support Vector Machines", IEEE Transactions on Neural Networks, 13(2), 464-472, 2002.

[10] Lee, G.H., Taur, J. S. and Tao, C.W., "A Robust Fuzzy Support Vector Machine for Two-class Pattern Classification", International Journal of Fuzzy Systems, 8(2), 76-87, 2006.

[11] Hsu, C.-C., Han, M.-F., Chang, S.-H., Chung, H.-Y., Fuzzy Support Vector Machines with the Uncertainty of Parameter C, Expert Systems With Application, 44, 6654-6658, 2009.

[12] Le, T., Tran, D., Ma, W. and Sharma, D., "A Novel Parameter Refinement Approach for One Class Support Vector Machine," submitted to PRICAI 2010 conference.

[13] Schuller, B.; Steidl, S.; Batliner, A.; Burkhardt, F.; Devillers, L.; Mueller, C.; Narayanan, S.: The INTERSPEECH 2010 Paralinguistic Challenge, INTERSPEECH (2010), ISCA, Makuhari, Japan, 2010.

[14] Burkhardt, F., Eckert, M., Johannsen, W. and Stegmann,J., "A Database of Age and Gender Annotated Telephone Speech," in Proc. 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta, 2010.

[15] Eyben, F., Wollmer, M. and Schuller, B. , "openEAR - Introducing the Munich Open-Source Emotion and Affect Recognition Toolkit," in Proc. ACII, Amsterdam, Netherlands, 576581, 2009.

[16] Moya, M.M., Koch, M.W., and Hostetler, L.D., "One-class classifier networks for target recognition applications", Proceedings of World Congress on Neural Networks, 797-801, 1991.

[17] Campbell, C. & Bennet, K.P., "A linear programming approach to novelty detection", *Advances in neural information processing*, 2001.

[18] Chang, C.-C. and Lin, C.-J., LIBSVM: A Library for Support Vector Machines. Downloadable at http://www.csie.ntu.edu.tw/ cjlin-libsvm