Evaluation of SVM Kernels and Conventional Machine Learning Algorithms for Speaker Identification

D. Ben Ayed Mezghani, S. Zribi Boujelbene, N. Ellouze
Dept. Electrical Engineering at the National School of Engineer of Tunis
National School of Engineer of Tunis (ENIT), Tunis – Tunisia
Dorrainsat@yahoo.fr, zribi_siwar@yahoo.fr, N.Ellouze@enit.rnu.tn

Abstract

One of the central problems in the study of Support vector machine (SVM) is kernel selection, that's based essentially on the problem of choosing a kernel function for a particular task and dataset. By contradiction to other machine learning algorithms, SVM focuses on maximizing the generalisation ability, which depends on the empirical risk and the complexity of the machine. In the following paper, we considered the problem of kernel selection of SVMs classifiers to achieve performance on text-independent speaker identification using the TIMIT corpus. We were focused on SVM trained using linear, polynomial and Radial Basic Function (RBF) kernels. A preliminary study has been made between SVM using the best choice of kernel and three other popular learning algorithms, namely Naive Bayes (NB), decision tree C4.5 and Multi Layer Perceptron (MLP). Results had revealed that SVM trained using polynomial kernel is the best choice for dealing with speaker identification tasks and that SVM is the best choice when compared to other algorithms.

Keywords: *Machine learning algorithms, Support Vector Machine, Naive Bayes, Decision Tree C4.5, Multi Layer Perceptron, kernels function, speaker identification.*

1. Introduction

Support vector machine (SVM) was the first proposed kernel based algorithm. It uses a kernel function to transform data from input space into a high dimensional feature space in which it searches for a separating hyperplane [2]. Linear, polynomial, RBF and others kernel functions are commonly used to transform input space into desired feature space. According to Vapnik [24], traditional learning algorithms for pattern recognition are based on the principle of Empirical Risk Minimization, in an attempt to optimize the performance of the learning set. Furthermore, SVM is based on the principle of Structure Risk Minimization by taking into account of the probability of misclassifying yet to be seen patterns for a fixed but unknown probability distribution of data. It uses a linear separating hyperplane to create a classifier, yet it is not easy to separate some problems in the original input space linearly. But it can easily transform the original input space into a high dimensional feature space non-linearly, where it is trivial to find an optimal linear separating hyperplane. This hyperplane is optimal in the sense of being a maximal margin classifier with respect to the learning set [1]. According to Joachims [6], SVM is the best machine learning algorithm for text classification by choosing the polynomial and RBF kernels. In [1] Ali and Abraham observed that SVM gives the maximum efficiency for different datasets for linear classification and compared his behaviour with three other machine learning algorithms. So, what is the best kernel for a particular problem such as speaker recognition and what is the best machine learning algorithm if we formulate a comparative study for the speaker recognition task?

In this paper, we were attempted to investigate the best choice among SVM kernels namely linear, polynomial and RBF kernels then we were attempted to make a preliminary study between SVM with the best choice of kernel and three popular machine learning algorithms including NB, C4.5 and MLP. These latest four algorithms were evaluated for closed-set text independent speaker identification using the TIMIT corpus, which provided high quality recordings of speech. The task was not straightforward especially since we were required the application of different types of kernels by using different feature datasets.

This paper was organized as follow: in section 2 the speaker identification task was briefly described. In section 3 an overview of four popular machine learning algorithms was defined including SVM, NB, C 4.5 and MLP. In section 4, simulations were presented and finally results were discussed in section 5.

2. Speaker identification

Speaker recognition, which can be classified into identification and verification task refers to the concept of recognizing a speaker by his/her voice [11]. Figure 1 shows the abstraction of an automatic speaker recognition system.

The objective of speaker verification is to verify the claimed identity of that speaker based on the voice samples of that speaker alone. Speaker identification deals with a situation where the person has to be identified as being one among a set of persons by using his/her voice samples. The speaker identification problem may be subdivided into closed-set and open-set. If the target speaker is assumed to be one of the registered speakers, the recognition task is a closed-set problem. If there is a possibility that the target speaker is none of the registered speakers, the task is called an open-set problem. In general, the open-set problem is much more challenging. In the closed-set task, the system makes a forced decision simply by choosing the best matching speaker from the speaker database. However, in the case of open-set identification, the system must have a predefined tolerance level so that the similarity degree between the unknown speaker and the best matching speaker is within this tolerance. Another distinguishing aspect of speaker identification systems is that they can either be text-dependent or text-independent depending on the application. In the text-dependent case, the input sentence or phrase is fixed for each speaker, whereas in the text-independent case, there is no restriction on the sentence or phrase to be spoken.

For such problem, a classifier consists of M speaker models (one for each speaker) and the decision logic necessary to render a decision. In the training phase, the feature vectors are used to create a model for each speaker. During the testing phase, when the test feature vector comes in, a number will be associated with each speaker model indicating the degree of match with that speaker's model. This is done for a set of feature vectors and the derived numbers can be used to find a likelihood score for each speaker's model. For the speaker identification task, the feature vectors of the test utterance will be passed through all the speakers' models and the scores are calculated. The model having the best score gives the speaker's identity (which is the decision). The output or score of the models may be a distortion measure or probability depending on the type of model. The identification success rate of the system is calculated as the ratio of the number of test cases for which the speaker is identified correctly to the total number of test cases for all the speakers. In open set problems, a scheme is used wherein a threshold value is needed in order to find out if the speaker is out of the set of M speakers. In closed set speaker identification, there is one source of error, namely,

when a speaker is not identified correctly. In the open set case, there are two additional sources of error. First, a speaker not in the set of M speakers is deemed to be within the set. The second is the opposite scenario when a speaker in the set is deemed to be outside the set.

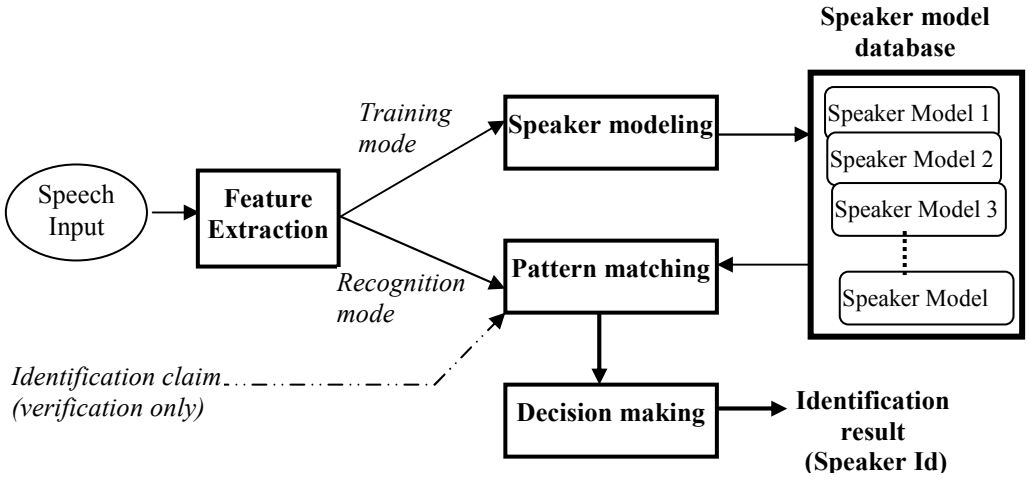

Figure 1. Components of automatic speaker recognition system.

3. Machine learning algorithms for speaker recognition

Machine learning is one of the hottest research areas of data mining. It has been widely adopted in speaker recognition task. This section gives a brief overview of four machine learning algorithms related to this study, namely SVM, NB, C4.5, and MLP.

3.1. Support Vector Machine

An SVM is a kernel machine method that makes its decisions by constructing a hyperplane that optimally separates two classes. The hyperplane is defined by $x \cdot w + b = 0$ where w is the normal to the plane.

For linearly separable data presented by $\{x_i, y_i\}, x_i \in \Re^d, y_i \in \{-1,1\}, i = 1...N.$ the optimal hyperplane is determined according to the maximum margin criterion. This is achieved by minimising

$$\|w\|_2^2 \, subject \;\; to \, (x_i \cdot w + b) y_i \geq 1, \forall i. \tag{1}$$

The solution for the optimal hyperplane w_0 is a linear combination of a small subset of data $x_s, s \in \{1...N\}$, known as the support vectors that satisfy $(x_s \cdot w_0 + b) y_s = 1.$

For not linearly separable data, no hyperplane exists for which all points that satisfy the inequality above. To overcome this problem, ζ_i are introduced and the object is then achieved by minimising

$$\frac{1}{2}\|w\|_2^2 + C\sum_i L(\zeta_i) subject\ to\ (x_i \cdot w + b)y_i \geq 1 - \zeta_i \tag{2}$$

Where L is the loss function, C is a hyper-parameter used to specify the effects of minimising the empirical risk and maximising the margin and the empirical risk associated with the marginal or misclassified points is presented by the term on the RHS. According to Burges [4], the dual formulation, which is more conveniently solved, of (2) with $L(\zeta_i) = \zeta_i$ is

$$\begin{aligned} &\max_{\alpha}\left(\sum_i \alpha_i + \sum_{i,j}\alpha_i\alpha_j y_i y_j x_i \cdot x_j\right) subject\ to \\ &0 \leq \alpha_i \leq C \\ &\sum_i \alpha_i y_i = 0 \end{aligned} \tag{3}$$

Where α_i is the Lagrange multiplier of the i^{th} constraint in the primal optimization problem. The dual can be solved using standard quadratic programming techniques. The optimal plane, w_0 is given by

$$w_0 = \sum_i \alpha_i y_i x_i \tag{4}$$

The extension to non-linear boundaries is achieved through the use of kernels that satisfy Mercer's condition. In practice, kernels commonly used are linear kernels polynomial kernels and radial basis function (RBF) kernels. The linear kernels take the form

$$\mathrm{K}(x_i, x_j) = (x_i \cdot x_j) \tag{5}$$

The polynomial kernels take the form

$$\mathrm{K}(x_i, x_j) = (x_i \cdot x_j + 1)^p \tag{6}$$

where p is the degree of the polynomial. Radial basis functions have received significant attention, most commonly with a Gaussian of the form

$$\mathrm{K}(x_i, x_j) = \exp\left[-\frac{1}{2}\left(\frac{\|x_i - x_j\|}{\sigma}\right)^2\right] \tag{7}$$

Where σ is the width of the radial basis function. Classical techniques utilizing radial basis functions employ some method of determining a subset of centers. Typically, a method of clustering is first employed to select a subset of centers. An attractive feature of the SVM is

that this selection is implicit, with each support vectors contributing one local Gaussian functions, and centred at that data point.

The dual for the non-linear case is thus

$$\max_{\alpha}\left(\sum_{i}\alpha_i+\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathrm{K}(x_i,x_j)\right) subject\ to$$
$$0\le\alpha_i\le C \tag{8}$$
$$\sum_i \alpha_i y_i = 0$$

For speaker recognition, the first approach in using SVM classifiers was implemented by Schmidt in [12]. Another approach became recently more popular, consists of using a combination of GMMs and SVMs. In [17], [19], [20] several types of combination were proposed.

3.2. Naive Bayes

Naive Bayes is the simplest form of Bayesian network, in which all attributes are independent given the value of the class variable. This is called conditional independence. In [16], Zhang suggested that, despite their simplicity, the NB classifier is one of the most efficient and effective inductive learning algorithm for classification and it has been found to perform surprisingly well.

The NB algorithm computes a discriminate function for each of n possible classes. Let E be an example vector, with a features $\{X_p,...X_a\}$, and $B_i(E)$ the discriminate function corresponding to the i^{th} class. The chosen class, C_k, is the one for which

$$B_k(E) > B_i(E) \forall i \neq k \tag{9}$$

The discriminate function $B_i(E)$ is defined as

$$B_i(E) = \Pr(Class = C_i)\prod_{j=1}^{a}\Pr(X_j = v_j | Class = C_i) \tag{10}$$

Where v_j is the value of the feature X_i in example E. The classification rule might be changed to reflect some desired operating conditions: in a two class problem the rule might be changed such that if one discriminate were above a given threshold, then that class would be assigned, regardless of the value of the other discriminate.

For speaker recognition, NB algorithm is largely used such in [10], [9], etc.

3.3. Decision tree C4.5

The decision tree C4.5 is admittedly the most widely classifier used for pattern recognition [23]. It uses an improved version of the ID3 algorithm. Among the merits of the C4.5 algorithms are: its applicability in a variety of learning problems, its computational efficiency and the human-readable format of the induced models. At the same time it provides comprehensible rules which are harder to interpret.

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of Information Entropy. The training data is a set $S = s_1, s_2,...$ of already classified samples. Each sample $s_i = x_1, x_2,...$ is a vector where $x_1, x_2,...$ represent attributes or

features of the sample. The training data is augmented with a vector $C = c_1, c_2, \ldots$ where $c_1, c_2, \ldots$ represent the class that each sample belongs to.

C4.5 uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. It examines the normalized Information Gain that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is the one used to make the decision. The algorithm then recurs on the smaller sub lists.

This algorithm has a few base cases; the most common case is when all the samples belong to the same class. Once this happens, a leaf node of the decision tree will be created. It might also happen that no information gain had given. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class. It might also that never seen any instances of a class, again.

For speaker recognition, the feature vectors are obtained from the training data for all speakers. Then, the data is labeled and a binary decision tree is trained for each speaker. The leaves of the binary decision tree identify the class label as follow: a one corresponds to the speaker and a zero corresponds to "not the speaker". For speaker identification, all feature vectors are applied to each decision tree for the test utterance. The labels are scored and the speaker having the maximum accumulated score is selected. For speaker recognition, decision tree classifier is used in several studies such as in [14], [13].

3.4. Multi Layer Perceptron

Perceptron networks are the most basic type of neural network and are particularly useful for many classification problems.

The network functions through combining the various inputs with some set of weights. This sum is then used as input for a single neuron's activation function. The output of the activation function is then taken to be the output of the network. Perceptrons with multiple outputs are composed of several independent perceptron networks each determining the value of a single output. That is, if the output is a three-dimensional vector (X_1, X_2, X_3), then each X_i is computed by a separate network and final vector is the combination of these outputs.

There are many known algorithms for use in training this type of network. Back propagation algorithm was the most popular algorithm characterized by it's optimality. It's an algorithm for modifying the weights of Multi Layer Perceptron based on incremental gradient descent of mean-square error. It aims at minimising the squared error cost function over a training set presented as follow:

$$E = \sum_{i=1}^{N} \left\{ \frac{1}{2} \sum_{j=1}^{j} \left(d_{ij} - y_{ij} \right)^2 \right\} \tag{11}$$

Where i indexes each pattern in the training set and j indexes each output variable (N patterns in the training set; each patter has J outputs); d_{ij} is the desired value of output j as given by example pattern i, and y_{ij} is the actual output value from the model.

For speaker recognition, test vectors, from training data, should have a "one" response for that speaker's MLP for a specific speaker, whereas from different speakers, test vectors should have a "zero" response. For speaker identification, all test vectors are applied to each MLP and the outputs of each vector are accumulated. The speaker is identified as a corresponding to the MLP with the maximum accumulated output [5].

4. Simulations

4.1. Speech Corpus

The speech corpus for the experiments reported in this paper is a subset of the DARPA TIMIT corpus. This set presented 47 speakers of the same DR1 dialect. The set of speakers included 16 females and 31 males. Each speaker prompted to read ten sentences. Two sentences have the prefix "sa" (sa1 and sa2). Sentences "sa1" and "sa2" are different, but they are the same across speakers. Three sentences have the prefix "si" and five have the prefix "sx". These eight later sentences are different from one another and different across the speakers.

4.2. Front-End Processing and Feature Extraction

Front-end processing of speech data from TIMIT corpus encompassed splitting each conversation into separate utterances, volume normalization and pre-emphasis. Later speech signal was split into frames. The Hamming window was applied to each frame. The frames of data corresponding to silence were removed from the utterances. The data were recorded at a sample rate of 16 KHZ and a resolution of 16 bits.

The feature extraction process consists of obtaining characteristic parameters of a signal to be used to classify the signal. For speaker recognition, the features extracted from a speech signal should be invariant with regard to the desired speaker while exhibiting a large deviation from the features of an imposter. The selection of speaker unique features from a speech signal is an ongoing issue. It has been found that certain features yield better performance for some applications than do other features. Thus far, no feature set has been found to allow perfect discrimination for all condition.

In this study, we used the classical parameterization based on 12 Mel frequency cepstral coefficients (MFCC), the energy and the first and the second derivatives Delta and Delta-Delta coefficient through MATLAB Toolbox. In fact, these coefficients are the current most commonly used in speaker identification task [7], [18]. In the one hand, Zunjingand and Zhigang [22] suggested that MFCC has been widely accepted as such a front-end for a typical speaker identification system as it is less vulnerable to noise perturbation, gives little session variability and is easy to extract. In the second hand, Zanuy and Chetouani [15] suggested that the use of energy can improve the robustness of the system in the sense that it is less affected by the transmission channel than the spectral characteristics, and therefore it is a potential candidate feature to be used as a complement of the spectral information. In the third hand, Nosratighods et al. [8] suggested that dynamic cepstral features such as Delta and Delta-Delta cepstral coefficients have been shown to play an essential role in capturing the transitional characteristics of the speech signal that can contribute to better recognition.

Experiments were done on four sets of data defined as follow (see table 1):

Table 1. Datasets and splitting.

Datasets	Size (ko)	Instances	Characteristics of every utterance
W1	135	470	middle frame
W3	398	1410	three middle frames
W5	660	2350	five middle frames
W7	922	3290	seven middle frames

4.3. Classification

Machine learning algorithms presented previously were evaluated and compared for speaker identification task. We have chosen the SVM, the MLP, the C4.5 and the NB algorithms from Weka data mining package where different kernels have been evaluated for training SVM whereas default settings have been used for the other three algorithms. We normalized the training and testing datasets, as data normalization is required for some kernels due to restricted domain, and may be advantageous for unrestricted kernels.

SVM algorithm was evaluated using different kernels. For SVM using polynomial kernel, the value of p was changed from 1 to 100. Only 1, 10 and 100 were chosen in this evaluation. For SVM using RBF kernel, the value of σ was changed from 0.01 to 1. Only 0.01 and 0.1 were chosen in this evaluation. To handle the SVM multi-class problem, we were considered the one versus one approach classifier. This approach avoids several problems encountered with other approaches: first it is much easier to separate two speakers than to separate one speaker from all others, second, the number of training vectors is roughly equal for both classes.

All experiments were carried out using a ten-fold cross validation approach. In fact, this approach has been unfortunately under-utilized in machine learning community [3]. In [21], Zribi Boujelbene et al. suggested that cross validation approach was the most powerful to estimate the generalization error of speech recognition systems. For our experiments, data were randomly partitioned into ten equally sized where 90% were used for training and the remaining 10% were used for testing. This technique was repeated for ten times, each time with a different test data. For each time i, the identification rate (IR_i) was computed by:

$$IR_i(\%) = \frac{Num.\,Correct\,Identif.\,vectors(time_i)}{Total\,Num.\,of\,Vectors(time_i)} \times 100. \quad (12)$$

The mean of the ten IR_i present the IR of all data.

5. Results and discussion

Our motivation was to analyse how much the different datasets results depending on the kernel function training SVMs and on the choice of the machine learning algorithms.

5.1. Performance evaluation of SVM kernels for speaker identification task

In this part of study, we were attempted to investigate the best choice among SVM kernels for speaker identification task. Table 2 presented the performance of SVMs trained with linear, polynomial and RBF kernels on the TIMIT corpus by using different datasets.

Table 2 showed that SVMs trained using linear kernel had achieved an IR between 5.11 % and 39.40 %. However SVMs trained using polynomial kernel with p = 1 had achieved an IR between 5.32 % and 38.89 %. We remarked that these two classifiers had almost the same behaviour whatever datasets used.

It can be seen that SVMs trained using polynomial kernel with p = 10 had reported a marked improvement in speaker identification performance gradually with increasing the frame numbers. Indeed the IR for W1 database was equal to 6.38 %, the IR for W3 database was equal to 74.89 %, the IR for W5 database was equal to 81.79 % and the IR for W7 database was equal to 77.81 %. We noted that the best IR was reported by using W5 database.

It can be seen also, that SVMs trained using polynomial kernel with $p = 100$ had reported the lowest IR equal to 2.13 % for all datasets. We suggested that SVMs trained using polynomial kernels performed worst by increasing the n value.

It can be noted that SVMs trained using RBF kernel with $\sigma = 0.01$ had achieved an IR between 4.68 % and 23.77 %, however, this IR was reported between 4.25 % and 68.51% where $\sigma = 0.1$. We concluded that SVMs trained using RBF kernel with $\sigma = 0.1$ performed better than SVMs trained using RBF kernel with $\sigma = 0.01$ for W3, W5 and W7. However SVMs trained using polynomial kernel with $p = 10$ performed significantly better than SVMs trained using RBF kernel with $\sigma = 0.1$ for all datasets.

Table 2. Performance evaluation of SVM kernels for speaker identification task (%)

Data	Liniaire	Polynomial ($p = 1$)	Polynomial ($p = 10$)	Polynomial ($p = 100$)	RBF ($\sigma = 0.01$)	RBF ($\sigma = 0.1$)
W1	5.11	5.32	6.38	2.13	4.68	4.25
W3	36.95	36.59	74.89	2.13	21.35	68.51
W5	39.40	38.89	**81.79**	2.13	22.98	28.25
W7	38.14	37.96	77.81	2.13	23.77	28.45

It can be easily observed that SVMs trained using polynomial kernel ($p = 10$) applied in W5 datasets had provided the best performance than other kernels.

5.2. Performance evaluation of SVM, NB, C4.5 and MLP for speaker identification task

In this part of study, we were attempted to investigate the best choice among four popular machine learning algorithms for speaker identification task. Table 3 presented the performance of SVM, NB, C4.5 and MLP on the TIMIT corpus by using different datasets.

Table 3. Performance evaluation of SVM, NB, C4.5 and MLP identification rate for speaker identification task (%)

Data	SVM	NB	C4.5	MLP
W1	6.38	30.21	4.68	6.81
W3	74.89	29.43	28.16	46.52
W5	**81.79**	31.79	35.91	50.85
W7	77.81	30.21	34.32	49.88

Table 3 showed that NB algorithm had achieved an IR between 29.43% and 30.21%. We remarked that this classifier had the same behavior whatever datasets used. However C4.5 algorithm had achieved an IR between 4.68% and 35.91% where the best performance was reported by W5 dataset. Table 3 showed also, that MLP algorithm had achieved an IR between 6.81% and 50.85% where the best performance was reported by W5 dataset. Whereas SVM training polynomial kernel **($p = 10$)** showed the highest IR compared to all these latest machine learning algorithms. This IR is equal to 81.79% assured by W5 dataset. So, we can conclude that the W5 dataset reported the best performance for all latest algorithms.

Figure 2 presented the IR average of SVM, NB, C4.5 and MLP classifiers for all the datasets. It showed that the SVM algorithm provided significantly the best performance than NB, C4.5 and MLP machine learning algorithms. In deed, the IR average for all datasets provided by C4.5 algorithm is equal to 25.77%, the IR average for all datasets provided by NB algorithm is equal to 30.41%, the IR average for all datasets provided by MLP algorithm is equal to 38.51% and the IR average for all datasets provided by SVM algorithm is equal to 60.22%.

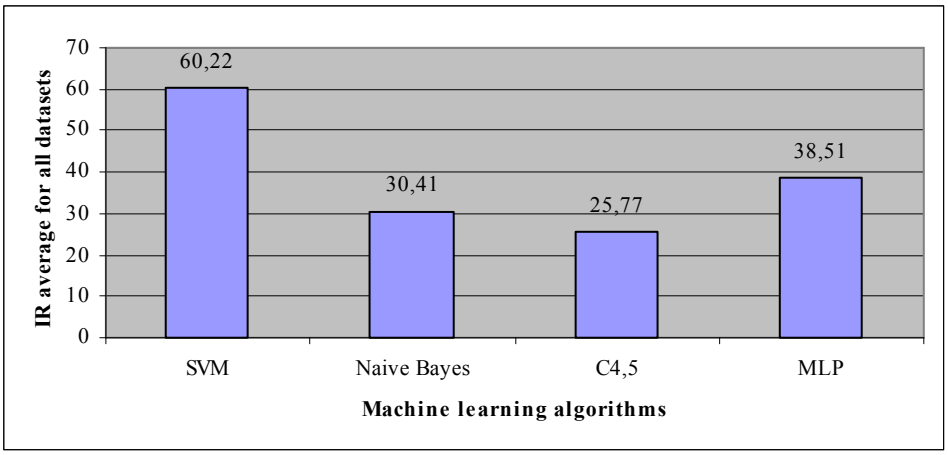

Figure 2. Comparative study of classifier performance for all datasets

6. Conclusion

In this paper, we have presented the performance of SVM algorithm using different kernels on four different datasets and a comparison was made with conventional machine learning algorithms to solve the speaker identification task. In deed, we were attempted to investigate the best choice among SVM kernels namely linear, polynomial and radial basis function (RBF) kernels for text independent speaker identification using the TIMIT corpus. Different degree of the polynomial kernel and different width of the RBF kernel were evaluated. To specify our feature space, we have explored all sentences pronounced by all male and female speakers to assure a multi-speaker environment. Four datasets were defined depending on the number of middle frame for each feature vector. W1 dataset was characterized by one middle frame, W3 dataset was characterized by three middle frames, W5 dataset was characterized by five middle frames and W7 dataset was characterized by seven middle frames. Then a comparative study was made to investigate the best choice among SVM trained using the best choice of kernel function, Naïve Bayes, C4.5 decision tree and MLP algorithms.

Our studies reveal that SVM polynomial kernel provided the best performance than other kernels even for large datasets and the resulting SVM algorithm was consequently very efficient and provided the best performance than NB, C4.5 and MLP algorithms. However, varying kernels is not the ideal solution for SVM optimization especially that the selection of the type of kernel is not an easy task even if the choice of these parameters has a significant effect on the performance of this algorithm. In addition, The NB, C4.5 and MLP algorithms can also be improved and adopted to our problem.

Thus, as a future work, we will focus on optimizing the NB, C4.5 and MLP systems by adopting its parameters and we will include other comparison criteria such as the complexity of different algorithms and the time of learning as a measure of classifier speed.

7. References

[1] A. B. M. S. Ali, and A. Abraham, "An Empirical Comparison of Kernel Selection for Support Vector Machines", Soft computing systems: design, management and applications, Ajith Abraham, Javier Ruiz-del-Solar,Mario Köppen (eds.), IOS Press Publisher, Amsterdam, 2002, pp. 321-330.

[2] M. Bak, "Support Vector Classifier with Linguistic Interpretation of the Kernel Matrix in Speaker Verification", Man-Machine Interactions, Krzysztof A. Cyran,Stanislaw Kozielski, James F. Peters (eds.), ISSN 1867-5662, Springer, 2009, pp 399-406.

[3] M. Bentoumi, G. Millerioux, G. Bloch, L. Oukhellou and P. Aknin, "Classification de Défauts de Rail par SVM," Congrès International IEEE de Signaux, Circuits et Systèmes SCS'04, Tunisie, 2004, pp. 242-245.

[4] C. J. C. Burges and B. Schölkopf, "Improving the Accuracy and Speed of Support Vector Learning Machines", Advances in Neural Information Processing Systems 9, Cambridge, MIT Press, 1997, pp. 375–381.

[5] M. Gerber, R. Beutler, B. Pfister, "Quasi Text-Independent Speaker-Verification based on Pattern Matching", INTERSPEECH 2007, Antwerp, Belgium, August 2007, pp. 1993-1996.

[6] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features", the 10th European Conference on Machine Learning ECML-98, 1998, pp. 137-142.

[7] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuyn and J. Sole-Casals, "Maximum Likelihood Linear Programming Data Fusion for Speaker Recognition," Speech Communication, 2008, doi: 10.1016/j.specom.2008.05.009.

[8] M. Nosratighods, E. Ambikairajah, and J. Epps, "Speaker Verification using A Novel Set of Dynamic Features", the 18th International Conference on Pattern Recognition, vol. 4, 2006, pp. 266-269.

[9] S. Pigeon, P. Druyts and P. Verlinde, "Applying Logistic Regression to the Fusion of the NIST'99 1-Speaker Submissions", Digital Signal Processing, Vol. 10.1-3, January 2000, pp. 237-248.

[10] P. Rose, "Technical Forensic Speaker Recognition: Evaluation, Types and Testing of Evidence", Computer Speech & Language, Vol. 20.2-3, April-July 2006, pp. 159-191.

[11] D. Reynolds, "An Overview of Automatic Speaker Recognition Technology", the International Conference on Acoustics, Speech, and Signal Processing ICASSP 02, Orlando, Florida, USA, 2002, pp. 4072–4075.

[12] M. Schmidt, "Identifying Speaker with Support Vector Networks", In Interface Proceedings, Sydney, 1996.

[13] W. Wang, Ping Lv, QingWei Zhao and YongHong Yan, "A Decision-Tree-Based Online Speaker Clustering," the Third Iberian Conference IbPRIA 07, Vol. 4477, Spain, June 2007, pp. 555–562.

[14] J. Yamagishi, T. Masuko, K. Tokuda, and T. Kobayashi, "A Training Method for Average Voice Model Based on Shared Decision Tree Context Clustering and Speaker Adaptive Training", the International Conference on Acoustics, Speech, and Signal Processing ICASSP 03, vol. 1, April 2003, pp.716-719.

[15] M. F. Zanuy and M. Chetouani, "Nonlinear predictive models: overview and possibilities in speaker recognition," Workshop on Nonlinear Speech Processing, Springer, 2005, pp. 170-189.

[16] H. Zhang, "The Optimality of Naive Bayes", the 17th International FLAIRS conference, Florida, USA, May 2004, pp. 17-19.

[17] S. Zribi Boujelbene, D. Ben Ayed Mezghani and N. Ellouze, "Applications of Combining Classifiers for Text-Independent Speaker Identification", the 16th IEEE International Conference on Electronics, Circuits and Systems ICECS 09, pp. 723-726, Hammamet-Tunisia, December 2009.

[18] S. Zribi Boujelbene, D. Ben Ayed Mezghani, and N. Ellouze, "Improved Feature data for Robust Speaker Identification using hybrid Gaussian Mixture Models - Sequential Minimal Optimization System", The International Review on Computers and Software, Vol. 4.3, ISSN: 1828-6003, May 2009, pp.344-350.

[19] S. Zribi Boujelbene, D. Ben Ayed Mezghani, and N. Ellouze, "Robust Text Independent Speaker Identification Using Hybrid GMM-SVM System", Journal of Convergence Information Technology – JDCTA, Vol. 3.2, ISSN: 1975-9339, June 2009, pp.103-110.

[20] S. Zribi Boujelbene, D. Ben Ayed Mezghani, and N. Ellouze, "Support Vector Machines approaches and its application to speaker identification", IEEE International Conference on Digital Eco-Systems and Technologies DEST-09, Turkey, Jun 2009, pp.662-667.

[21] S. Zribi Boujelbene, D. Ben Ayed Mezghani and N. Ellouze, "Vowel Phoneme Classification Using SMO Algorithm for Training Support Vector Machines", the IEEE International Conference on Information and Communication Technologies: from Theory to Applications ICTTA-08, Syria, 2008, pp. 1-5.

[22] W. Zunjingand, and C. Zhigang, "Improved MFCC-Based Feature for Robust Speaker Identification", Tsinghua Science and Technology, Vol. 10.2, 2005, pp. 158-161.

[23] Quinlan J. R., C.4.5: programs for machine learning, Morgan kaufmann, 1993.

[24] Vapnik V., The nature of statistical learning theory, Springer, N.Y., 1995.

Authors

D. Ben Ayed Mezghani received computer science engineering degree in 1995 from the National School of Computer Science (ENSI-Tunisia), the MS degree in electrical engineering (signal processing) in 1997 from the National School of Engineer of Tunis (ENIT-Tunisia), the Ph. D. degree in electrical engineering (signal processing) in 2003 from (ENIT-Tunisia).
She is currently an associate professor in the computer science department at the High Institute of Computer Science of Tunis (ISI-Tunisia). Her research interests include fuzzy logic, support vector machines, artificial intelligence, pattern recognition, speech recognition and speaker identification.
E-mail:Dorra.mezghani@isi.rnu.tn, dorrainsat@yahoo.fr

S. Zribi Boujelbene received a university diploma in computer science in 1999 from the High Institute of Management of Tunis (ISG-Tunisia), the MS degree in electrical engineering (signal processing) in 2004 from the National School of Engineer of Tunis (ENIT-Tunisia) and prepared a Ph. D. degree in electrical engineering (signal processing) from (ENIT-Tunisia).
She is currently a teacher in the computer science department at the Faculty of Humanities and Social Sciences of Tunis (FSHST-Tunisia). Her research interests include data mining approaches such as decision tree, support vector machines, neural networks, fuzzy logic, and pattern recognition such as speech recognition and speaker recognition.
E-mail: zribi_siwar@yahoo.fr, zribi.siwar@planet.tn

N. Ellouze received a Ph.D. degree in 1977 from l'Institut National Polytechnique at Paul Sabatier University (Toulouse-France), and Electronic Engineer Diploma from ENSEEIHT in 1968 at the same University.
In 1978, Dr. Ellouze joined the Department of Electrical Engineering at the National School of Engineer of Tunis (ENIT–Tunisia), as assistant professor in statistic, electronic, signal processing and computer architecture. In 1990, he became Professor in signal processing; digital signal processing and stochastic process. He has also served as director of electrical department at ENIT from 1978 to 1983. General Manager and President of the Research Institute on Informatics and Telecommunication IRSIT from 1987-1990, and President of the Institut in 1990-1994. He is now Director of Signal Processing Research Laboratory LSTS at ENIT, and is in charge of Control and Signal Processing Master degree at ENIT.
Pr Ellouze is IEEE fellow since 1987; he directed multiple Masters and Thesis and published over 200 scientific papers both in journals and proceedings. He is chief editor of the scientific journal Annales Maghrébines de l'Ingénieur. His research interest include neural networks and fuzzy classification, pattern recognition, signal processing and image processing applied in biomedical, multimedia, and man machine communication.
E-mail:N.Ellouze@enit.rnu.tn