SVM-RFE based Feature Selection for Network Speaker Recogntion task

Meriem FEDILA
Speech Communication & Signal Processing Lab.
Faculty of Electronics and Computer Science, USTHB
Algiers, Algeria.
fedila_m@yahoo.fr

Abderrahmane AMROUCHE
Speech Communication & Signal Processing Lab.
Faculty of Electronics and Computer Science, USTHB
Algiers, Algeria.
namrouche@usthb.dz

Abstract— **This paper investigates the impact of Support Vector Machine Recursive Feature Elimination SVM- RFE method used for feature selection and feature ranking on speaker recognition performance in network environment. The motivation behind reducing the dimension of the feature set is by the fact that features are not all equally important to identify a speaker. In the present work, we thought to use SVM- RFE based feature selection to remove the irrelevant features influenced by speech coding algorithms, transmission errors and environmental noise of decoded speech. We find that the SVM-RFE selection method achieves comparable performance on network speaker recognition (NSR) system, while it obtains excellent performance with only few features. Results demonstrate the effectiveness of the feature selection method on the transcoded TIMIT database obtained using G722.2 speech coder together with the 6.60Kbit/s, 8.85 Kbit/s, 12.65 Kbit/s and 23.85 Kbit/s bit-rates.**

Keywords-SVM-RFE; NSR; MFCC; G722.2; Feature Selection.

I. INTRODUCTION

As the demand for mobile communications is continuously increasing, it is expected that an increasing number of transactions using speaker recognition will take place through the mobile cellular network. Furthermore, the process of coding and decoding modifies the speech signal; it is likely to have an influence on speaker recognition performance, together with other perturbations introduced by the mobile cellular network (channel errors, background noise). In fact, improve the performance and robustness of automatic speaker recognition in mobile communication systems has become an active topic and a number of techniques has been proposed to enhance the degradation introduced by speech coders (GSM AMR-NB, AMR-WB…) and channel errors. For instance, in [1] and [2], some studies have been done consisting of techniques to ameliorate the performance accuracy of the system. But, the performance is still poorer than that achieved by using uncoded speech.

In this paper we thought to work on the features extraction stage of decoded speech. The aim of this paper is to improve the performance of NSR system by the choice of the most relevant features that reduce the degradation caused by decoded speech. For this, we have investigated the use of feature selection technique.

Feature selection provides effective ways to discover relevant features for many learning tasks [3]. Using only the relevant features, we can perform data mining in reduced spaces, thereby producing more stable learning models (which often lead to more accurate prediction) in shorter time. Such models are also easier to understand and to apply. The support vector machine recursive feature elimination (SVM-RFE) is one of the most effective feature selection methods which have been successfully used in selecting features for classifications task.

II. FEATURE SELECTION

A. Feature selection Definition

The definition of feature selection differs according to the authors. Although its advantage is general, it consists on the search for a sufficiently reduced subset of **d** features out of the total number of available ones **D** without significantly degrading or even improving in some cases the performance of the resulting classifier when using either set of features. This search is driven by a certain measure of criterion function which is used to assess the validity of each feature subset. Figure 1 shows the general procedure of feature selection.

B. Advantages of feature selection

- It removes the redundant, irrelevant or noisy data.
- It reduces the dimensionality of the feature space, to increase algorithm speed;
- The immediate effects for data analysis tasks are speeding up the running time of the learning algorithms.

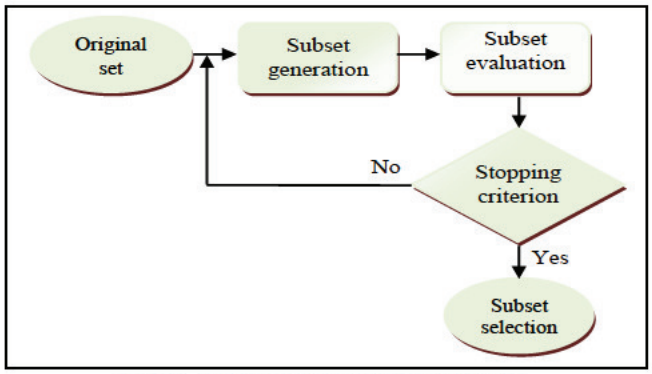

Figure1. General Procedure of feature selection.

- Improving the data quality.
- Increasing the accuracy of the resulting model.

C. Types of Feature Selection Methods

There are many feature selection algorithms with numerous ways to measure relevance and redundancy of features. We broadly categorize them into three types [4]:

1) Filter Methods: These methods select features based on discriminating criteria that are relatively independent of classification. They are also known as the variable ranking methods.

2) Wrapper Methods: Whereas the filter approach that completely ignores the influence of feature selection on performance of the classifier, Wrapper methods utilize the classifier as a black box to score the subsets of features based on their predictive power. Wrapper methods based on SVM have been widely studied in machine-learning community. SVM-RFE (Support Vector Machine Recursive Feature Elimination), a wrapper method applied to recursively remove insignificant features from subsets of features.

3) Embedded Methods: Embedded Methods perform feature selection in the process of training and are usually specific to given learning machines, where the search is guided by the learning process.

III. Support Vector Machine Recursive Feature Elimination (SVM-RFE)

SVM is a classification algorithm based on statistical learning theory [5]. It is designed to separate vectors in a two class problem. It is also known as maximum margin classifier. The SVM problem can be reduced to find the hyperplane that maximizes the distance from it to the nearest examples in each class. An SVM classifier has the general form:

$$f(x) = \left[\sum_{t=i}^{N} \alpha_i y_i K(x_i, x) + b \right]. \quad (1)$$

Where α_i are the training data. Each point of x_i belongs to one of the two classes identified by the label y_i {-1,1} . The coefficients α_i and b are the solutions of a quadratic programming problem. α_i are non-zero for support vectors (SV) and are zero otherwise. K is the kernel function and it used when data are not linearly separable in the finite dimensional space.

Support vector machine recursive feature elimination (SVM-RFE) is known as an excellent feature selection algorithm derived from the classical SVM. SVM-RFE [6] is a wrapper feature selection method which generates the ranking of features using backward feature elimination. Its basic idea is to eliminate redundant features and evaluate the contribution of each feature to the classification error in the sense of a maximum margin criterion. The features are eliminated according to a criterion related to their support to the discrimination function, and the SVM is retrained at each step. RFE SVM is a weight based method; at each step, the coefficients of the weight vector of a linear SVM are used as the feature ranking criterion. The feature with the smallest contribution to $||w||^2$ is removed. The SVM-RFE algorithm can be broken into four steps:

- Train the SVM classifier with the current feature set
- Compute the contribution of each feature
- Eliminate the feature with smallest contribution to the norm of w from the current feature set.
- Start over again from the step (1) until a desired number of features is reached.

IV. Experimental Protocol

A. Description of the database

The TIMIT database [7] contains speech from 630 speakers (438 men and 192 women), each of them speaking 10 sentences. The speech signal is recorded at 16 kHz and 16 bit pulse code modulated (PCM). The text material consists of 2342 sentences, divided into 2 dialect sentences (SA sentences), 450 phonetically compact sentences (SX sentences) and 1890 phonetically diverse sentences (SI sentences). Each speaker reads the two SA sentences, 5 of the SX sentences and 3 of the SI sentences.

In our experience, we used the "long training / short test" protocol for Independent-text Speaker Identification on the TIMIT database. There are a total of 90 speakers, 34 female and 56 male. For training each speaker model, we concatenated the features corresponding to the two sentences from "SA" portion of the corpus, five sentences from the "SX", and one from "SI", for a total of eight sentences. We used two different sentences from the "SI" portion in the speaker identification system for testing each speaker (90x2=180 test patterns of 3.2 seconds each, in average).

In order to transcode TIMIT database, the software library ITU-T [8] was used to simulate the G.722.2 codec. G.722.2 is a wideband speech codec, with a sampling frequency of 16 kHz, it also known as AMR-WB and consists of nine source codec modes with bit-rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.60 Kbit/s. For each bit-rate we obtained the resynthesized (transcoded) speech by decoding of the transmitted coded speech (figure 2).

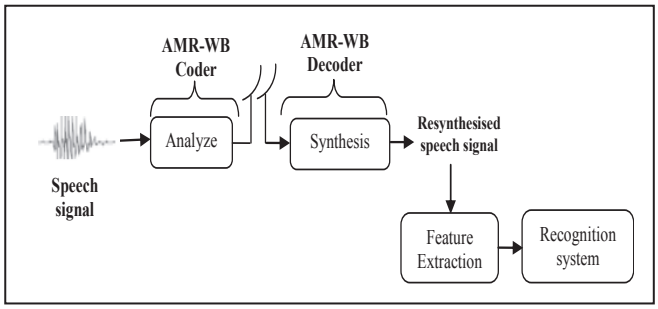

Figure 2. Network Speaker Recognition Architecture.

B. Baseline system

Our baseline system is based on NSR architecture shown in figure 2, a total of 21 MFCCs is extracted every 20 ms from the resynthesized speech; we added the log-energy, 0'th order cepstral coefficient, delta (Δ) and acceleration (ΔΔ) coefficients, forming a 63 dimension feature vector. These parameters are then used as input to the GMM system for the modeling phase (with GMM of 16 Gaussians).

For Classification phase, the performances of our system are measured by the CIR (Correct Identification Ratio) defined as the ratio of number of positive identified tests to the total number of accuracy tests.

V. EXPERIMENT RESULTS

A. Network Speaker Recognition using G.722.2

First, we test the effect of the resynthesized speech on performance of NSR and for previous bit allocation of G722.2 (6.60 Kbit/s, 8.85 Kbit/s, 12.65 Kbit/s and 23.85 Kbit/s). Results for Correct Identification Ratio (CIR) are shown in Figure 3.

The results show performance degradation when the bit-rate decreases. The CIR is less than 70% for bite-rate of 6.60 Kbit/s. Another interesting observation, especially from the high bit-rate (23.85 Kbit/s) results, is that NSR performances are more than 98%.

B. SVM-RFE

In this section, we evaluated a SVM-RFE method that used for feature selection on our NSR system. TABLE I presents the feature ranking in ascending order as it was acquired by the Support Vector Machine Recursive Feature Elimination (SVMRFE) for various bit-rates.

Experiments with selected feature vectors, from each ranking list, were performed. Figure 4 presents comparative experimental results for the feature vectors (for each vector, we removed the irrelevant features as ordered in TABLE I).

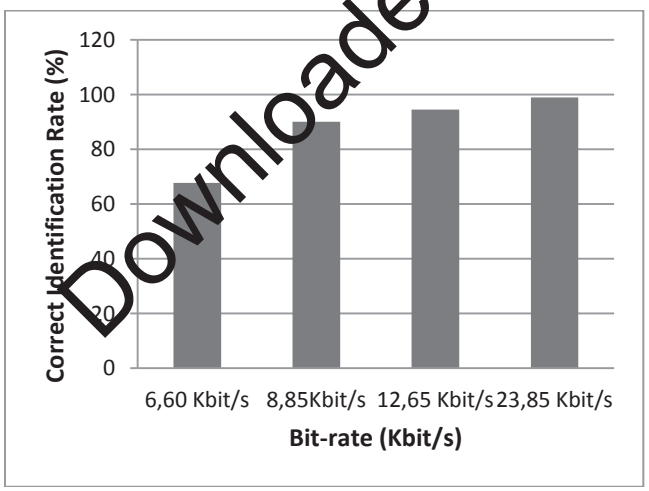

Figure 3. Correct Identification Rate on % dependent on bit-rates 6.60 Kbit/s, 8.85Kbit/s, 12.65Kbit/s and 23.85Kbit/s).

TABLE I. FEATURE RANKING IN DESCENDING ORDER FOR EACH BIT-RATE.

#	*6.60 Kbit/s*	*8.85 Kbit/s*	*12.65 Kbit/s*	*23.85 Kbit/s*
1	MFCC(9)	MFCC(19)'	MFCC(2)''	MFCC(10)'
2	MFCC(10)	MFCC(1)'	MFCC(13)	MFCC(5)''
3	MFCC(14)'	MFCC(0)	MFCC(0)	MFCC(5)
4	MFCC(1)	Log E''	MFCC(10)	Log E'
5	MFCC(7)'	MFCC(2)''	MFCC(0)'	MFCC(1)
6	MFCC(1)'	MFCC(3)'	MFCC(5)	MFCC(1)
7	MFCC(15)''	MFCC(8)''	MFCC(1)	MFCC(1)''
8	MFCC(0)''	MFCC(1)''	MFCC(5)''	MFCC(6)''
9	MFCC(2)''	MFCC(5)	MFCC(8)''	MFCC(7)
10	MFCC(1)''	MFCC(7)	MFCC(0)''	MFCC(3)''
11	MFCC(9)''	MFCC(3)''	MFCC(1)	MFCC(0)'
12	MFCC(15)'	MFCC(18)'	MFCC(13)	Log E''
13	MFCC(19)'	MFCC(16)	Log E''	MFCC(14)
14	MFCC(18)'	MFCC(8)	MFCC(14)''	MFCC(0)''
15	MFCC(3)''	MFCC(11)	Log E'	MFCC(0)
16	MFCC(8)	MFCC(5)''	MFCC(9)'	MFCC(19)'
17	Log E'	MFCC(9)'	MFCC(16)'	MFCC(2)''
18	MFCC(0)'	MFCC(0)	MFCC(7)	MFCC(11)'
19	MFCC(10)''	MFCC(0)'	MFCC(8)	MFCC(13)'
20	MFCC(4)	MFCC(9)''	MFCC(11)''	MFCC(7)
21	MFCC(5)	Log E'	MFCC(1)'	MFCC(9)'
22	MFCC(11)''	MFCC(11)''	MFCC(15)''	MFCC(18)''
23	MFCC(6)	MFCC(3)'	MFCC(10)'	MFCC(15)'
24	MFCC(14)	MFCC(3)	MFCC(1)''	MFCC(10)''
25	MFCC(0)	Log E	MFCC(3)'	MFCC(1)'
26	MFCC(10)'	MFCC(3)''	MFCC(3)	MFCC(4)
27	MFCC(3)'	MFCC(7)'	MFCC(11)	MFCC(13)''
28	MFCC(8)'	MFCC(4)'	Log E	MFCC(5)'
29	Log E	MFCC(13)'	MFCC(6)''	Log E
30	MFCC(4)	MFCC(17)	MFCC(15)	MFCC(2)
31	MFCC(2)	MFCC(12)'	MFCC(17)''	MFCC(12)'
32	MFCC(11)''	MFCC(13)''	MFCC(11)'	MFCC(13)
33	MFCC(19)''	MFCC(4)''	MFCC(4)''	MFCC(3)
34	MFCC(18)	MFCC(12)	MFCC(9)''	MFCC(3)'
35	MFCC(17)''	MFCC(12)	MFCC(19)''	MFCC(4)'
36	MFCC(11)	MFCC(13)	MFCC(6)'	MFCC(17)''
37	MFCC(12)	MFCC(16)''	MFCC(18)''	MFCC(6)
38	MFCC(13)'	MFCC(17)'	MFCC(14)	MFCC(18)'
39	MFCC(6)	MFCC(4)	MFCC(2)	MFCC(6)'
40	MFCC(18)''	MFCC(2)	MFCC(13)''	MFCC(17)'
41	MFCC(17)'	MFCC(11)'	MFCC(18)'	MFCC(8)
42	MFCC(12)'	MFCC(17)'	MFCC(17)	MFCC(11)''
43	MFCC(5)'	MFCC(10)'	MFCC(12)	MFCC(14)'
44	MFCC(2)'	MFCC(10)''	MFCC(8)'	MFCC(8)'
45	MFCC(7)''	MFCC(15)'	MFCC(16)''	MFCC(16)'
46	MFCC(16)''	MFCC(19)''	MFCC(12)'	MFCC(4)''
47	MFCC(9)'	MFCC(2)'	MFCC(18)	MFCC(9)''
48	MFCC(5)	MFCC(6)''	MFCC(2)'	MFCC(2)'
49	MFCC(19)	MFCC(6)'	MFCC(9)	MFCC(14)''
50	MFCC(6)'	MFCC(18)	MFCC(19)'	MFCC(12)
51	Log E"	MFCC(14)''	MFCC(12)''	MFCC(19)
52	MFCC(3)	MFCC(5)'	MFCC(15)'	MFCC(18)
53	MFCC(11)'	MFCC(8)'	MFCC(10)''	MFCC(16)''
54	MFCC(14)''	MFCC(15)	MFCC(7)''	MFCC(8)''
55	MFCC(5)''	MFCC(7)''	MFCC(17)'	MFCC(19)''
56	MFCC(4)''	MFCC(19)'	MFCC(7)'	MFCC(11)
57	MFCC(12)''	MFCC(18)''	MFCC(6)	MFCC(15)''
58	MFCC(60)''	MFCC(9)	MFCC(4)'	MFCC(10)
59	MFCC(16)'	MFCC(12)''	MFCC(19)	MFCC(15)
60	MFCC(14)'	MFCC(10)''	MFCC(3)''	MFCC(17)
61	MFCC(17)	MFCC(5)''	MFCC(5)'	MFCC(16)
62	MFCC(16)	MFCC(18)	MFCC(16)	MFCC(9)
63	MFCC(7)	MFCC(19)	MFCC(4)	MFCC(12)''

As presented in the table, there was no common agreement among all ranking features for different bite-rate. For instance, 6.60Kbit/s presented MFCC(7) as the most informative feature, followed by MFCC(16), for 8.85Kbit/s MFCC(19) is presented as the most relevant parameters. MFCC(16) was selected two times at second position for 6.60 Kbit/s and 12.65 Kbit/s bite-rate. MFCC(0) for 8.85 Kbit/s and 12.65 Kbit/s bite-rate, was placed on the sixtieth position as least relevant parameters.

From this table, we removed n irrelevant features with n=1, 2…, M (M is the number of initial features MFCC, M=63). As show in figure 4, the SVM-RFE method provides the best performance for the subset down-19 features for bit-rates of 8.80 Kbit/s, 12.65 Kbit/s and 23.85 Kbit/s, and only 6.60 Kbit/s has its best subset at down-15 features. This figure illustrate that variable selection helps the classifier to be less influenced by indiscriminative variables.

Table 2 summarizes the Correct Identification Rate for the best subset for bit-rates of 6.60Kbit/s, 8.85 Kbit/s, 12.65 Kbit/s and 23.85 Kbit/s. (We only keep the most significant features so "r" is the number of relevant features).

VI. CONCLUSION

In this work, we applied the idea of feature selection to speaker recognition, specially, for network speaker recognition task (NSR). We have investigated relevance of selected features obtained from SVM-RFE ranking technique. These relevant features were compared to the reference feature set. The experimental results showed that the performance of NSR system are degraded using transcoded TIMIT database (these degradation are caused by speech coding algorithms, transmission errors, bit-rates and environmental noise). But using SVM-RFE selection method, results showed the improvement brought by the discriminative cepstral features for several bit-rates of G722.2 speech coder.

TABLE II. CORRECT IDENTIFICATION RATE ON % FOR THE BEST SUBSET FOR EACH BIT-RATE.

	M=63 n=0	*r=15 n=48*	*r=19 n=44*
6.60kbit/s	67.77%	**88.88%**	-
8.85kbit/s	90.00%	-	**98.88%**
12.65kbit/s	94.44%	-	**100%**
23.85kbit/s	98.88%	**100%**	**100%**

REFERENCES

[1] M. Fedila, A. Amrouche, "Automatic speaker recognition for mobile communications using AMR-WB speech coding", International Conference on Information Sciences, Signal Processing and their Applications (ISSPA), pp. 1034-1038, Montreal, Canada, 3-5 July 2012.

[2] M. Chowdhury, S. Selouani and D. O'shaughnessy, "Text-independent distributed speaker identification and verification using GMM-UBM speaker models for mobile communications," International conference on Information Science, Signal Processing and their Applications (ISSPA), pp. 57- 60, 2010.

[3] T. Ganchev , P. Zervas , N. Fakotakis , and G. Kokkinakis "Benchmarking feature selection techniques on the speaker verification task" Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP'06), pp. 314-318, 2006.

[4] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", Journal of Machine Learning Research 3, pp.1157–1182, 2003.

[5] V. Vapnik, "Statistical learning theory, John Wiley and Sons", New York, 1998

[6] I. Guyon, J. Weston, S. Barnhill, and N. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[7] W. Fisher, V. Zue, J. Bernstein, D. Pallet, "An acousticphonetic database", JASA, suppl. A, Vol. 81(S92), 1986.

[8] ITU-T Recommendation G.722.2, "Wideband Coding of Speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", 2003.

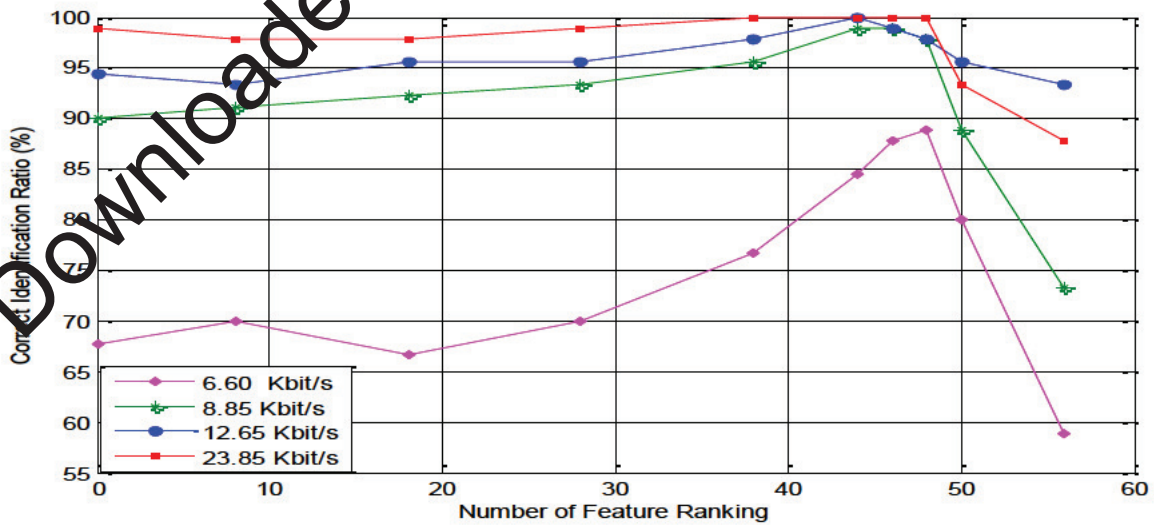

Figure 4. Correct Identification Rate on % dependent on the number of selected features for bit-rates of 6.60 Kbit/s, 8.85Kbit/s, 12.65Kbit/s and 23.85Kbit/s. The x-axis represents the number of chosen features