

# Différentes variantes GMM-SMOs pour l'identification du locuteur

Siwar Zribi Boujelbene\*\*\*\*, Dorra Ben Ayed Mezghani\*\*\*\*\*, Nouredine Ellouze\*\*\*

\*\*\* UR Signal Image et Reconnaissance de Formes, Ecole National des Ingénieurs de Tunis  
- ENIT, BP-37 Campus Universitaire 1002, Tunis, Tunisie

\*zribi\_siwar@yahoo.fr

\*\* dorra.mezghani@isi.rnu.tn

\*\*\* N.ellouze@enit.rnu.tn

**Résumé.** Dans cet article, nous présentons différentes variantes GMM-SMOs pour l'identification du locuteur en mode indépendant du texte. Pour mettre en œuvre les différents systèmes, nous avons opté une représentation multi-gaussienne de l'espace des caractéristiques basées sur l'algorithme Expectation Maximisation (EM). Ces nouvelles représentations constituent les vecteurs d'entrées pour entraîner les supports vecteurs machines (SVMs) par l'algorithme de type Optimisation par Minimisation Séquentielle (SMO).

## 1 Introduction

Les SVMs sont des nouvelles techniques de l'apprentissage statistique (Vapnik, 1995). Leur utilisation en reconnaissance du locuteur été très prometteuse. Dans ce papier, nous nous intéressons à mettre en place une combinaison des modèles GMMs et SVMs à différents noyaux. Les systèmes proposés ont la capacité de modélisation multi-gaussiennes (Reynolds et Rose, 1995) et l'efficacité de décision des SVMs de type SMO (Platt, 1999). Différentes variantes GMM-SMOs sont implémentés et comparés.

Notre article est organisé comme suit : les différents systèmes GMM-SMOs proposés sont présentés dans la section 2 et l'analyse des résultats est donnés par la section 3.

## 2 Systèmes GMM-SMOs pour l'identification du locuteur

Différentes variantes ont été conçues. Les points en communs des différents systèmes résident au niveau de la phase de paramétrisation, qui consiste à déterminer l'espace des caractéristiques extraite du corpus TIMIT, et de la phase de modélisation, qui consiste à discriminer les classes des locuteurs. Durant cette phase, nous avons utilisé la modélisation GMM basé sur l'algorithme EM ce qui permet de générer des nouvelles représentations de l'espace élaboré. Lors de la phase de décision, nous avons utilisés quatre variantes de noyau : noyau polynomial de degré 1, noyau polynomial de degré 10, noyau polynomial de degré 100 et noyau gaussien RBF. Pour étudier la similarité des caractéristiques d'une classe de locuteur avec toutes les autres classes des locuteurs, nous avons exploité la méthode un-contre-un.

Pour valider nos systèmes, nous avons adopté la technique de validation croisée k-fold avec  $k = 10$ . Le locuteur identifié est celui pour lequel la similarité est la plus grande.

### 3 Analyse des résultats

Pour discriminer notre espace des caractéristiques, nous avons varié le nombre d'époque de l'algorithme EM. Les valeurs d'époques retenus sont: 1 000, 10 000, 50 000 et 100 000.

La figure suivante permet de faire une étude comparative entre les différents systèmes. Il est à noter que le meilleur taux d'erreur d'identification présenté par nos systèmes est donné par le système à noyau polynomial de degré 1 caractérisé par 50 000 époques qui est égale à 8.33 %. Si nous comparons ce taux avec les taux donnés par les autres systèmes ayant le même nombre d'époques, nous trouvons que ce taux est visiblement plus faible. En effet, dans les mêmes conditions, le système à noyau polynomiale de degrés 10 présente un taux d'erreur de 16,15 %, le système à noyau polynomiale de degrés 100 présente un taux d'erreur de 19,27 % alors que le système à noyau RBF présente un taux d'erreur de 35,94 %.

Nous constatons aussi, que les systèmes caractérisés par 100 000 époques, donnent des taux d'erreurs qui varient entre 9,90 % et 10,42 % quelques soit le noyau. Nous pouvons dire à ce niveau expérimental, que les systèmes proposés ont presque la même performance.

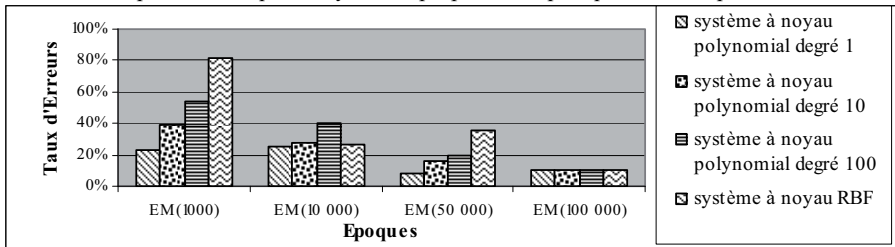


FIG. 1 – Taux d'identification des différents systèmes proposés

### Références

- Reynolds, D. A. et R. Rose (1995). Robust text independent speaker identification using gaussian mixture models, *IEEE Trans. Speech and Audio Process*, 3, 72-83.
- Platt, J.C. (1999). Fast training of support vector machines using sequential minimal optimization, *Advances in Kernel Methods - Support Vector Learning*.
- Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

### Summary

In this article, different variants GMM-SMOs are used for text independent speaker identification. To implement the various systems, we chose a multi-Gaussian representation of the characteristics space based on the Expectation Maximization (EM) algorithm. These representations present the input vectors to train SVMs by using SMO algorithm which allows the computation of similarity between all GMM models.