

Modélisation par mélange de lois Gaussiennes (GMM) et Perceptron multicouches (MLP) sur des données de sons de voyelles enregistrées dans un environnement sans bruit

B. Dato

Université Toulouse III - Paul Sabatier
118 route de Narbonne, 31062 TOULOUSE CEDEX 9
bruno.dato.meneses@gmail.com

20 novembre 2016

Résumé

Les méthodes d'apprentissage supervisées de Modélisation par Mélange de lois Gaussiennes et de Perceptron Multicouches permettent de classifier tout types de données notamment des données audio de voyelles enregistrées dans un environnement sans bruit. Après un pré-traitement des ces données à l'aide de leur cepstre réel puis d'une analyse discriminante linéaire on obtient à l'aide de ces deux méthodes de classification, des scores presque parfait pour 8 voyelles sur 9. On constate aussi que la méthode du Perceptron Multicouches est une méthode plus malléable que celle des mélanges de Gaussiennes.

Mots Clef

Apprentissage supervisé, Modélisation par Mélange de lois Gaussiennes, Perceptron Multicouches, Transformation de Fourier rapide, Cepstre, Analyse Discriminante Linéaire.

Abstract

Supervised learning methods such as Gaussian Mixture Models and Multi-layers enable to classify all sorts of data especially audio data of vowels recorded in a non-noisy environment. After a pre-treatment of those data using real cepstrum and linear discriminant analysis, thanks to this two classification methods, we almost get perfect accuracy for 8 vowels out of 9. We also note that Multi-layer Perceptron is a more malleable method than the Gaussian Mixture one.

Keywords

Supervised learning, Gaussian Mixture Models, Multi-layer Perceptron, Fast Fourier transform, Cepstrum, Linear Discriminant Analysis.

1 Introduction

L'automatisation d'une tâche de perception réalisé usuellement par le cerveau humain, c'est à dire l'apprentissage automatique, est une théorie qui s'est développé depuis les années 1960. Cette théorie se divise en 2 catégories :

- L'apprentissage supervisé : l'ensemble d'apprentissage est collecté en connaissant la forme de chacune des observations, le modèle est donc établi à partir d'un ensemble connu de classes.
- L'apprentissage non supervisé : l'ensemble des données acquises n'est pas étiqueté par des formes connues, le modèle est défini à partir d'une recherche aveugle des classes.

Nous présenterons et comparerons ici deux méthodes de classification supervisées qui sont la Modélisation par Mélange de lois Gaussiennes (GMM) et le Perceptron Multicouches (MLP). La première étant issue d'une approche bayésienne (Bayes 1763, 1939) et la seconde étant basée sur les réseaux de neurones (Rozenblatt 1958).

Nous utiliseront comme jeu de données, des fichier .wav de sons de voyelles enregistrés dans un environnement propre non bruité. Les données sont composées de 9 voyelles différentes et pour chaque voyelles nous disposons de 100 enregistrements différents.

2 Pré-traitement

Avant tout apprentissage supervisé ou non, on effectue tout d'abord un pré-traitement afin de représenter les données dans des espaces plus propices à la classification. C'est à dire des espaces qui permettent de faire apparaître les classes en regroupant les données suivant leurs affinités.

2.1 Transformation de Fourier

Pour représenter ces données, nous avons tout d'abord effectué une transformation de Fourier (FFT) de chacun des fichiers audio après un fenêtrage de type hamming pour supprimer des effets de bord.

$$spectre_{freq} = FFT(spectre_{temp} * fenetragement_{hamming})$$

La transformée de Fourier nous permet de représenter nos données dans le domaine fréquentiel qui est beaucoup plus exploitable que le domaine temporel. Puis nous avons réduit la dimension du spectre obtenu à 32. On peut alors

observer sur la figure 1 une partie de l'ensemble des données dont on dispose pour notre classification. Le spectre de chaque fichier étant de dimension 32, on ne peut représenter dans le plan que 2 dimension de ce spectre, nous avons choisis le 2ème et la 3ème qui sont les plus significatives.

On peut observer alors que certaines voyelles se détachent des autres mais l'allure générale reste tout de même assez ramassée.

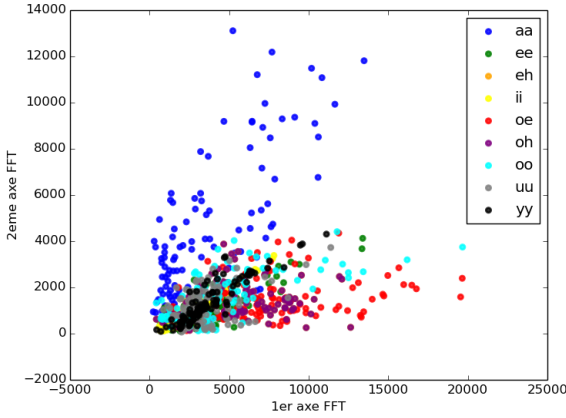


FIGURE 1 – Représentation des classes voyelles par rapport à la 2ème et à la 3ème dimension de la FFT réduite des données audio après un fenêtrage de type hamming

2.2 Cepstre

Afin de séparer d'avantage les classes, nous avons ensuite utilisé le cepstre réel qui lui permet de représenter les données dans un espace analogue au domaine temporel :

$$cepstre_{rel} = |FFT^{-1}(\log(|FFT(Spectre_{temp} * fenetragement_{hamming})))|$$

On peut alors voir sur la figure 2 la représentation des données du cepstre réel centré et réduit dans ce nouveau domaine sur 2 dimensions, les données sont moins ramassées mais la différenciation des classes reste assez difficile dans le plan.

2.3 Analyse Discriminante Linéaire

Enfin, le dernier pré-traitement que nous avons utilisé est l'Analyse Discriminante Linéaire (LDA) sur le cepstre réel réduit (on garde seulement les dimensions 2 à 33). Cette méthode d'analyse permet d'utiliser la connaissance des classes pour les représenter dans un nouveau repère qui maximise l'inertie interclasses et minimise l'inertie intra-classes. On obtient alors une représentation des classes (figure 3) sur laquelle on peut nettement voir apparaître les groupes de données correspondant aux différentes classes de voyelles sur 2 dimensions seulement de la LDA. Même

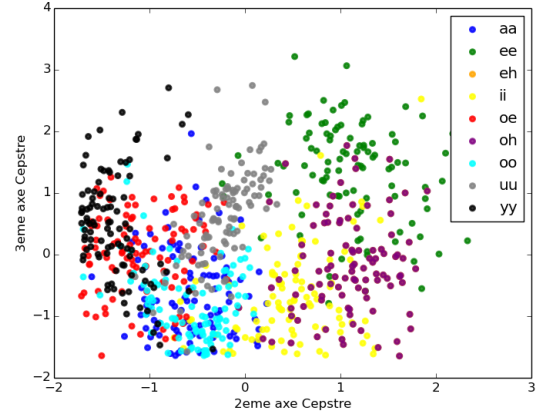


FIGURE 2 – Représentation des classes voyelles par rapport à la 2ème et à la 3ème dimension du cepstre réel centré réduit des données audio après un fenêtrage de type hamming

si certains groupes de données semblent se superposer, on rappelle qu'en réalité, l'analyse discriminante linéaire s'effectue sur plus de 2 dimensions, on effet, si on trace une troisième dimension, on peut voir les classes plus distinctement encore (figure 4).

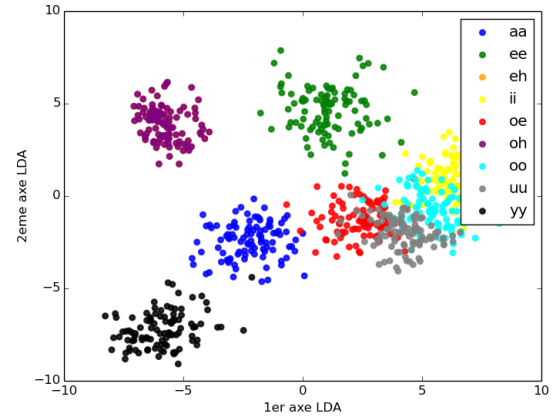


FIGURE 3 – Représentation des classes voyelles par rapport à la 1ère et à la 2ème dimension de l'analyse discriminante linéaire du cepstre réel centré réduit des données audio après un fenêtrage de type hamming

3 Modélisation par mélange de lois Gaussiennes GMM

Maintenant que l'on dispose des données dans un domaine où elles sont plus facilement différenciables, on peut appliquer une première méthode de classification supervisée, la modélisation par mélange de lois Gaussiennes. Un modèle de mélange de lois Gaussiennes est un modèle

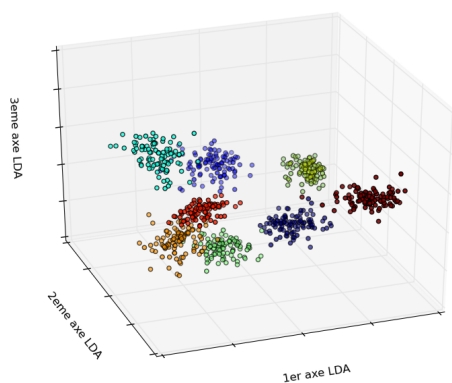


FIGURE 4 – Représentation des classes voyelles par rapport à la 1ère, 2ème et 3ème dimension de l’analyse discriminante linéaire du cepstre réel centré réduit des données audio après un fenêtrage de type hamming

probabiliste qui suppose que chaque donnée appartient à un mélange fini de Gaussiennes de paramètres inconnus. Les algorithmes de mélange de lois Gaussiennes permettent alors de trouver les paramètres correspondant à chacune des classes. Afin de pouvoir tester notre classification nous avons créé le modèle à l’aide de 75% des données, les données d’apprentissage. Puis nous l’avons testé avec les 25% restants. On peut voir sur la figure 5 les différents jeux de gaussiennes obtenus à l’aide des données d’apprentissage pour différents types de covariances.

En moyenne, on trouve des scores d’apprentissage et de test respectivement d’environ 89% et 88% quelque soient les types de covariances choisies pour les modèles.

4 Perceptron Multicouches MLP

La seconde méthode de classification supervisée que nous avons utilisée est celle du Perceptron Multicouches (MLP), c’est un réseau de neurones avec une ou plusieurs couches cachées qui lui permettent de faire un apprentissage plus approfondi qu’un réseau de neurones simple. Nous avons utilisé un réseau avec 3 couches cachées comprenant 64 neurones chacune, un taux d’apprentissage constant de 0.1 et des fonctions d’activation de type relu. Avec ces paramètres on trouve un score d’apprentissage d’environ 89% et un score de test d’environ 87%. Ces résultats sont très semblables aux résultats obtenus pour les méthodes de GMM.

5 Comparaison GMM et MLP

Afin de mieux pouvoir comparer les deux méthodes de classification GMM et MLP, nous avons utilisé les matrices de confusion obtenues pour les données de test sur ces 2 méthodes (figures 6 et 7). On remarque que les matrices sont très proches et donnent de bons résultats pour les 2 méthodes, il y a cependant une classe (oh) qui pour les tests, est prise pour une autre.

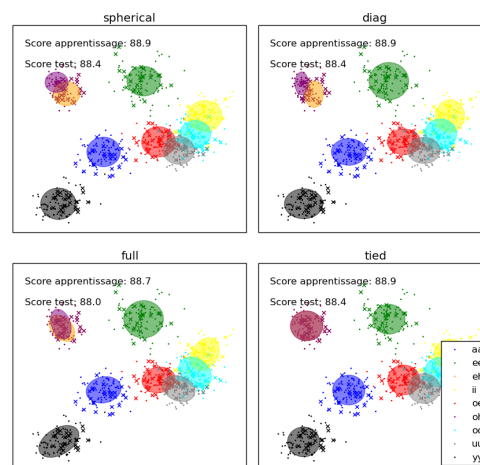


FIGURE 5 – Représentation dans 2 dimensions des classes voyelles modélisées par des lois Gaussiennes pour 4 différents types de covariances sur le cepstre des données pré-traitées par LDA. Les points représentent les données d’apprentissage (75%) et les croix représentent les données de test (25%)

Si maintenant, au lieu d’utiliser 75% des données pour l’apprentissage, on en utilise 50%, on obtient alors de nouvelles matrices de confusion figures 8 et 9. Les résultats pour la méthode GMM sont sensiblement les mêmes mais pour le réseau de neurones, on obtient une matrice moins diagonale que précédemment. Tout en gardant 50% des données en apprentissage et 50% des données en test, on multiplie par 2 le nombre de neurones des couches cachées. On obtient la matrice de confusion figure 10 qui est à nouveau satisfaisante.

6 Conclusion

Après un pré-traitement des données à l’aide d’une analyse discriminante linéaire appliquée sur le cepstre réel de nos données, lorsque l’on applique les méthodes d’apprentissage supervisées qui sont la modélisation par mélange de lois Gaussiennes et le Perceptron Multicouches, les modèles obtenus permettent de classer des voyelles enregistrées dans un environnement sans bruit presque parfaitement.

Si on réduit alors le domaine d’apprentissage, on constate que le Perceptron Multicouches perd en efficacité mais en jouant sur ces paramètres comme le nombre de neurones par couche cachée, il est possible d’obtenir à nouveau des résultats satisfaisants. Il est aussi possible d’utiliser d’autres fonctions d’activation et de modifier le taux d’apprentissage, nous n’avons pas effectué de tels tests dans cette étude. On retiendra que la méthode MLP semble plus malléable concernant ses paramètres que la méthode GMM qui ne permet de jouer principalement que sur les types de covariances.

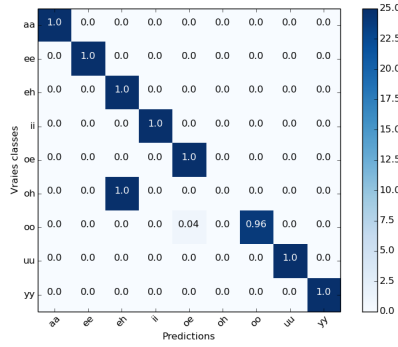


FIGURE 6 – Matrice de confusion normalisée de la méthode GMM pour une covariance de type "tied" sur les données de test (25%)

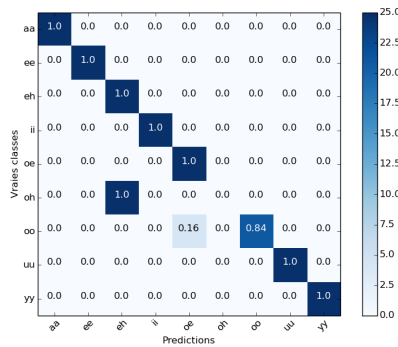


FIGURE 7 – Matrice de confusion normalisée de la méthode MLP sur les données de test (25%) avec 3 couches cachées de 64 neurones

Références

- [1] scikit-learn.org, consulté le 20 novembre 2016.
- [2] J. Farinas, Cours de Reconnaissance des Formes et Apprentissage Automatique, Université Paul Sabatier, 2016.

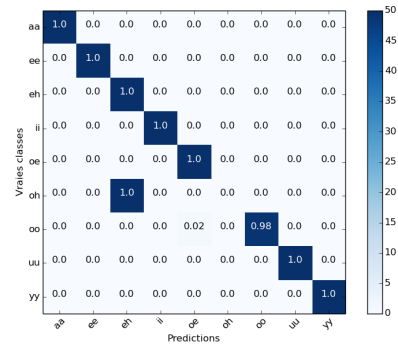


FIGURE 8 – Matrice de confusion normalisée de la méthode GMM pour une covariance de type "tied" sur les données de test (50%)

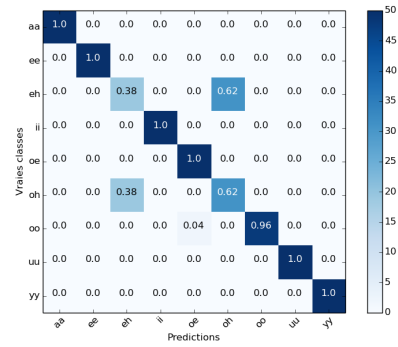


FIGURE 9 – Matrice de confusion normalisée de la méthode MLP sur les données de test (50%) avec 3 couches cachées de 64 neurones

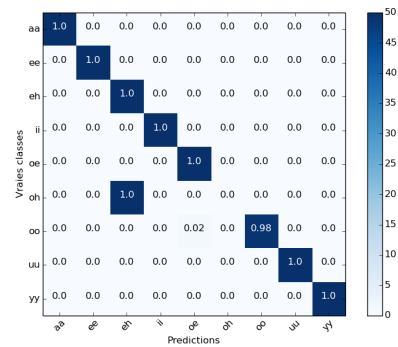


FIGURE 10 – Matrice de confusion normalisée de la méthode MLP sur les données de test (50%) avec 3 couches cachées de 128 neurones