

BE reconnaissance des formes

Veysseire Daniel

Fabre Michaël

Université Paul Sabatier

10 novembre 2014

Résumé

Cet article vise à comparer l'efficacité de deux méthodes de classification (méthode de classification par loi normal multidimensionnel et méthode des K-Plus Proche voisin), ainsi que les choix de paramétrisation des données (FFT, cepstre, MFCC), principalement dans le cadre de la reconnaissance de la parole.

Dans un premier temps, nous ferons une présentation théorique de ces méthodes et paramétrisations. Dans un deuxième temps nous présenterons le protocole expérimental mis en place afin de comparer leurs efficacités.

Nous interpréterons ensuite les résultats obtenus puis nous finirons par une conclusion sur l'efficacité des différentes méthodes et paramétrisations.

Mots Clef

Méthodes de classification, reconnaissance de la parole, loi normale, K plus proche voisins, paramétrisation, FFT, Cepstre, MFCC, apprentissage supervisé.

Abstract

This paper aims to compare the efficiency of two methods of classification (method of classification with normal distribution multidimensional and Nearest neighbor search (NNS)), and the choice of parameterization (FFT, cepstrum, MFCC), mainly in the context of the speech recognition. Primary, we will make a theoretical presentation of these methods and parameterizations. in Secondly, we present the experimental protocol implemented to compare their efficiencies. Finally we interpret the results then finish with a conclusion on the efficiency of these different methods and parameterizations.

Keywords

methods of classification, speech recognition, normal distribution, Nearest neighbor search, NNS, parameterization, FFT, Cepstrum, MFCC, Supervised learning.

1 Introduction

La reconnaissance automatique de la parole est une technique informatique qui permet d'analyser un signal de parole.

On se place ici dans le cas où on essaye de reconnaître chaque syllabe individuellement. On dispose d'une référence de 1000 éléments sonore de 64ms échantillonnés à 16KHz et quantifiés sur 16 bits. On a ainsi 100 échantillons pour chacune des dix syllabes suivantes :

[a],[e],[ε],[ə],[ɪ],[ø],[ɔ],[o],[u],[y]

correspondant aux classes :

'aa','ee','eh','eu','ii','oe','oh','oo','uu','yy' ;

2 paramétrisations et méthodes

2.1 Les différentes paramétrisations

Nous allons utiliser différentes paramétrisations des données et les comparer pour ne conserver que celles qui offrent les meilleurs résultats.

Transformé de Fourier Rapide (FFT)

La transformée de Fourier Rapide est un algorithme permettant de traiter un signal afin d'obtenir son spectre. Le spectre d'un signal nous fournit l'intensité de chacune des plages de fréquences pendant un intervalle de temps t. Elle s'effectue sur un certain nombre de points ; augmenter ce nombre de points diminue la taille des plages de fréquences, et augmente le nombre de plages. On ne garde que la valeur absolue du résultat pour ne pas manipuler des nombres complexes.

En générale on effectue plusieurs FFT sur le signal partitionné, à l'aide de fenêtres glissantes, afin d'obtenir l'intensité des fréquences à plusieurs instants t. Puis on utilise des algorithmes comme le DTW (Dynamic time warping). Mais dans le cas présent dans cette étude, les échantillons sont extrêmement courts (64ms avec une fréquence d'échantillonnage de 16KHz). Utiliser une fenêtre glissante ne s'avère pas nécessaire. On est donc dans un cas simplifié, on ne cherche qu'à comparer des voyelles prononcées dans un temps très court. Une simple FFT sur tout le

signal est donc suffisante, on obtient ainsi un vecteur de taille variable selon le nombre de point sur lesquels on a réalisé la FFT. On comparera par la suite ces vecteurs entre eux (e.g par distance euclidienne). On effectue souvent un lissage du signal par Hamming lorsqu'il y a un recouvrement de fenêtre pour éviter de trop grandes discontinuités entre les fenêtres. Ce serait donc une erreur de faire un lissage par Hamming ici, puisqu'on n'a pas utilisé de fenêtres glissantes.

Le cepstre et les MFCC

Le cepstre est obtenu à partir du spectre. On effectue la transformée inverse du logarithme de la transformée de Fourier (ou spectre) obtenu précédemment. En pratique on ne garde que la valeur absolue du résultat. On obtient ainsi une transformation du signal dans un domaine analogue au domaine temporel. "Les MFCC (Mel-Frequency Cepstral Coefficients) sont des coefficients cepstraux calculés par une transformée en cosinus discrète appliquée au spectre de puissance d'un signal. Les bandes de fréquence de ce spectre sont espacées logarithmiquement selon l'échelle Mel" (wikipedia). Les MFCC sont proches du cepstre, mais diffèrent par l'utilisation de l'échelle Mel, échelle basée sur la perception humaine. Pour calculer ces MFCC j'ai utilisé la fonction MELCEPST disponible sur la toolbox voicebox. Elle réalise une RFFT (DFT of real data, DFT = Discrete Fourier Transform) sur le signal lissé par une fonction hanning adapté à la fréquence d'échantillonnage. Une fois la DFT appliqué, on multiplie la partie réel avec la partie conjugué obtenue. La fonction MELBANKM sert ensuite à calculer la matrice de passage à l'échelle MEL.

On applique un log adapté à la valeur max des données sur ces données obtenues après passage à l'échelle MEL. Puis on refait une RDCT (Discrete cosine transform of real data). On obtiens ainsi les différents coefficients cepstraux de Mel (ce sont des vecteurs).

2.2 Les différentes méthodes de classifications

Comme dit précédemment nous allons comparer les deux méthodes de classifications. Pour classer des données, il faut effectuer au préalable un apprentissage supervisé à partir de données de références. Il y a donc une phase d'apprentissage et une phase de reconnaissance.

classification par loi normale multidimensionnel

Pour utiliser la méthode de classification par loi normale (ou loi gaussienne) multidimensionnel, on suppose que chacune des composantes des vecteurs obtenus par paramétrisation suit une distribution aléatoire. Cette classification prend en paramètre la moyenne et la matrice de variance-covariance des données d'apprentissage. La matrice de variance-covariance est une matrice carrée de taille $N \times N$ (N le nombre de composante du vecteur). Chaque élément placé ligne i et colonne j dans la matrice vaut $\text{cov}(X_i, X_j)$ avec X_i la i ème composante du vecteur. Ainsi sur la

diagonal se situent les variances de chaque composante du vecteur. La covariance se calcule à l'aide de la formule suivante :

$$\text{Cov}(x,y) = E(XY) - E(X)E(Y)$$

La matrice de covariance permet de prendre en compte l'éloignement des données à la moyenne, leur dispersion.

classification par les K-plus proche voisin

La méthode de classification des K-Plus proche voisin est relativement simple. Pour chaque vecteur de test, nous allons trier tous les vecteurs d'apprentissages en fonction de leur distance à ce vecteur. Puis on ne conserve que les K plus proche (K étant un entier choisis au préalable), et si une classe est représenté majoritairement on attribue cette classe à ce vecteur. Sinon on rajoute à la liste le prochain vecteur d'apprentissage le plus proche.

3 protocole expérimental

Les en-têtes sont également en 12 points gras.

Il n'y a pas nécessairement d'espacement entre les paragraphes.

Les références à la Bibliographie peuvent être de la forme [2] où [1]. Les numéros correspondent à l'ordre d'apparition dans la bibliographie, pas dans le texte. L'ordre alphabétique est conseillé.

4 Le coin L^AT_EX

Pour les utilisateurs de L^AT_EX, ce patron est minimaliste et vous aurez besoin de votre manuel L^AT_EX pour insérer équations et images.

Pour les images le « paquet » `graphicx` est très bien.

Les fichiers de style nécessaires pour la compilation L^AT_EX sont :

- `a4.sty` (pour L^AT_EX, mais pas L^AT_EX 2_ε)
- `french.sty` (ou Babel français)
- `rfia2000.sty`

Vous devriez avoir les deux premiers dans votre installation de L^AT_EX. Le dernier contient la définition des marges et vous devrez le récupérer.

Annexe

Merci de votre participation.

nous concacter

wedg@hotmail.fr

Références

- [1] U. Nexpert, *Le livre*, Son Editeur, 1929.
- [2] I. Troiseu-Pami, Un article intéressant, *Journal de Spi-rou*, Vol. 17, pp. 1-100, 1987.