BE reconnaissance des formes

Veysseire Daniel Fabre Michaël

Université Paul Sabatier

6 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 reconnaisance 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 éfficacités.

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

Mots Clef

Methodes de classification, reconnaisance 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 échantillonés à 16KHz et quantifiés sur 16 bits. On a ainsi 100 échantillons pour chacune des dix syllabes suivantes :
[ɑ],[e],[ɛ],[ə],[ɪ],[ø],[ɔ],[o],[u],[y]
correspondant aux classes :
'aa','ee','eh','eu','ii','oe','oh','oo','uu','yy' ;

1.1 Les différentes paramétrisations

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

Transformé de Fourrier Rapide (FFT)
La transformée de Fourrier 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 extremement courts (64ms avec une fréquence d'échantillonage 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 glissante.

Le cepstre et les MFCC

Le cepstre est obtenu à partir du spectre. On effectue la transformée inverse du logarithme de la transformée de Fourrier (ou spectre) obtenu précedement. 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 logarithmement selon l'échelle Mel" (wikipédia). Les MFCC sont proches du cepstre, mais diffère par l'utilisation de l'échelle Mel, échelle basée sur la perception humaine.

1.2 Les différentes méthodes de classifications

Comme dit précédement nous allons comparer les deux méthodes de classifications Pour classifier 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*N (N le nombre de composante du vecteur). Chaque élément placé ligne i et colonne j dans la matrice vaut cov(Xi, Xj) avec Xi la ieme composante du vecteur. Ainsi sur la digonal on a les variances de chaque composante. La covariance se calcule à l'aide de la formule suivante :

(mettre la formule E(XY) - E(X)E(Y))

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

lois normales

16	0	0	3	0	1	0	0	0	0
0	17	0	0	3	0	0	0	0	0
0	0	19	0	0	0	1	0	0	0
0	0	0	14	0	3	3	0	0	0
0	1	0	0	18	1	0	0	0	0
0	0	0	1	0	15	3	1	0	0
0	0	0	0	0	1	17	2	0	0
0	0	0	0	0	0	0	17	3	0
0	0	0	0	0	0	0	0	20	0
0	0	0	0	0	0	0	1	0	19

classification par les K-plus proche voisin

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.

2 Le coin LaTeX

Pour les utilisateurs de LaTeX, ce patron est minimaliste et vous aurez besoin de votre manuel LaTeXpour insérer équations et images.

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

Les fichiers de style nécessaires pour la compilation LaTeX sont :

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

Vous devriez avoir les deux premiers dans votre instalation de LaTeX. 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 Spirou*, Vol. 17, pp. 1-100, 1987.