

**M1 - Informatique****Université Paul Sabatier
31062 Toulouse Cedex 09****IAN - Introduction à l'Audio Numérique**

Travaux dirigés n°4

I Application de la Programmation Dynamique (DTW) pour l'alignement entre deux chaînes

Soit $V = \{A, B, D, E, R\}$ l'ensemble des caractères utilisés pour écrire des chaînes.

Question 1 :

On donne l'alignement temporel suivant entre deux chaînes C1 (la référence) et C2 (la chaîne à reconnaître), en déduire où se situent les insertions, délétions (ou suppressions) et substitutions.

C1 = A A A B B B A A A A

C2 = A R B B D B A E A

Question 2 :

On se propose maintenant d'utiliser l'algorithme de DTW pour trouver le chemin de coût minimal entre la chaîne de référence Ref=AAABBBAAAA et la chaîne en entrée Obs=ARBBDABEA

On considère que les coûts d'insertion et de délétion sont égaux à 1. De plus on considère que la comparaison de 2 caractères se traduit par une substitution de coût 0 s'ils sont identiques ou de coût 1 sinon.

- 1) Calculer le chemin de coût minimal.
- 2) Expliquer le chemin trouvé en termes d'insertions, omissions et substitutions.
- 3) Comparer les résultats des questions 1 et 2.2. Qu'en conclure sur l'alignement proposé à la question 1 ?

II Reconnaissance de la parole par Dynamic Time Warping (DTW ou Programmation Dynamique)

La tâche de reconnaissance envisagée est la reconnaissance orale des nombres « cinq », « vingt » et « cent ».

Le signal acoustique est analysé classiquement par fenêtre de 20 ms, toutes les 10 ms (on parle de trames). Ce pré-traitement consiste à étiqueter chaque trame de signal à partir d'un dictionnaire de références acoustiques :

$$D = \{ X, C, U, U_x, V \}$$

avec les interprétations suivantes :

X représente une trame de bruit non voisé, C une trame de silence, U une trame d'une zone consonantique voisée, U_x une trame d'une zone consonantique 'voisée et bruitée', et V une trame d'une zone formantique.

On suppose que les nombres « cinq », « vingt » et « cent » sont ainsi codés par : « XVVCX », « U_xVV » et « XVV ».

Dans le tableau 1, on dispose des distances entre les références acoustiques.

Question 1 : En s'appuyant sur ces distances et en utilisant des contraintes locales de poids (1,2,1), donner pour chacun des mots de référence, le meilleur chemin ainsi que le score obtenu par l'algorithme de DTW en considérant la suite d'observations "XXV U_x CX". Quel nombre est reconnu ?

| Distance | X | C | U | U_x | V |
|----------|---|---|---|-------|---|
| X | 0 | 1 | 1 | 1 | 2 |
| C | 1 | 0 | 1 | 1 | 2 |
| U | 1 | 1 | 0 | 1 | 1 |
| U_x | 1 | 1 | 1 | 0 | 2 |
| V | 2 | 2 | 1 | 2 | 0 |

tableau 1 : tableau des distances (symétriques) entre les références acoustiques

Question 2 : On considère maintenant que l'on peut reconnaître soit « cinq », soit « vingt », soit « cent », soit « cent cinq », soit « vingt cinq ». Mettre en œuvre l'algorithme de DTW pour les mots enchainés (ou connectés) sur la suite d'observations " $U_xU_xVXXVU_xXXV$ ". Quel nombre est reconnu ?

Question 3 : Est on assuré d'avoir tous les représentants en mode isolé pour envisager les mots connectés tels que 20-5?