Rapport de TP - Apprentissage

IDRISSOU SOULER Hamzath

April 27, 2015

1 nombre voisins

1.1 Question 1.2:

Nous allons dans un premier temps essayer de faire varier le nombre de voisins à regarder afin de pouvoir déterminer correctement le caractère dont on ne connaît que la forme. Pour cela nous allons avoir besoin de la distance entre deux formes. Les formes étant numérisé dans un tableau de 784 octets, il faudra donc faire le calcul de la distance dans la dimension 784 ! Pour cela nous utiliserons la formule suivante : Soit d la distance entre un point α et un point β:

$$Z = \sqrt{\sum_{i=0}^{783}(x_i - y_i)^2}$$

où $x_i = i^e$ case de la forme connue et $y_i = i^e$ case re rechercher.

Une fois, toutes les fonctions utilitaires :génération d'ensemble d'apprentissage et test, recherche des k-plus proches voisins, détermination d'une forme en fonction d'un ensemble de voisins... des tests seront mis en place afin de déterminer le taux d'erreurs et donc l'efficacité de notre programme.

```
1  public static void statisticWithKChange(int maxK, int nbModels, int
       modelSize, int testSize) throws NoMoreDataException {
2    List<List<LabelledData>> models = glyphGenerator.getNbGroupGlyph(nbModels
       , modelSize);
3    List<LabelledData> test = glyphGenerator.getNewGroupGlyph(testSize);
4
5    double EMT = 0;
6    double EMS = 0;
7    maxK++;
8
9    for(int k = 1; k < maxK; k++) {
10     EMT = 0;
11     EMS = 0;
12
13     for (List<LabelledData> curModel : models) {
14       EMS += Utils.erreurMoyenne(curModel, curModel, k);
15       EMT += Utils.erreurMoyenne(curModel, test, k);
16     }
17     EMT /= nbModels;
18     EMS /= nbModels;
19     System.out.println(k + "\t" + EMT + "\t" + EMS);
20   }
21 }
```

Resultat figure 1.

plus k devient grand, plus le taux d'erreurs est elevé que cela soit sur l'ensemble de test ou sur l'ensemble sur les ensembles sources. On remarque aussi qu'il y a un point particulier avec k = 3 où le taux d'erreurs est minimale pour l'ensemble de test.

2 taille d'ensemble d'apprentissage

varions la taille des ensembles d'apprentissage pour voir si cela a un impact sur le taux d'erreurs. On fixera k, le nombre de voisins recherchés, à 3 comme nous l'avons vu avant cette valeur était la meilleur. On se propose de réaliser cette fois ce programme qui est très semblable au précédent :

```
1  public static void statisticWithModelChange(int k, int nbModels, int
       minModelSize, int maxModelSize, int testSize) throws NoMoreDataException
       {
2    if (minModelSize > maxModelSize) {
3      throw new IllegalArgumentException("Min > Max");
4    }
5    List<LabelledData> test = glyphGenerator.getNewGroupGlyph(testSize);
6
7    double EMT = 0;
8    double EMS = 0;
9
10   for(int i = minModelSize; i <= maxModelSize; i++) {
11     EMT = 0;
12     EMS = 0;
13
14     List<List<LabelledData>> models = glyphGenerator.
           getNbGroupGlyphWithoutRetail(nbModels, i);
15
16     for (List<LabelledData> curModel : models) {
17       EMS += Utils.erreurMoyenne(curModel, curModel, k);
18       EMT += Utils.erreurMoyenne(curModel, test, k);
19     }
20     EMT /= nbModels;
21     EMS /= nbModels;
22     System.out.println(i + "\t" + EMT + "\t" + EMS);
23   }
24 }
```

résultats de la Figure 2.

remarque :le taux d'erreurs, peut importe sur les ensembles d'apprentissage ou les données de test, décroit avec un ensemble plus grand de données de base.

3 Problème du nombre de voisins égaux

Une solution, a ce problème consisterai à faire la somme des distances pour les caractères A et pour les caractères B, ensuite on compare les deux sommes et l'on si ce sont globalement les A ou les B qui sont les plus proches.

4 conclusion

Nous avons donc appris au cours de ce TP que pour réduire le taux d'erreurs sur notre programme de reconnaissance de caractères (formes). Il fallait tout d'abord trouver un k optimal pour nous on a trouvé 3. Enfin

nous avons vu que nous pouvions encore améliorer notre programme lors de conflits pour le choix du caractère représenté par la forme que nous recherchons.

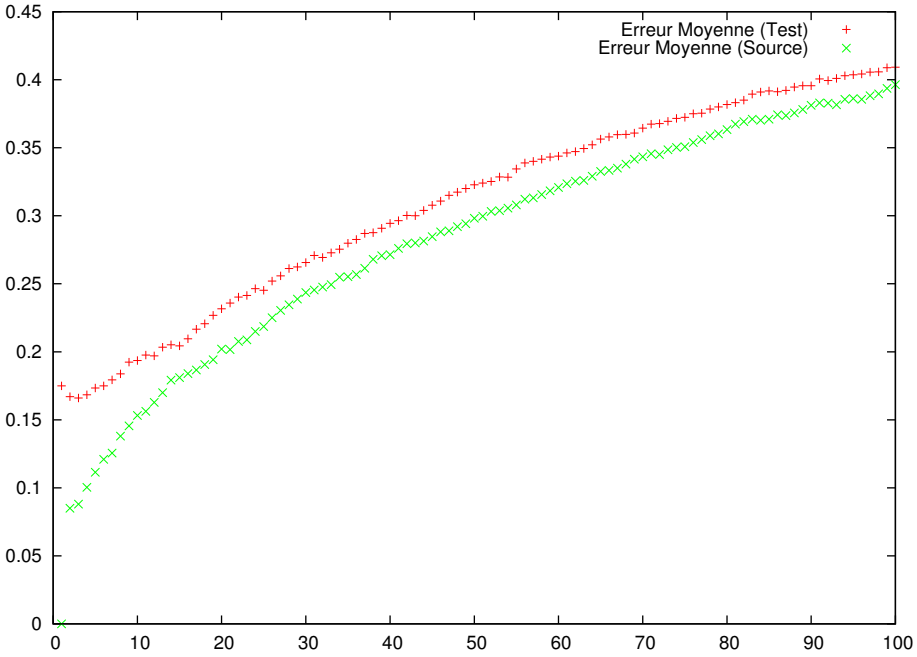

Figure 1: Résultat avec variation k

r2.pdf

Figure 2: Résultat du test avec variation de la taille des ensembles d'apprentisage