De la programmation à la data intelligence : projet : rapport

# Phase 3

Objectif :

## Utiliser les modèles produits à partir du trainset pour classer les différents mouvements du testSet. Ensuite, déterminer le pourcentage de mouvements du testSet correctement classés pour chacune des 6 catégories

## Diagramme d’action :

### evaluationTest

Fonction où les différents mouvements du testSet seront comparés un par un à chacun des 6 modèles qui se trouvent dans pattern.csv. Si un mouvement est bien identifié, ‘estimateClasses’ sera incrémenté de 1 à la cellule qui correspond à ce modèle.Une image contenant texte, capture d’écran, document, Police

Description générée automatiquement

### convertFiletoTable

Fonction qui sert à convertir le fichier pattern.csv en tableau de tableaux, afin de permettre l’utilisation des modèlesUne image contenant texte, Police, ligne, capture d’écran

Description générée automatiquement

### displayResultsForEachClass

onction pour afficher les résultats après avoir classé l'ensemble des mouvements de testSet par rapport aux modèles (pattern.csv).

Éléments affichés :

* Le numéro/type du mouvement
* Le nombre de mouvements qui ont été correctement classés
* Le nombre de mouvements de ce type dans le testSet
* Le pourcentage = (nombre de mouvements correctement classés / nombre de mouvements de ce type)
* À la toute fin : le pourcentage de mouvements correctement classés dans le testSetUne image contenant texte, Police, capture d’écran, ligne

  Description générée automatiquement

## Résultats

Une image contenant texte, Police, capture d’écran, noir

Description générée automatiquement

On peut constater que de manière générale, on peut classer correctement un peu moins d'une fois sur deux un mouvement du testSet en utilisant les patterns.

Ce qui n'est pas vraiment efficace ou suffisant.

# Phase : amélioration

### Amélioration (testée) : remplacer les valeurs aberrantes

### Principe

### Remplacer les valeurs qui dépassent trois fois l'écart type par rapport à la valeur moyenne par une valeur qui correspond davantage à une valeur attendue.

### Valeur essayée : la moyenne

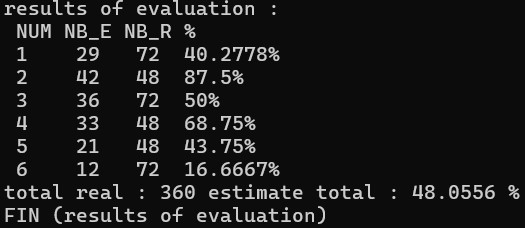
Une image contenant texte, Police, capture d’écran, noir

Description générée automatiquement

* Nous pouvons remarquer une augmentation du pourcentage de correspondance correcte.
* Mais l'effet est assez limité, avec moins de 1 % d'amélioration.

### Explication

Bien que la moyenne soit la valeur qui se rapproche le plus de n'importe quel point qu'on peut attendre dans la courbe d'un mouvement X,Y,Z, cela signifie également qu'elle ne sera presque jamais celle attendue. Elle sera toujours plus grande ou plus petite. Sauf dans les rares cas où la courbe et la moyenne se croisent. Cependant, selon moi et les résultats obtenus, c'est quand même mieux que les aberrations, qui sont également incorrectes mais de manière plus aléatoire que la moyenne, laquelle a un lien avec la valeur attendue.

Valeur essayée : la moyenne des deux valeurs, avant et après l'aberration

## Augmentation de la précision par rapport au remplacement par la moyenne générale.

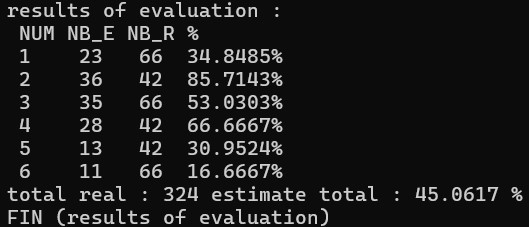
## L'augmentation est de moins de 1 %, ce qui est vraiment minime.

### Amélioration (testée) : une autre façon de former les ensembles d'entraînement et de test.

### Principe :

### Au lieu de prendre les 300 premiers vecteurs d'accélération pour le trainset et les 60 suivants pour le testset, on divise les 360 fichiers en deux paquets. Pour 60 des fichiers, on utilise les 600 premiers vecteurs d'accélération pour le testset, et pour les 300 autres fichiers, on utilise les 600 premiers vecteurs d'accélération dans le trainset.

### Résultat :



* On peut remarquer une diminution du pourcentage de correspondances.
* Ici, le remplacement des aberrations par la moyenne des voisins est également appliqué.
* Pour moi, cela semble être l'approche la plus logique. Sinon, on a des patterns de 600 colonnes dont on n'utilise que les 60 premières, ce qui semble assez inutile, étant donné qu'on ignore la majorité des données.
* Mais je suppose que le fait qu'on ait moins de valeurs pour les patterns fait qu'ils sont plus différents en moyenne par rapport aux mouvements auxquels ils sont liés, ce qui explique que l'on trouve moins de correspondances. Là encore, la différence de résultat est de moins de 2 %.

Amélioration (hypothétique) : différencier par rapport à l'amplitude et à la période.

Les différents types de mouvements se comportent globalement de manière ondulatoire, ce qui nous permettrait de calculer une amplitude et une fréquence moyennes (utilisées comme pattern) que l'on pourrait utiliser pour comparer chaque mouvement du testset.

À partir de là, nous pourrions déterminer à quel mouvement appartient chaque mouvement testé.

## Amélioration (testée) : utiliser une autre variable que l'accélération.

|  |  |
| --- | --- |
| Gravity (les trois) | 20,2216 |
| Rotation rate (les trois) | 40,7202 |
| Attitude (les trois) | 28,5319 |
| taux de rotation x | 41,2742 |
| taux de rotation y | 46,5374 |
| taux de rotation z | 36,5651 |
| attitude\_pitch | 39,8892 |
| attitude\_roll | 19,3906 |
| attitude\_yaw | 21,0526 |
| acceleration\_x | 47,0914 |
| acceleration\_y | 48,4765 |
| **acceleration\_z** | **50,9695** |
| gravity x | 23,2687 |
| gravity y | 34,903 |
| gravity z | 34,349 |

Résultat :

On peut remarquer que de façon générale, ce sont bien les vecteurs d'accélération qui ont le meilleur taux de correspondance. Et parmi les trois, c'est le vecteur d'accélération en z qui nous permet d'avoir légèrement plus de la moitié de bonnes correspondances. Une augmentation de plus de 2,5 %.

Cela nous permet de dire qu'effectivement, les vecteurs d'accélération sont les meilleurs pour obtenir le meilleur taux de correspondance.

Et si l'on voulait en impliquer un second, on pourrait prendre le taux de rotation qui est en seconde position par rapport aux autres.