

Rapport

Contexte

Dans le cours de numérique, on a demandé de construire 3 phrases au total. Voici la tentative d'amélioration de la phrase 3.

Solution de base

La solution construite au cours a été paramètre avec les distances euclidiennes avec la colonne accélération des x, accélération des y et accélération des z plus 600 colonnes dans le trainset et 60 colonnes dans le testset avec les valeurs extrêmes retirées.

Cette solution a obtenu :

23	1	44	0	0	1	23	69	33,33%
5	42	1	0	0	0	42	48	87,50%
13	0	40	3	0	15	40	71	56,34%
0	0	0	4	44	0	4	48	8,33%
0	0	0	12	36	0	36	48	75,00%
47	2	22	0	0	1	1	72	1,39%
						146	356	41,0112%

Une moyenne de résultats de 41%. On vas essayer d'améliorer ca dans ce rapport.

Approche

Pour cette tentative, j'ai eu 7 grands axes d'améliorations listées ci-dessous :

- **Les colonnes prises dans le trainset** : prendre plus les colonnes.
- **Le couplement des colonnes prises** : découpler les colonnes accélérations.
- **Le nombre de colonnes du testset** : augmenter le nombre de colonnes.
- **Les critères de valeurs de extrêmes** : aucune valeur extrême n'est extrême « aucun-extrême », les valeurs sont centrées réduites avant d'être vérifiées « réduit-extrême », les valeurs sont « bornées » par des valeurs fixes « bornée-extrêmes ».
- **La méthode de filtrage de donnée** : les valeurs extrêmes sont remplacées par la moyenne « vers-moyenne », s'il y a au moins une valeur extrême sur la ligne alors cette ligne est complément ignorée « un-défaut ».
- **La méthode d'évaluation** : la distance entre 2 droites « distance euclidienne », comparer l'amplitude et la fréquence du modèle « cyclique », comparer l'histogramme du modèle « histogramme ».
- **Le calcul des moyennes et des écart-types pour chaque colonne** : pour la méthode filtrage de donnée « vers-moyenne » où on remplace les valeurs extrêmes par la moyenne. C'est la moyennes et écart-types devrait être vérifiées mais si on rajoutait d'autre colonnes il faudrait calculer nous même les moyennes et écart-types de ces nouvelles colonnes.

Développement des axes

Pour ces modifications, je vais partir de la solution de base :

Modification de la solution de base « combinaison de colonne » :

[découpler les colonnes accélérations].

En sachant qu'il y a 12 colonnes, l'ordre ne compte pas, qu'il n'y pas de remise et qu'on peut prendre 1 comme 12 colonnes Il y a $\sum_{i=1}^{12} C_{12}^i = 4096$ possibilités. En fait 4096, parce qu'on peut visualiser les combinaisons comme l'incrémentation d'un chiffre binaire à 12 bits où chaque bit représenterais la présence de la colonne ou non et l'exploration de chaque combinaison comme l'incrémentacion.

Modification de la solution de base « rocher-caillou » :

[aucune valeur extrême n'est extrême], [les valeurs sont centrées réduites avant d'être vérifiées], [les valeurs sont « bornées » par des valeurs fixes].

J'ai trouver 3 manières qu'on pourrait considirer une valeur comme extremes. Ca fait 3 possibilités.

Modification de la solution de base « scan » :

[augmenter le nombre de colonnes] dans le testset.

En fessant varier la proportion de colonnes de testset en fonction du trainset, on peut voir le nombre de colonnes optimales. Je vais prendre 1, 0.1 et 0.5 pour ces proportions.

Modification de la solution de base « scan » :

[augmenter le nombre de colonnes] dans le trainset

On va aussi regarder quel serait le nombre de colonne optimal pour le trainset, on va prendre x1, x2, x3 plus de colonnes.

Les autre idées ne seront pas prises parce que je manque de temps.

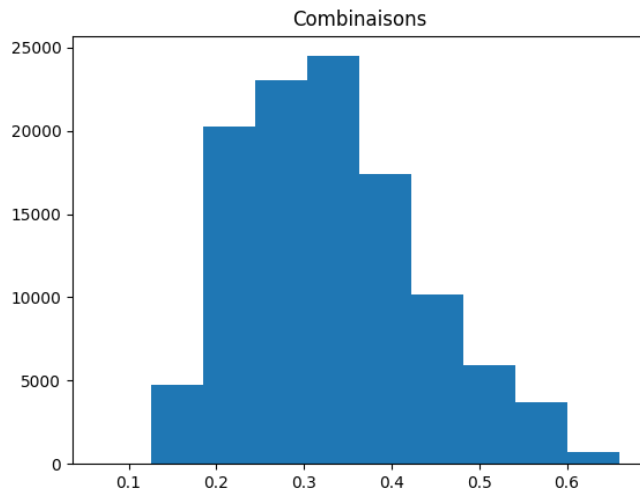
Methode d'optimations de la performance

Avec tous ces possibilités, on a $4096 * 3 * 3 * 3 = 110\,592$ combinaisons possibles. Avec une combinaison qui prend en moyenne 3 secondes, ca prendra 3 jours 20 heures.

Evidement tous les résultats ont été obtenu sur un serveur Linux compilé avec gcc.

Sur 110 592 combinaisons voici ce qu'on a :

Voici un histogramme qui represente le nombre de combinaisons en fonction de la moyenne des résultats sur tous les mouvements :



On peut observer la plupart des combinaisons sont entre 10% et 40% de précision et que le minimum est de 10% et le maximum est de 66% de précision

Il y a 6 meilleurs combinaisons :

Combinaisons de colonnes	Nombre de colonnes du trainset	Nombre de colonnes du testset	Fonction d'extrême	Résultat
011101010000	1800	180	aucun-extrême	66,05%
011101100000	1800	180	bornée-extrêmes	66,05%
110101010000	1800	180	aucun-extrême	66,05%
111000010000	1800	180	aucun-extrême	66,05%
111101010000	1800	180	aucun-extrême	66,05%
111101100000	1800	180	bornée-extrêmes	66,05%

Les 12 colonnes ont été encodées sous formes d'un chiffre binaires :

Valeur de la colonne	Nom de la colonne
0	ATTITUDE ROLL
1	ATTITUDE PITCH
2	ATTITUDE YAW
4	GRAVITY X
8	GRAVITY Y
16	GRAVITY Z
32	ROTATION X
64	ROTATION Y
128	ROTATION Z
256	ACCELERATION X
512	ACCELERATION Y
1024	ACCELERATION Z

Si on regarde le tableau des meilleures combinaisons, on pourrait croire que 1800 colonnes pour le trainset et 180 colonnes pour testset est meilleurs que les autres variations.

Mais si on regarde le minimum, le maximum, le maximum des résultats obtenu en fonctions du nombre de colonnes du trainset et du nombre des colonnes du testset, on obtient :

Colonnes trainset	Colonnes testset	Min	Max	Moyenne
600	60	15%	62%	33%
	300	15%	62%	36%
	600	18%	60%	36%
1200	120	13%	63%	34%
	600	10%	60%	33%
	1200	12%	56%	31%
1800	180	9%	66%	34%
	900	6%	60%	32%
	1800	12%	52%	26%

Ce qui nous intéresse, c'est la colonne des maximums. On peut voir que généralement pour chaque colonne du trainset et que les colonnes du testset augmente, au plus le maximum diminue mais que le minimum augmente. Peut-être que comparer 1800 colonnes du trainset avec 3 colonnes du testset sauterait les résultats au plafond ?

Si on observe le minimum, le maximum et la moyenne des résultats obtenus en fonctions des fonctions d'extrêmes :

Fonction extrêmes	Min	Max	Moyenne
bornée-extrêmes	6%	66%	37, 4%
réduit-extrême	12%	56%	24%
aucun-extrême	13%	66%	37, 5%

La fonction « réduit-extrême » se démarque parce que son minimum, maximum et moyenne sont toutes inférieurs à l'autre techniques. Sinon, les 2 autres sont presque équivalentes. Il y a juste le minimum de « bornée-extrêmes » qui est inférieur à « aucun-extrême ». Donc cette dernière est la meilleure à choisir.

Conclusion