

# Dissimilarités entre jeux de données

-

## Analyse dimensionnelle des résultats

15 février 2017

On présente ici en détail l'analyse complète des différents facteurs identifiés.

### 1 Facteurs secondaires

On qualifie de secondaires les facteurs pouvant influencer sur la performance au méta-niveau mais ne faisant pas partie intégrante des dissimilarités. On étudie ici leur impact sur la performance au méta-niveau et leurs relations avec les différentes dissimilarités.

**Nombre de voisins considérés** L'algorithme 2 estime la performance des classifieurs sur un nouveau jeu de données  $D$  selon leur performance sur les  $k$  plus proche voisins de  $D$ , au sens de la dissimilarité évaluée. Ce nombre  $k$  prend ici des valeurs communes pour des ensembles de l'ordre de la centaine [1] :  $k \in \{3, 5, 10\}$ . Pour étudier l'impact de ce facteur  $k$ , on observe le comportement des différentes dissimilarités pour chaque valeur de  $k$ , en moyennant selon les autres dimensions. Cette performance moyenne au méta-niveau est présentée en Figure 1.



Pour ce faire, on fait appel au test *post-hoc* de Nemenyi. Ce dernier permet de comparer les échantillons deux à deux, déterminant lesquels sont significativement différents, tout en contrôlant le risque global d'erreur type 1. En effet, comparer de nombreux échantillons rend exponentiel le risque de commettre au moins une erreur de type 1 (de valider une différence infondée). Le test de Nemenyi permet de maîtriser ce risque quel que soit le nombre d'échantillon. Dans nos expériences, on contraindra ce risque à la valeur courante de 0.05, ce qui signifie que chaque test de Nemenyi effectué a au plus 5% de chance de discriminer deux échantillons qui n'étaient pas réellement discernables. Le test résulte ainsi en une *différence critique CD*, représentant la différence nécessaire entre le *rang moyen* de deux échantillons échantillon (statistique produite par le test de Friedman) pour pouvoir les considérer significativement différents. Ceci est représenté en Figure 2, où l'on classe les différents échantillons par rang moyen. Ceux trop proches pour pouvoir être considérés significativement différents sont liés entre eux. Ici, cela signifie que l'échantillon  $k = 3$  est significativement moins bon que les échantillons  $k = 10$  et  $k = 5$ , mais que ces derniers ne sont pas suffisamment éloignés pour être jugés significativement différents (sans prendre le risque d'avoir plus de 5% de chances de se tromper). La valeur indiquée à côté de l'identifiant de l'échantillon est sa performance moyenne, qui permet de constater les écarts de moyenne parfois très faibles entre échantillons jugés différents.

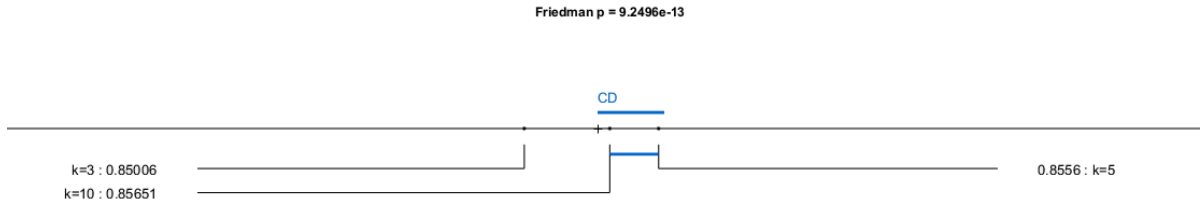


FIGURE 2 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes valeurs de  $k$ .

Les résultats du test de Nemenyi en Figure 2 valident donc notre première tendance : choisir  $k = 3$  mène à des performances significativement inférieure. Pour valider la seconde, on répète les tests de Friedman et Nemenyi en se limitant cette fois aux dissimilarités utilisant *DMFf\_full*. Les résultats, présentés en Figure 3, montrent bien que  $k = 10$  y est significativement meilleur que  $k = 5$ , lui-même significativement meilleur que  $k = 3$ .

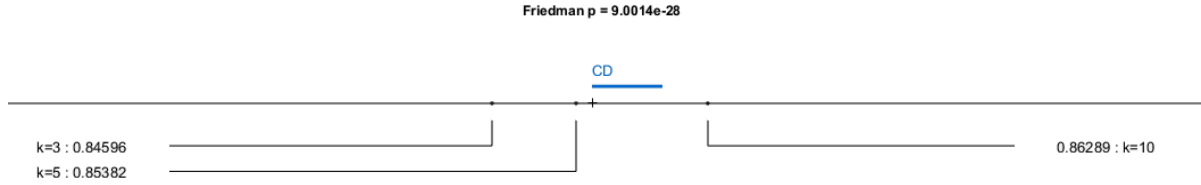


FIGURE 3 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes valeurs de  $k$ , pour les dissimilarités utilisant *DMFf\_full*.

Ces différents résultats nous permettent d'écarter  $k = 3$ , et pointe vers l'utilisation de  $k = 10$ , en particulier en conjonction avec *DMFf\_full*. Ceci pourrait indiquer la nécessité de considérer davantage de voisins si l'on souhaite exploiter de grands ensembles de méta-attributs des attributs.

Cette méthodologie de test, par étude des distributions puis validation par Friedman et Nemenyi, guidera ainsi notre étude des résultats, et sera reproduite selon nos différentes dimensions pour évaluer l'impact de chaque facteur identifié plus tôt.

**Méthode d'estimation de performance d'un classifieur selon celle de ses voisins** Pour estimer la performance d'un classifieur sur un nouveau jeu de données  $D$ , l'algorithme 2 peut utiliser les deux fonctions proposées : soit "**mean**", la performance moyenne du classifieur sur les  $k$  plus proches voisins de  $D$ , soit "**weighted**", moyenne cette fois pondérée par la dissimilarité entre  $D$  et ses voisins. On présente en Figure 4 la moyenne de performance au méta-niveau des dissimilarités selon la méthode d'estimation utilisée.

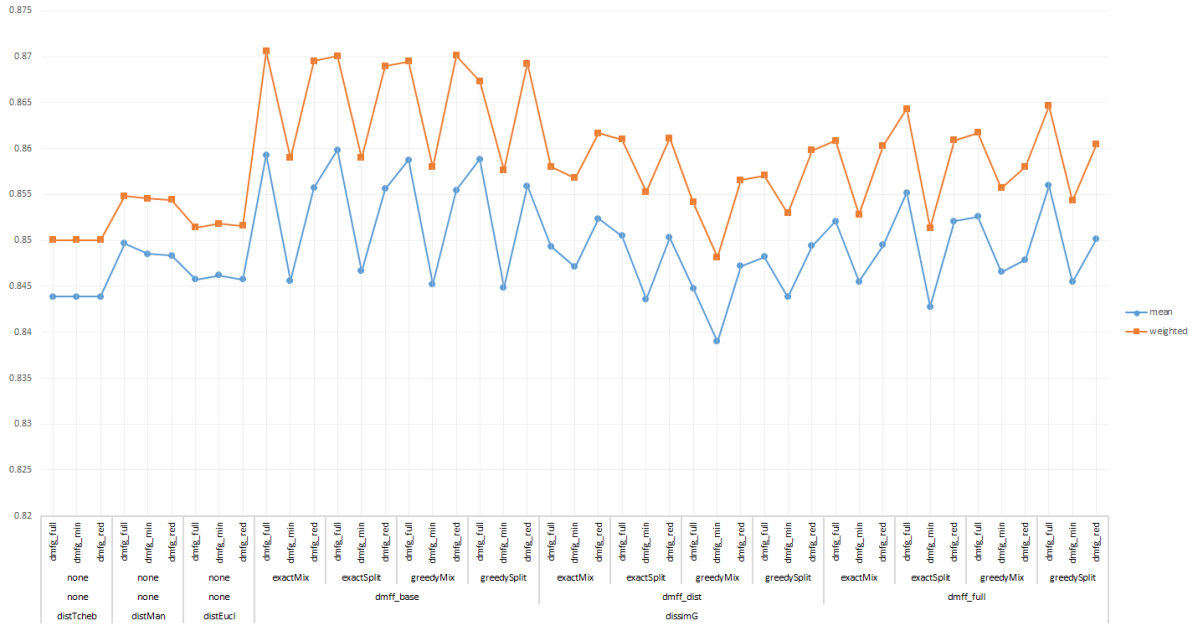


FIGURE 4 – Moyenne des performances au méta-niveau selon la méthode d’estimation utilisée.

On remarque une importante similitude entre ces courbes, avec une différence quasi-constante à l’avantage de la méthode **”weighted”**. Comme précédemment on valide cette tendance par un test de Friedman avec post-hoc Nemenyi en Figure 5.

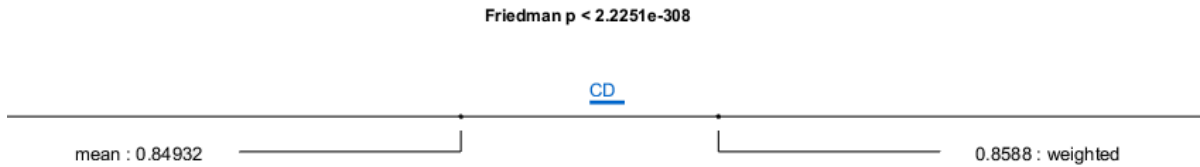


FIGURE 5 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes méthodes d’estimation.

Le test valide bien la supériorité globale de la méthode **”weighted”**. Ce résultat conforte l’intérêt perçu des dissimilarités, et leur utilité pour l’estimation de performance au niveau de base.

**Critères de performance de base** Notre critère de performance au méta-niveau se définit par rapport à un critère de performance de base que l’on cherche à optimiser par le travail au méta-niveau. Dans cette expérience, le problème de base est celui de la classification supervisée, pour lequel on a retenu quatre

critères de performance à la sémantique distincte. On présente en Figure 6 la moyenne de performance au méta-niveau des dissimilarités et de la *baseline* selon le critères de performance de base utilisé (la *baseline* apparaît à gauche, suivie des différentes dissimilarités).

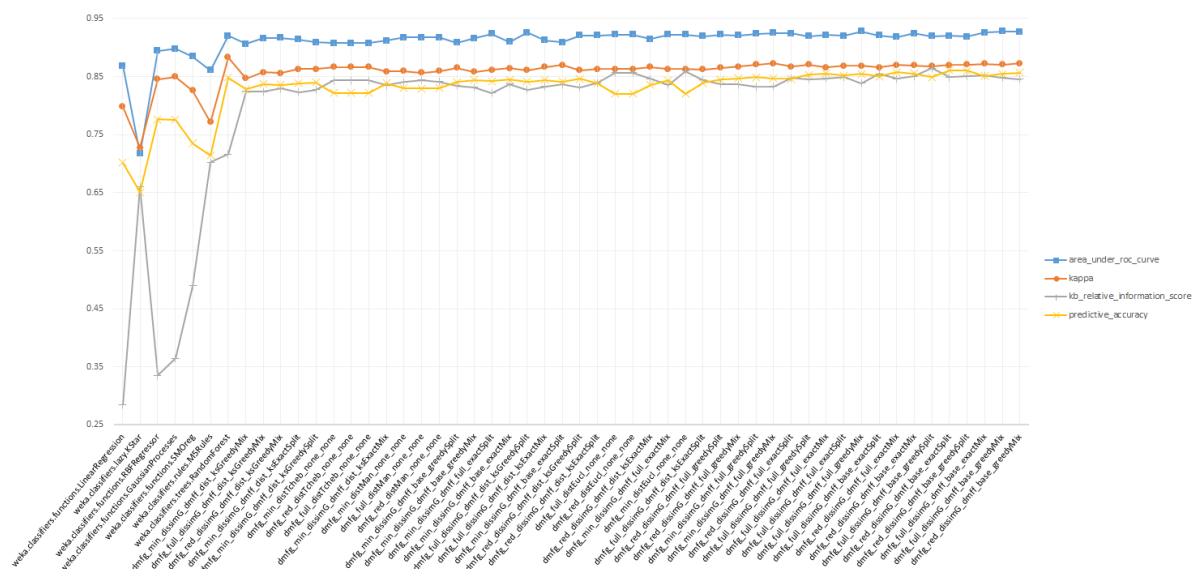


FIGURE 6 – Moyenne des performances au méta-niveau selon le critères de performance de base utilisé.

La performance au méta-niveau semble donc meilleure pour optimiser l'aire sous la courbe de *ROC* que pour le *kappa* de Cohen, lui-même meilleur que la précision et l'*information score* de Kononenko et Bratko. On ne peut pas réellement discriminer entre ces deux derniers, mais l'*information score* semble devancer la précision pour les dissimilarités utilisant de simples distances. On vérifie ces tendances respectivement en Figures 7 et 8.

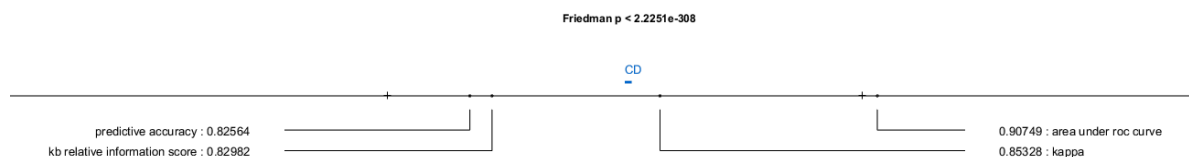


FIGURE 7 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents critères de performance de base.



FIGURE 8 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents critères de performance de base, pour les dissimilarités utilisant des distances simples.

Les Figures 7 et 8 nous assurent donc de l'ordonnancement global des critères par les performances au méta-niveau qu'ils induisent, en effet plus marqué pour les dissimilarités utilisant des distances simples.

Une étude plus approfondie de la relation critère - performance se concentrant sur les dissimilarités semble faire apparaître une inversion de tendance entre l'*information score* et la précision lorsque l'on utilise *DMFg\_full* (voir Figure 9). On tente de valider cette tendance en Figure 10.

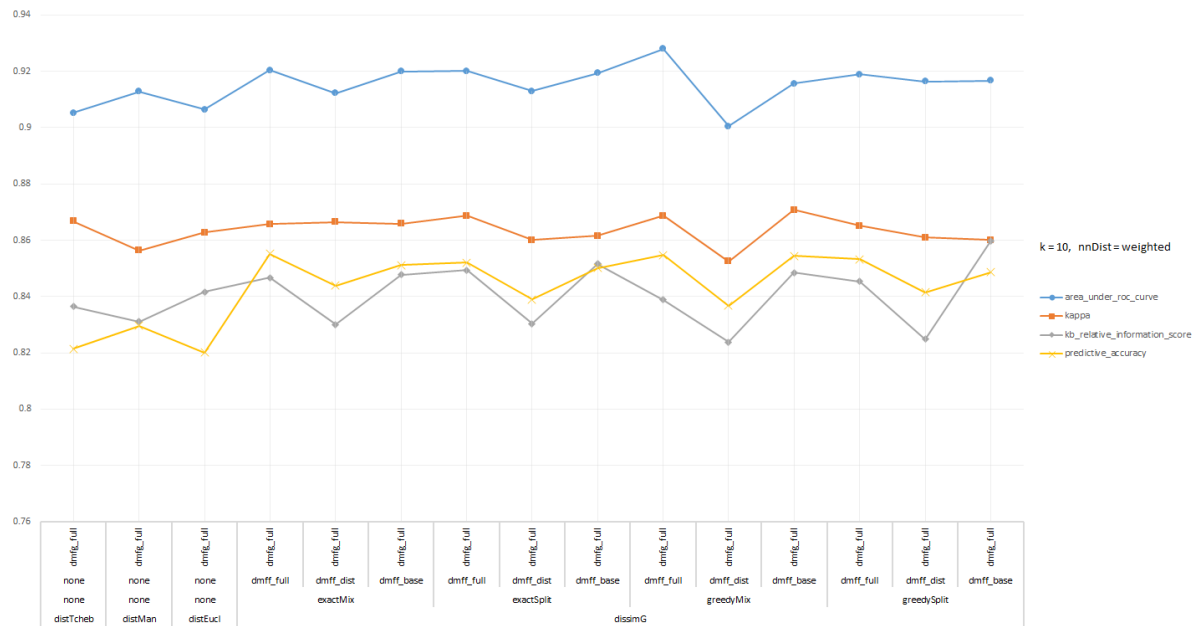


FIGURE 9 – Moyenne des performances au méta-niveau selon le critères de performance de base utilisé, pour les dissimilarités utilisant *DMFg\_full*.



FIGURE 10 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents critères de performance de base, pour les dissimilarités utilisant *DMFg-full*.

Le test de Nemenyi en Figure 10 ne parvient pas à déceler d'inversion de tendance. L'ordonnancement en performance induite des différents critères semble donc valide dans une portion significative des cas de test, et aucun contre-exemple significatif n'a pu être trouvé. Tous nos éléments pointent donc vers l'aire sous la courbe de ROC, dont l'utilisation comme critère de performance de base semble mener aux meilleures performances au méta-niveau, et qui est généralement reconnu comme un bon critère de performance en classification.

**Nombre de recommandations** Comme indiqué précédemment, le nombre  $n$  de recommandations effectués par les algorithmes 2 et 3 prend ses valeurs dans l'ensemble  $\{1, 3, 5\}$ . L'introduction de ce facteur avait pour objectif de "lisser" les résultats (réduire leur variance). On présente tout d'abord en Figure 11 la moyenne de performance au méta-niveau des dissimilarités et de la *baseline* selon le nombre de recommandations effectuées.

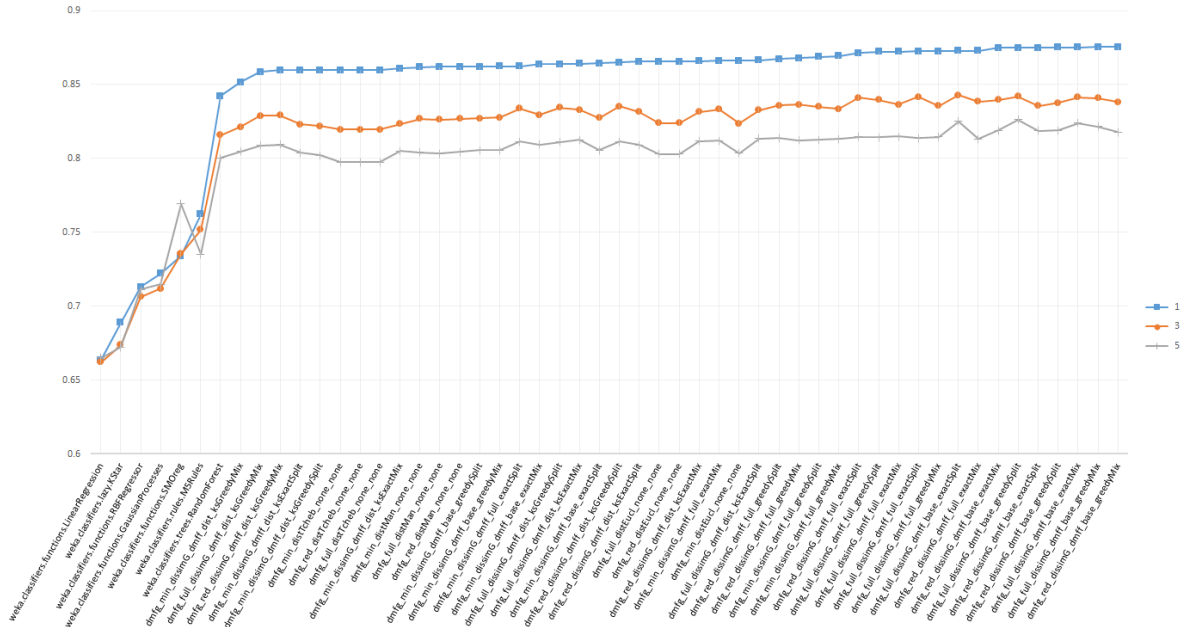


FIGURE 11 – Moyenne des performances au méta-niveau selon le nombre de recommandations.

Aucune tendance ne se distingue sur les algorithmes de la *baseline*, mais



sur les dissimilarités, il semble qu'augmenter le nombre de recommandations impacte négativement les performances. On vérifie cela en Figure 12.

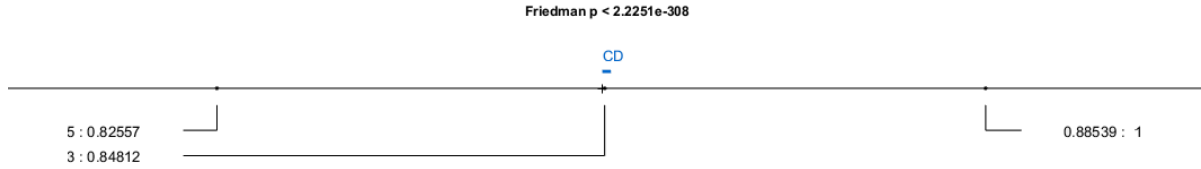


FIGURE 12 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes valeurs du nombre de recommandations.

Augmenter le nombre de recommandations impacte donc bien négativement les performances au méta-niveau. Ce compromis était attendu, le but étant de contrôler la variance des résultats. On étudie donc en Figure 13 la variance des performances au méta-niveau selon le nombre de recommandations.

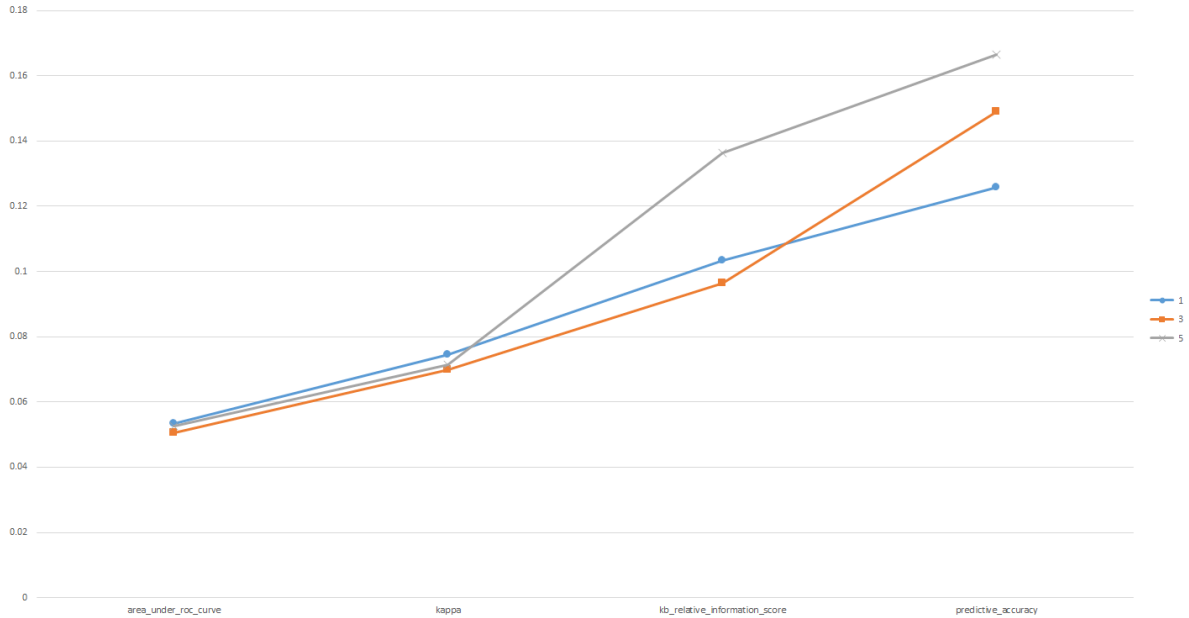


FIGURE 13 – Variance des performances au méta-niveau par critère selon le nombre de recommandations.

On observe donc les variances les plus basses sur le critère d'*AUC*, sans que le nombre de recommandations ne semble l'impacter. Le seul critère où la multiplication des recommandations a l'effet escompté et limite la variance, est celui de précision, mais les valeurs y restent peu intéressantes. Cette étude de variance conforte donc notre choix du critère d'aire sous la courbe de *ROC*, mais rend apparent le peu d'intérêt de la multiplication des recommandations, du point de vue des performances. D'autres contraintes du domaine de la recommandation peuvent justifier la multiplication des recommandations, mais la qualité des solutions trouvées semble suffisamment élevée pour que cela entraîne une perte significative de performance. Ce *tradeoff* sera bien sûr à considérer, mais

dépasse le cadre de la présente étude.

**Espace optimal** Ces études préliminaires nous ont permis de gagner une meilleure compréhension de l’impact des facteurs secondaires, et en particulier de déterminer un ”espace optimal”, ensemble de valeurs des facteurs secondaires maximisant la performance au méta-niveau. Dans certaines des observations suivantes, on se placera dans cet espace optimal pour analyser le comportement des dissimilarités dans ce qui serait le plus proche d’un futur cas d’application réel. Comme espace optimal, on retient donc les valeurs de  $k = 10$ ,  $\text{nnDist}=\text{weighted}$ ,  $\text{criterion}=\text{AUC}$ , et  $n = 1$ . D’autre part, on prendra toujours la performance au méta-niveau pour  $n = 1$  (sauf mention contraire explicite).

## 2 Facteurs primaires

On qualifie de facteurs primaires les éléments fonctionnels variables des dissimilarités. On étudiera ici leur impact sur les performances au méta-niveau et les potentielles interactions entre eux.

**Méta-attributs généraux** Nos ensembles de méta-attributs généraux diffèrent par le nombre de méta-attributs de *landmarking* employés. *DMFg\_full* en contient 27 (pour 62 autres méta-attributs), *DMFg\_red* seulement 9, et *DMFg\_min* aucun. On présente en Figure 14 les performances moyennes au méta-niveau pour les dissimilarités utilisant ces ensembles, selon leurs autres facteurs primaires. L’intérêt des *landmarkers* y semble avéré, et on le vérifie en Figure 15.

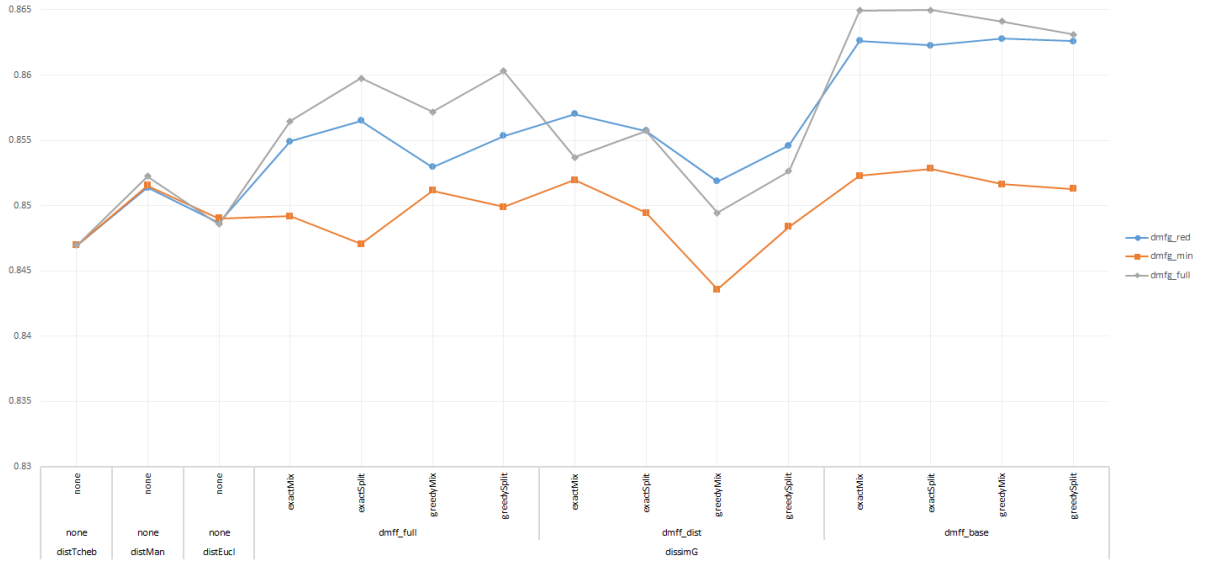


FIGURE 14 – Moyenne des performances au méta-niveau selon l’ensemble de méta-attributs généraux utilisé.

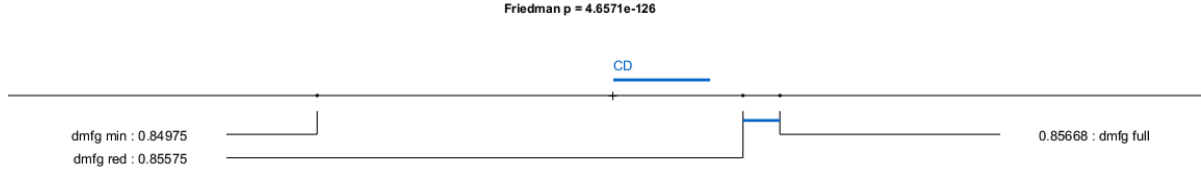


FIGURE 15 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents ensembles de méta-attributs généraux utilisés.

*DMFg\_full* et *DMFg\_red* distancent significativement *DMFg\_min*, validant l'intérêt des méta-attributs de *landmarking*, mais aucune différence significative n'est établie entre eux. On peut rejoindre ici des travaux de sélection d'attributs insistant sur l'importance de la diversité au sein des attributs [3]. Il pourrait ainsi être intéressant d'étudier l'impact de la *diversité* des *landmarkers* et autres méta-attributs généraux choisis sur la performance au méta-niveau.

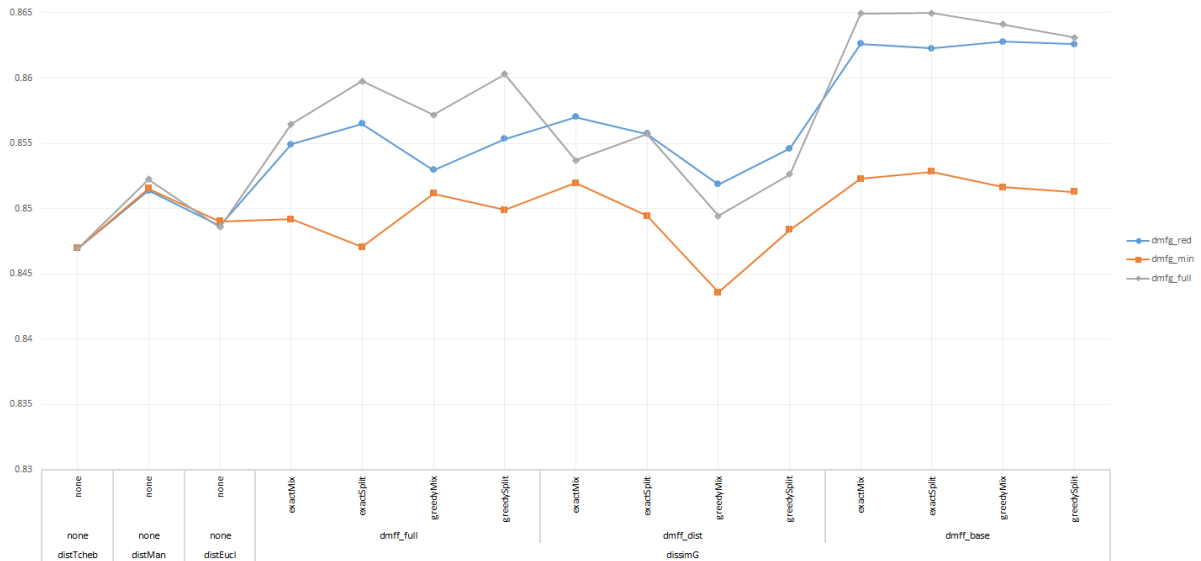


FIGURE 16 – Moyenne des performances au méta-niveau selon l'ensemble de méta-attributs généraux utilisé, sur l'espace optimal.

Si l'on restreint l'étude à l'espace optimal (Figure 16), on ne décèle plus de tendance particulière. De même, le test statistique (Figure 17) révèle uniquement la domination de *DMFg\_red* sur *DMFg\_min*, insistant sur l'intérêt de la présence de *landmarkers*, nonobstant leur nombre.

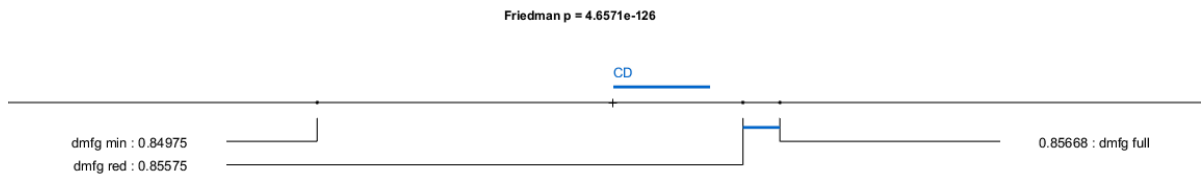


FIGURE 17 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents ensembles de méta-attributs généraux utilisés, sur l'espace optimal.

**Dissimilarité sur les méta-attributs généraux** Les dissimilarités sur les méta-attributs généraux considérées sont un panel de distances adjoint de la dissimilarité normalisée par la borne supérieure. On observe donc en Figure 18 comment se comporte la dissimilarité normalisée par la borne supérieure par rapport aux distances classiques dans notre espace optimal.

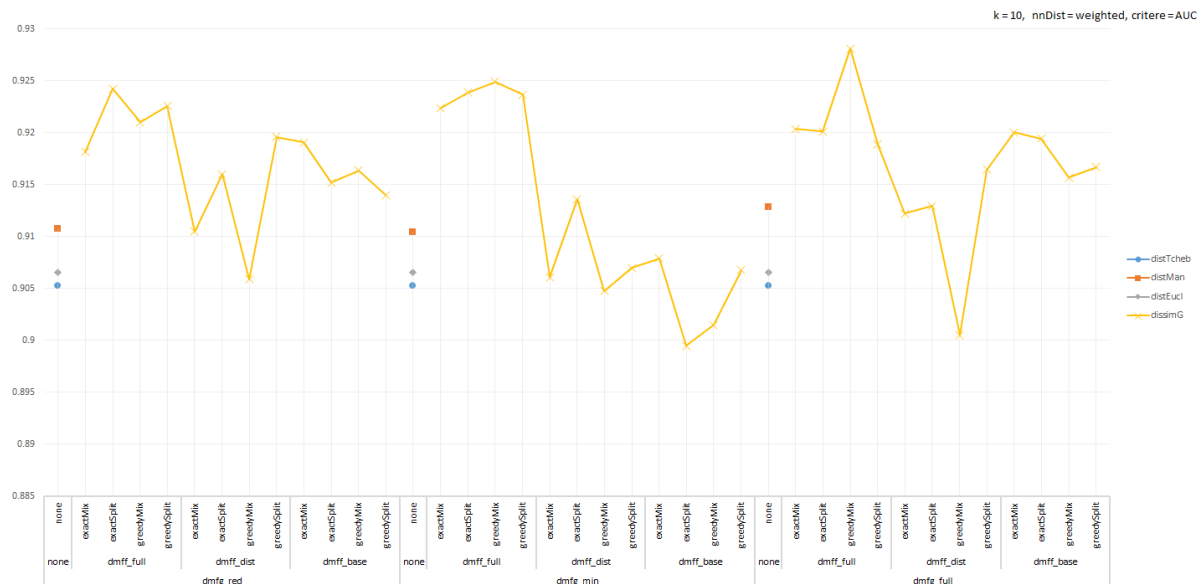


FIGURE 18 – Moyenne des performances au méta-niveau selon la dissimilarité sur les méta-attributs généraux.

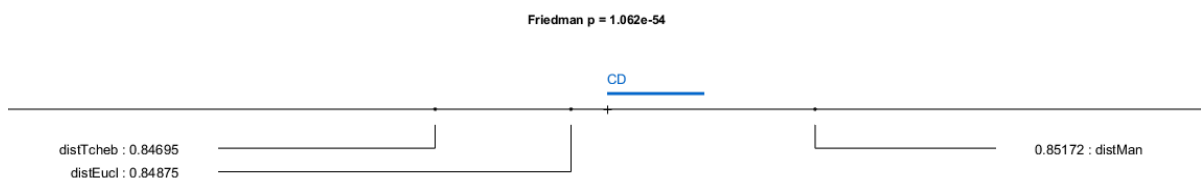


FIGURE 19 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes dissimilarité sur les méta-attributs généraux.

L'asymétrie du problème rend impossible la comparaison statistique de *dis-*

$simG$  et des distances (on a besoin de mesures appariées, différant selon une seule dimension), mais on peut observer en Figure 18 la dominance de  $dissimG$  dans une majorité de cas. De plus, une comparaison entre elles des distances confirme la dominance de la distance de Manhattan sur les deux autres (voir Figure 19), ce qui conforte son choix comme brique de base de la dissimilarité normalisée par la borne supérieure.

Le test de Mann-Whitney, conçu pour permettre la comparaison d'échantillons de taille différente, est ici une alternative valide au test de Friedman pour comparer les distances à la dissimilarité. On présente en Table 1 les résultats de tests de Mann-Whitney pour l'hypothèse  $H_0$  : la probabilité qu'une observation de l'échantillon " $G=dissimG$ " soit supérieure à une observation de l'échantillon " $G \neq dissimG$ " est égale à la probabilité qu'une observation de l'échantillon " $G \neq dissimG$ " soit supérieure à une observation de l'échantillon " $G=dissimG$ " (et de même avec  $G=distMan$ ).

TABLE 1 – Résultats des tests de Mann-Whitney comparant les distances à la dissimilarité.

Échantillon	Taille	Moyenne	$p$ -value
$G \neq dissimG$	85320	0.8491	2.12E-34
$G=dissimG$	341280	0.8553	
$G=distMan$	28440	0.8517	4.97E-04
$G=dissimG$	341280	0.8553	

On peut donc dans les deux cas valider la supériorité de  $dissimG$  sur les distances, bien qu'elle soit en grande partie une distance de Manhattan. En particulier, la dominance de  $dissimG$  sur la distance de Manhattan assure de l'intérêt de la normalisation par la borne supérieure et de la prise en compte des valeurs de méta-attributs manquantes par dissimilarité à l'absence de valeur.

**Méta-attributs des attributs** Nos ensembles de méta-attributs des attributs diffèrent par l'adjonction de mesures normalisées.  $DMFf\_base$  comprend un ensemble de méta-attributs adaptés aux attributs numériques et nominaux, auquel  $DMFf\_full$  ajoute quelques méta-attributs normalisés par le nombre d'instances.  $DMFf\_dist$ , l'ensemble des distributions des attributs, ne représente pas à proprement parler un ensemble de méta-attributs des attributs, mais peut être considéré comme tel par la dissimilarité  $\delta_{KS}^\sigma$  utilisant le test de Kolmogorov-Smirnov, comme décrit précédemment.

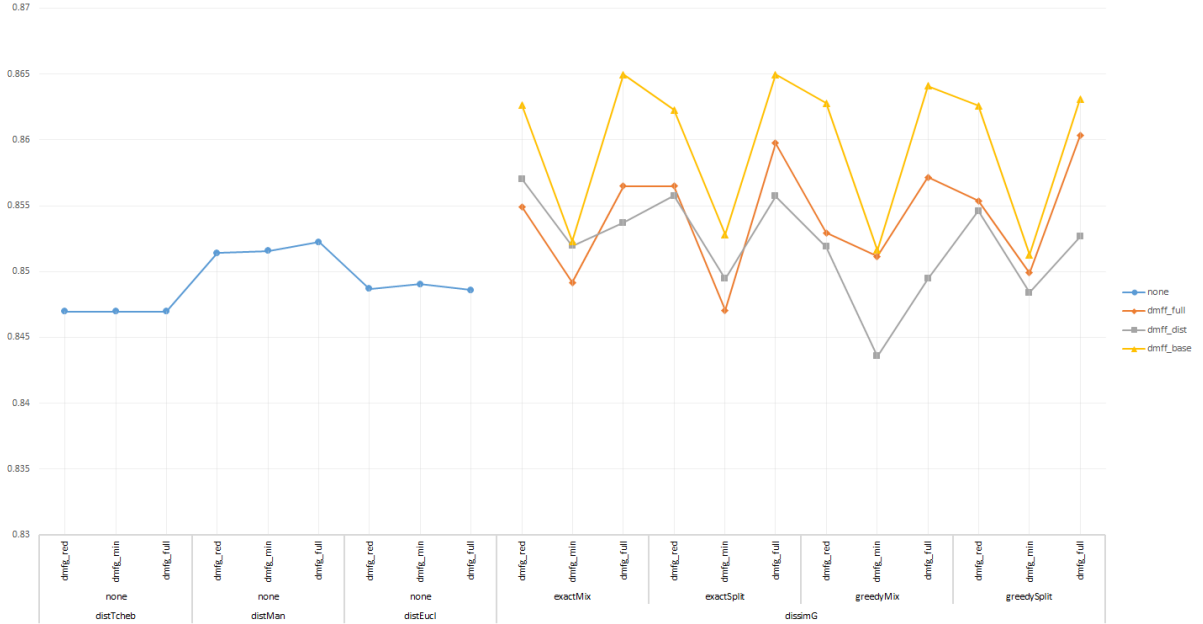


FIGURE 20 – Moyenne des performances au méta-niveau selon l'ensemble de méta-attributs des attributs utilisé.

On présente en Figure 20 les performances moyennes au méta-niveau pour les dissimilarités utilisant ces ensembles, selon leurs autres facteurs primaires. Comme précédemment, les différents ensembles surclassent la baseline (*none*), mais l'asymétrie ne permet pas de le valider par le test de Friedman. En revanche, on peut observer une certaine tendance de *DMFf\_base* à mieux se comporter que les autres ensembles, ce que l'on vérifie en Figure 21.

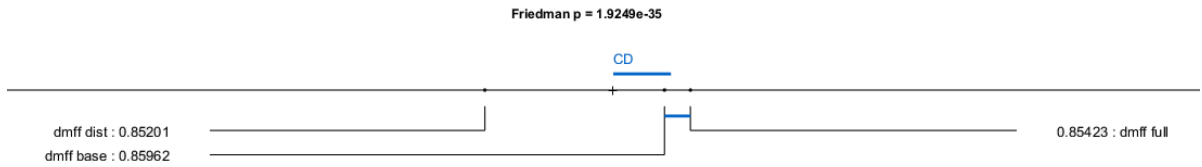


FIGURE 21 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents ensemble de méta-attributs des attributs.

Le résultat observé ne coïncide pas complètement avec la tendance attendue : le test ne permet pas de conclure à une différence significative entre *DMFf\_base* et *DMFf\_full*. En revanche il établit que tous deux sont significativement supérieurs à *DMFf\_dist*. On se place ensuite dans le cas particulier des valeurs optimales de facteurs secondaires isolées dans la section précédente ( $k = 10$ ,  $\text{nnDist}=\text{weighted}$ ,  $\text{criterion}=\text{AUC}$ ,  $n = 1$ ), pour répéter l'observation (Figure 22). *DMFf\_full* semble alors clairement dominer, ce que l'on va chercher à confirmer avec le test en Figure 23.



On présente en Table 2 les résultats de ce test sur l'espace complet, puis en se limitant au critère d'aire sous la courbe, et enfin en se limitant à l'espace optimal.

TABLE 2 – Résultats des tests de Mann-Whitney comparant les dissimilarités proposées à celles de la baseline.

Échantillon	Taille	Moyenne	<i>p-value</i>
DMFf= <i>none</i>	85320	0.8491	2.12E-34
DMFf≠ <i>none</i>	341280	0.8553	
DMFf= <i>none</i> , criterion= <i>AUC</i>	21330	0.9058	1.04E-51
DMFf≠ <i>none</i> , criterion= <i>AUC</i>	85320	0.9079	
DMFf= <i>none</i> , criterion= <i>AUC</i> , $k = 10$ , nnDist= <i>weighted</i>	3555	0.9077	9.53E-07
DMFf≠ <i>none</i> , criterion= <i>AUC</i> , $k = 10$ , nnDist= <i>weighted</i>	14220	0.9152	

On constate bien dans tous les cas une *p-value* suffisante pour écarter  $H_0$  : les échantillons diffèrent donc significativement. La performance moyenne des dissimilarités utilisant des méta-attributs des attributs étant toujours supérieure à celles de la *baseline*, on peut bien conclure à leur dominance. L'augmentation de la *p-value* sur l'espace optimal était attendue, dans la mesure où la réduction de la taille des échantillons réduits d'autant la puissance du test. Mais on peut remarquer que se limiter au critère d'aire sous la courbe permet au contraire de la réduire davantage. Ceci met à nouveau en valeur l'adéquation des dissimilarités proposées avec ce critère, leur avantage y étant plus affirmé malgré la perte de puissance du test.

**Dissimilarité sur les attributs** Les dissimilarités sur les méta-attributs des attributs diffèrent par le *mapping*  $\sigma$  qu'elles emploient. Les distinctions principales sont entre les *mappings* "*greedy*" et "*exact*" utilisant respectivement un algorithme glouton et exact pour l'appariement, et entre les "*split*" et "*mix*" considérant séparément ou non les attributs numériques et nominaux.

On présente en Figure 24 les performances moyennes au méta-niveau pour les dissimilarités construite sur ces différents *mappings*. Une simple observation ne révèle pas de tendance forte, et seul un test statistique (Figure 25) révèle la dominance de l'*exactSplit* sur le *greedyMix*. Ajoutant à cela les temps d'exécution présentés précédemment, on pourra favoriser l'utilisation de *mappings* "*split*", environ 1.5 fois plus coûteux que le *greedyMix* pour le *greedySplit*, et 2 fois plus coûteux pour l'*exactSplit*.



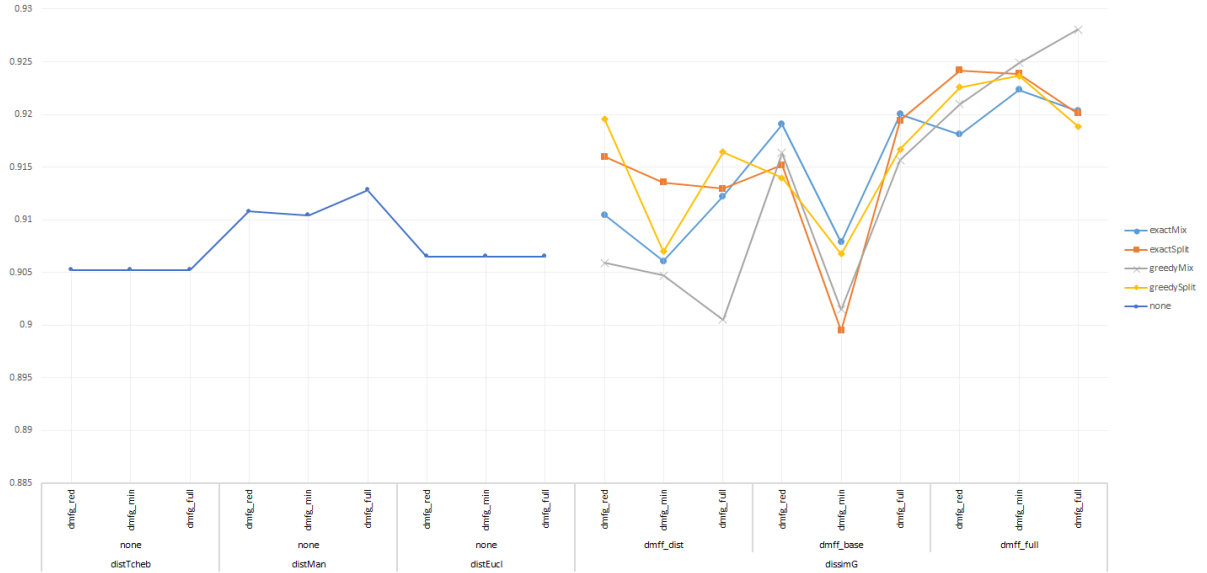


FIGURE 24 – Moyenne des performances au méta-niveau selon la dissimilarité sur les méta-attributs des attributs.

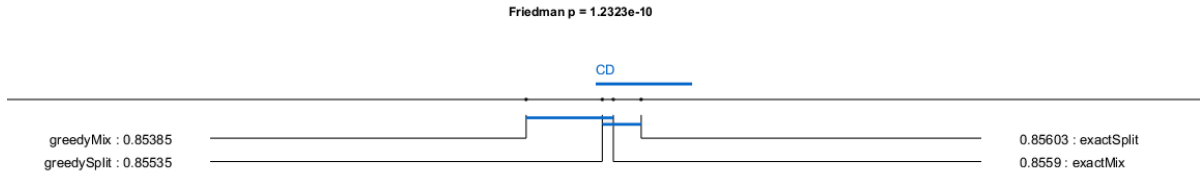


FIGURE 25 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes dissimilarités sur les méta-attributs des attributs.

On vérifie en Figure 26 si ce résultat varie lorsque l'on considère uniquement les dissimilarités  $\delta_{KS}^\sigma$  utilisant le test de Kolmogorov-Smirnov au lieu de méta-attributs des attributs. On arrive cette fois à établir la supériorité des deux méthodes "exact" sur le *greedyMix*, ce qui est très raisonnable dans la mesure où  $\delta_{KS}^\sigma$  n'utilise qu'une seule valeur (le résultat du *KS-test*) pour discriminer, et serait donc plus sensible aux variations "aléatoires" induites par le *mapping* "greedy".

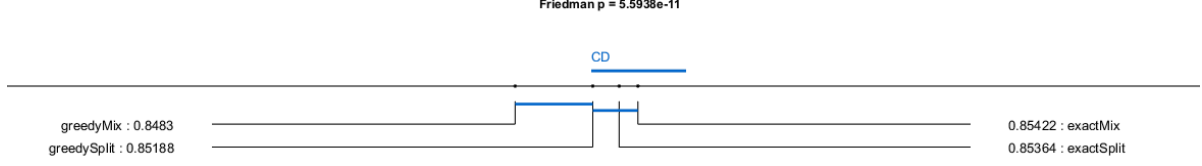


FIGURE 26 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes dissimilarités sur les méta-attributs des attributs, et utilisant  $\delta_{KS}^\sigma$ .

Le résultat précédent peut laisser penser que le choix du *mapping* a plus d'influence sur les dissimilarités  $\delta_{KS}^\sigma$ , ce que l'on peut tenter de confirmer par un test sur les dissimilarités  $\delta_F^\sigma$  (c'est à dire n'utilisant pas le *KS-test*). Leur nombre supérieur confère au test en Figure 27 une puissance supérieure, ce qui devrait permettre de trouver plus facilement les différences significatives. Or, le test ne parvient à mettre en évidence aucune différence significative, ce qui nous confirme bien une sensibilité supérieure de  $\delta_{KS}^\sigma$  au choix du *mapping*.

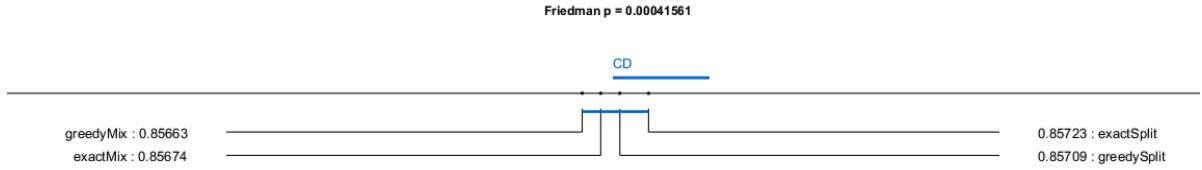


FIGURE 27 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différentes dissimilarités sur les méta-attributs des attributs, et n'utilisant pas  $\delta_{KS}^\sigma$ .

Ces résultats caractérisent donc un léger avantage à la fois en performance et en stabilité de  $\delta_F^\sigma$  par rapport à  $\delta_{KS}^\sigma$ , et pointent vers l'utilisation de *mappings* *greedySplit* ou *exactSplit*.

### 3 Comparatif global

La comparaison à la *baseline* forme le dernier spectre d'observations à mener. On étudie ainsi en Figure 28 les performances des dissimilarités et des algorithmes de la *baseline*, par critère de performance de base. Les dissimilarités semblent bien dominer la *baseline*, ce qui sera à confirmer par des tests d'hypothèse, et présentent d'autre part des performances beaucoup plus stables d'un critère à l'autre (pas de gouffre pour l'*information score*, en particulier).

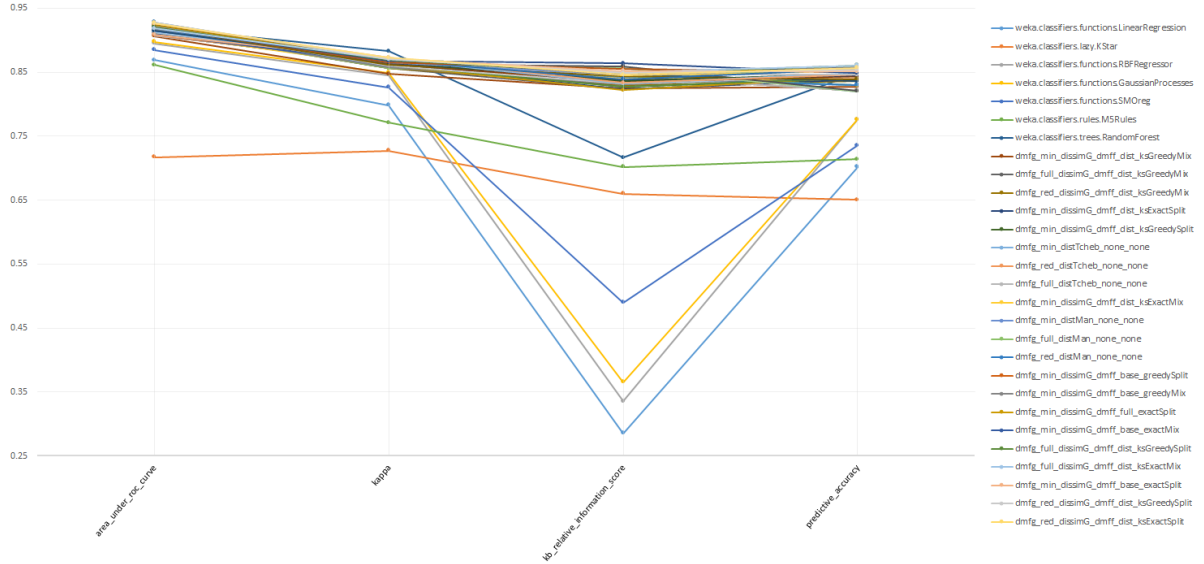


FIGURE 28 – Performances moyennes des différents algorithmes au méta-niveau, par critère de performance de base.

L'algorithme *KStar* (en orange, second de la liste) de la *baseline* semble présenter un profil de performance similaire aux dissimilarités (écart quasi-constant d'environ 20%), ce qui est raisonnable dans la mesure où il s'agit d'un algorithme proche de *kNN*, mais utilisant une distance basée sur l'entropie [4]. Cette stabilité peut caractériser une meilleure capacité à utiliser des critères de base non-normalisés (l'*information score* peut par exemple prendre de très grandes valeurs), qui est un avantage de la structure "*instance based*" des dissimilarités.

On valide ensuite l'amélioration par rapport à la baseline par des tests d'hypothèse. Se limiter à l'espace optimal réduit trop le nombre de résultats à comparer, ce qui affecte la puissance des tests et empêche de tirer des conclusions. On se limite donc en Figure 29 à comparer les différents algorithmes sur le critère d'aire sous la courbe, qui était notre meilleur candidat.

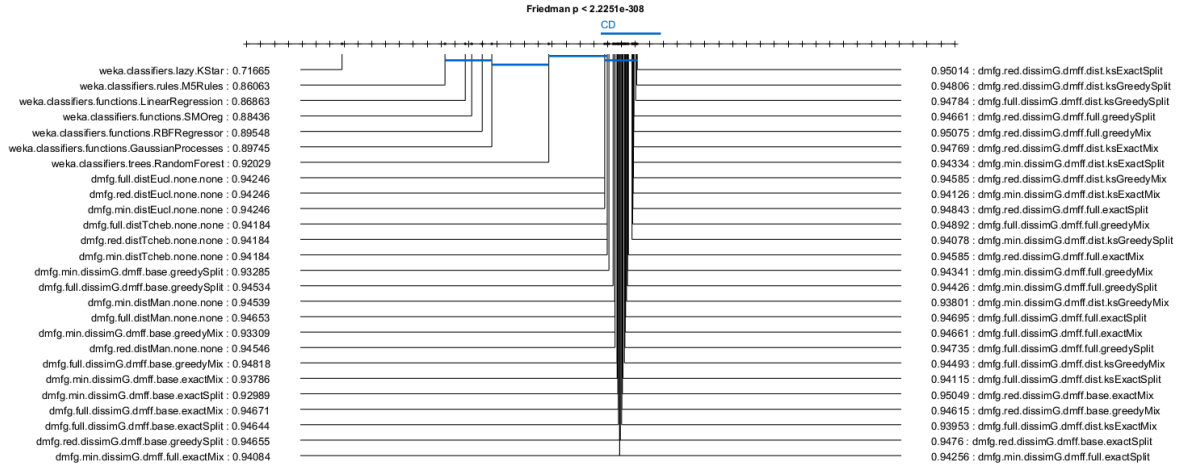


FIGURE 29 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents algorithmes au méta-niveau, pour le critère d'aire sous la courbe.

On valide donc bien une différence significative entre nos dissimilarités et la *baseline* d'algorithme traditionnels. Seules quelques dissimilarités basées sur des distances simples (donc de la *baseline*) ne peuvent être considérées significativement supérieures au meilleur algorithme de la *baseline* (RandomForest [2]). Le test ne suffit malheureusement pas à établir de différence significative entre ces dissimilarités de la *baseline* et les autres, mais cette différence a déjà pu être constatée dans la section précédente.

Un test plus puissant utilisant tous les critères de performance de base sur l'espace des algorithmes au niveau méta (Figure 30) confirme la supériorité des dissimilarités par rapport à tous les algorithmes traditionnels de la *baseline*, mais là encore ne met pas en évidence de différence significative entre les dissimilarités.

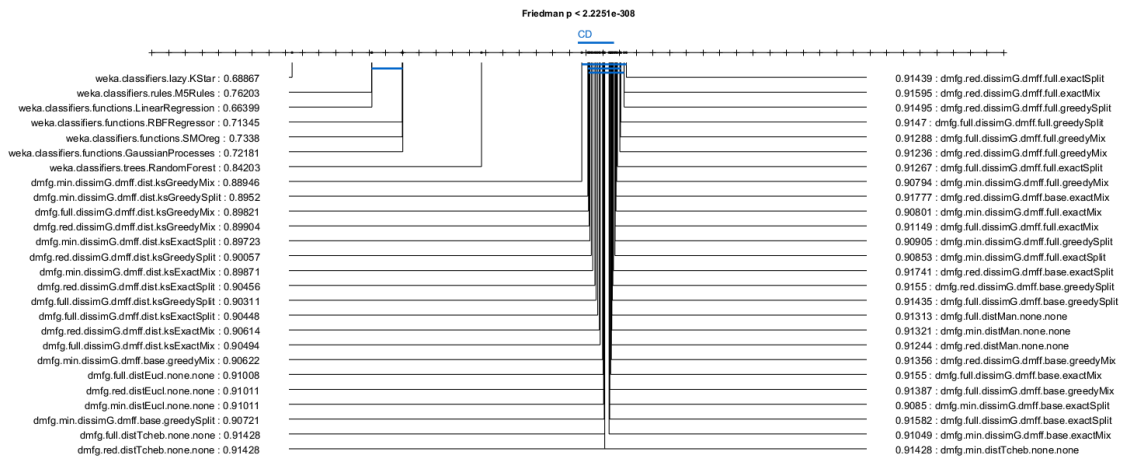


FIGURE 30 – Résultats du test de Friedman avec *post-hoc* de Nemenyi entre les échantillons représentant les différents algorithmes au méta-niveau.

Ces deux figures (29 et 30) nous permettent par ailleurs d’investiguer les méthodes ayant obtenu les meilleures performances moyennes au méta-niveau. Sur l’ensemble des critères (30), on trouve une dominance (dont on ne peut établir la signification à ce niveau) de l’ensemble de méta-attributs généraux comportant un nombre réduit de *landmarkers*, et l’ensemble de méta-attributs des attributs complet. Si l’on se restreint à l’aire sous la courbe, on voit figurer en tête les dissimilarités comparant les distributions des attributs par le test de Kolmogorov-Smirnov, avec une performance au méta-niveau dépassant les 0.95. Cela signifie que ces dissimilarité ont permis de recommander des algorithmes en moyenne 95% aussi bons que le meilleur disponible !

## Références

- [1] GEAPA Batista and Diego Furtado Silva. How k-nearest neighbor parameters affect its performance. In *Argentine Symposium on Artificial Intelligence*, pages 1–12, 2009.
- [2] Leo Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.
- [3] Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Luján. Conditional likelihood maximisation : a unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research*, 13(1) :27–66, 2012.
- [4] John G Cleary, Leonard E Trigg, et al. K\* : An instance-based learner using an entropic distance measure. In *Proceedings of the 12th International Conference on Machine learning*, volume 5, pages 108–114, 1995.