Utilisation des graphes de proximité dans le cadre de l'apprentissage basé sur les voisins

Sylvain Ferrandiz*, Marc Boullé**
*2 avenue Pierre Marzin, 22307 Lannion
sylvain.ferrandiz@rd.francetelecom.com,
**2 avenue Pierre Marzin, 22307 Lannion
marc.boulle@rd.francetelecom.com,

Résumé. La classification suivant les plus proches voisins est une règle simple et attractive, basée sur une définition paramétrique du voisinage. Les graphes de proximité, quant à eux, induisent des notions plus souples de voisinage. Il s'agit ici d'effectuer la substitution.

Les variantes obtenues, peu testées dans la bibliographie, ont été soumises à une expérimentation intensive, sur bases de données de l'UCI et de France Télécom. On a ainsi considéré divers types de prétraitement des données et plusieurs catégories de graphes. De plus, on a caractérisé les effets du "piège de la dimension" sur le comportement théorique de tous les graphes présentés, une quantification empirique du phénomène ayant été réalisée.

Il ressort de notre étude que l'utilisation du voisinage de Gabriel provoque une amélioration en moyenne et que le prétraitement basé sur la statistique de rang est le plus adéquate. Quoiqu'il arrive, des précautions doivent être prises en grande dimension.

1 Introduction

La classification constitue un problème d'apprentissage classique qui se présente comme suit. On cherche à prédire la valeur d'un attribut cible (ou variable endogène), prédiction basée sur un ensemble S de n instances structurées en vecteurs d'un produit cartésien de d attributs descriptifs (ou variables exogènes) et un ensemble E d'étiquettes correspondantes. On suppose ici les attributs descriptifs tous à valeurs réelles.

Une méthode de classification, proposée dans [Fix et Hodges, 1951], est la règle k-NN. Elle consiste, pour une instance x à étiqueter, à effectuer un vote à la majorité sur les k plus proches voisins (au sens euclidien) de x dans S. Un résultat de [Cover et Hart, 1967], étendu à toutes distributions par Stone et Devroye, affirme que le plus proche voisin de x contient plus de la moitié de l'information sur x. Autrement dit, si L^* désigne l'erreur bayésienne (optimale) et $L_{1-\mathrm{NN}}$ l'erreur de la règle 1-NN, on a $L^* \leq L_{1-\mathrm{NN}} \leq 2L^*$.

De plus, la règle k-NN, dans le cas à deux modalités cibles, est asymptotiquement optimale [Stone, 1977]. Ainsi, si n devient infini et si la suite (k_n) du nombre de voisins pris en compte tend vers l'infini suffisament lentement par rapport à n (i.e $k_n/n \to 0$), l'erreur $L_{\mathrm{k_n}-\mathrm{NN}}$ converge en probabilité vers L^*.

La classification repose donc sur le voisinage de l'instance à étiqueter. Dans la règle k-NN, celui-ci est défini de manière globale et paramétrique. L'introduction des graphes

de proximité permet d'envisager d'autres caractérisations du voisinage d'un point, non paramétriques et locales.

L'organisation du papier est la suivante. Tout d'abord, on présente les graphes de proximité sélectionnés. On étudie également les effets de l'accroissement du nombre d'attributs, sur données synthétiques et réelles. Tout ceci est l'objet de la section 2. Ensuite, dans la section 3, on décrit les méthodes obtenues par changement du critère de voisinage. Enfin, dans la section 4, on examine les résultats obtenus.

2 Etude des graphes de proximité

2.1 Présentation des graphes

Commençons par les graphes de voisinage relatif, de Gabriel et d'influence rectangulaire. Soient s, t deux instances de S. Au couple (s, t) on associe les voisinages suivants :

- $U_{(s,t)}$: intersection des sphères centrées en s et t de rayon la distance entre s et t,
- $V_{(s,t)}$: sphère de diamètre (s, t),
- $W_{(s,t)}$: plus petit hyperrectangle dont les faces sont perpendiculaires à un axe et dont s et t sont les sommets.

Le couple (s, t) est alors une arête du graphe si le voisinage associé ne contient aucun autre point de S. Avec $U_{(s,t)}$ (resp. $V_{(s,t)}$, resp. $W_{(s,t)}$), on obtient le graphe de voisinage relatif (resp. de Gabriel, resp. d'influence rectangulaire), abrégé en RNG (resp. GG, resp. RIG). De l'inclusion des voisinages, on déduit que RNG est un sous-graphe de GG qui lui-même est contenu dans RIG. Le test d'appartenance d'un troisième point au voisinage nécessitant d opérations, le coût de construction de chacun de ces graphes est un $O(dn^3)$. Une heuristique, proposée dans [Toussaint *et al.*, 1985], permet de diminuer la quantité de tests d'appartenance à effectuer et, par voie de conséquence, la complexité.

Pour le graphe d'influence sphérique, abrégé en SIG, on procède autrement. A tout point de S, on commence par associer sa sphère d'influence, soit la sphère centrée en ce point de rayon la plus petite distance entre ce point et tout autre point de S. Dès lors, un couple (s, t) forme une arête du graphe si les sphères d'influence de s et t s'intersectent. La construction du graphe demande donc un $O(dn^2)$ opérations au total. Notons que, contrairement aux autres graphes présentés ici, SIG peut être non connexe.

Terminons par le graphe de Delaunay, abrégé en DG. $d + 1$ points de S forment un polyhèdre de Delaunay si la sphère qui leur est circonscrite ne contient aucun autre point de S. Alors, tout couple parmi ces points forme une arête du graphe de Delaunay. La complexité de calcul de ce graphe est un $O(n^{d+2})$. Notons que DG contient GG.

2.2 Comportement en grande dimension

Il est reconnu que la notion de plus proche voisin perd parfois son sens en grande dimension. Nous allons voir qu'il en est de même pour la notion de voisin sur le graphe,

en exploitant un résultat de [Beyer *et al.*, 1999]. Posons quelques notations. Soient $P_{0,d}, ..., P_{n,d}$ des points de $\mathbb{R}^d$ distribués indépendamment et de même loi. On désigne par δ_d une application quelconque de $\mathbb{R}^d \times \mathbb{R}^d$ dans $\mathbb{R}_+$ (appelée dans la suite à jouer le rôle du carré de la distance euclidienne, mais pouvant être n'importe quelle mesure de similitude) et on note

$$DMIN_d = \min_{1 \leq i \leq n} \delta_d(P_{0,d}, P_{i,d}),\ DMAX_d = \max_{1 \leq i \leq n} \delta_d(P_{0,d}, P_{i,d}).$$

Comme la loi de $\delta_p(P_{i,d}, P_{j,d})$ est indépendante du choix de (i, j), on note W_d une variable de même loi. On suppose W_d d'espérance et de variance finies.

Théorème 1 . – *Si le rapport de l'écart-type par l'espérance de W_d tend vers 0 avec la dimension, alors $DMAX_d/DMIN_d$ converge en probabilité vers* 1.

Lorsque δ_d désigne par exemple le carré de la distance euclidienne sur $\mathbb{R}^d$, cela signifie que, pour un ε fixé quelconque, la probabilité que l'écart entre le point le plus éloigné de $P_{0,d}$ et le point le plus proche soit inférieur à ε tend vers 1 avec la dimension. De plus, lorsque les composantes de chaque points sont indépendantes, d'après la loi faible des grands nombres, on est sous les hypothèses du théorème.

Ainsi, sous condition, la position de trois points tend à devenir équilatérale. En conséquence, le graphe de Gabriel tend à devenir complet. Par inclusion, il en est de même pour les graphes de Delaunay et d'influence rectangulaire. En ce qui concerne le graphe de voisinage relatif, la position équilatérale est une position limite vis-à-vis du voisinage de deux points. Celle-ci se jouant de plus en plus sûrement à un ε près, les voisins sur le graphe sont peu stables. Enfin, les rapports de distances convergeant vers 1, l'intersection de deux sphères d'influence devient de plus en plus probable et le graphe d'influence sphérique tend à devenir complet.

2.3 Caractérisation empirique

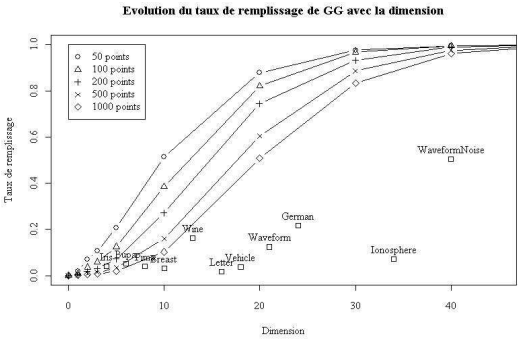

FIG. 1 – Taux de remplissage du graphe de Gabriel

Pour quantifier le "piège de la dimension" sur les graphes de proximité, on a évalué leur taux de remplissage, c'est-à-dire le rapport entre leur nombre d'arêtes et celui du

graphe complet. Les jeux de données sont de deux types : réels (bases de l'UCI) et synthétiques. Pour les seconds, on a généré des nuages de points dans l'hypercube de Hamming uniformément et indépendamment par composante, pour diverses valeurs de la dimension d et du nombre d'instances n. Dans ce cas, l'hypothèse du théorème 1 est vérifiée et on s'attend à ce que le taux de remplissage converge vers 1, sauf pour RNG.

Une description des bases de données utilisées est donnée en annexe. Le principal intérêt de la figure 1 réside dans l'illustration du fait que les distributions réelles sont souvent loin de vérifier la condition du théorème 1. Le comportement de SIG est sensiblement le même, avec une tendance à mailler légèrement plus et plus vite.

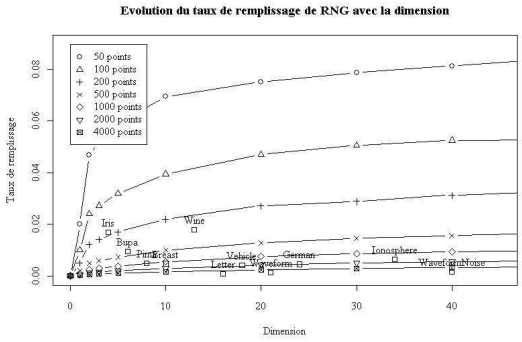

FIG. 2 – Taux de remplissage du graphe de voisinage relatif

En revanche, à la vue de la figure 2 et du faible taux de remplissage, on s'aperçoit que RNG est fortement biaisé et favorise les voisinages lacunaires. Ceci, couplé avec le fait que les voisinages peuvent devenir instables (même si les distributions réelles ne semblent pas favoriser ce genre de cas), laisse entrevoir un comportement moins favorable par la suite.

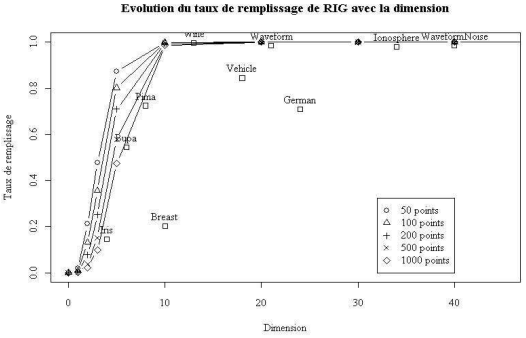

FIG. 3 – Taux de remplissage du graphe d'influence rectangulaire

Un regard à la figure 3 suffit pour remarquer que le graphe d'influence rectangu-

laire est inutilisable en pratique. Bien qu'on s'y attende, ce graphe maille beaucoup et beaucoup trop vite. De plus, et c'est rédhibitoire, il est complet (ou proche de la complétude) pour la plupart des jeux de données réelles utilisés.

3 Apprentissage basé sur les voisins

3.1 Règle de classification

Dans la règle de k-NN usuelle, une instance est étiquetée suivant un vote à la majorité sur ses k plus proches voisins au sens euclidien. L'introduction d'un graphe de proximité permet d'envisager la modification naturelle suivante :

- Pour tout x à classifier
  - Trouver les voisins $y_1, ..., y_k$ de x sur le graphe basé sur $E \bigcup \{x\}$
  - Classifier x suivant un vote à la majorité sur $y_1, ..., y_k$

Le comportement asymptotique de cette régle, qu'on notera G-NN, a été étudié dans [Devroye *et al.*, 1996] pour les graphes de Gabriel et d'influence rectangulaire. Il ressort que la règle G-NN est asymptotiquement optimale lorsqu'elle utilise RIG et asymptotiquement meilleure que 1-NN lorsqu'elle utilise GG. En ce qui concerne l'étude empirique, elle n'a été menée que pour GG et RNG, dans (Sànchez 2000).

3.2 Réduction de la base de données

Un des défauts de la règle k-NN est une phase d'apprentissage réduite à sa plus simple expression : on stocke tous les exemples. Comme de plus chaque nouvel étiquetage nécessite le parcours de tous les exemples, il a été rapidement envisagé de réduire le nombre d'observations stockées. En effet, en vue d'appliquer la règle 1-NN, il est inutile de conserver les instances centrales. De nombreuses méthodes de détection ont été proposées, mais peu se sont révélées complètement satisfaisantes. On parle à leur sujet de règles de compression.

Afin de conserver les frontières de décision de la règle 1-NN, celles-ci étant définies par le diagramme de Voronoi de l'ensemble des exemples, il convient d'éliminer les instances de même étiquette que toutes les cellules de Voronoi adjacentes. Par dualité, cela revient à considérer les voisins de l'instance sur le graphe de Delaunay. Cette méthode à été proposée par [Toussaint *et al.*, 1985]. Il a de plus été envisagé d'utiliser un sous-graphe de DG (GG ou RNG). Nous sommes allés plus loin en incluant dans notre étude les graphes d'influence.

3.3 Elimination des instances mal étiquetées

Dans l'objectif de réduire le nombre de données stockées, il peut également être intéressant de détecter les exemples mal étiquetés, exemples qui polluent la source d'expérience. Les règles obtenues sont dites d'édition. Le problème réside dans la bonne définition du mauvais étiquetage de l'exemple. Dans [Wilson, 1972], il est proposé de

considérer une observation bruitée si son étiquette n'est pas celle obtenue par un vote à la majorité sur ses k plus proches voisins.

L'idée de remplacer le voisinage paramétrique par un voisinage dérivé d'une structure de graphe a été proposée et testée dans [Sànchez, 2000], seulement pour les graphes de Gabriel et de voisinage relatif. Nous avons ici également pris en compte les graphes d'influence.

Notons que la définition choisie du mauvais étiquetage amène à considérer les instances frontalières comme des instances bruitées. En effet, il est assez probable qu'une majorité des voisins d'une instance frontalière appartiennent à une autre classe. Ainsi, ce type d'élimination rogne le bord des régions et modifie la frontière de décision de manière incontrôlée.

3.4 Prétraitement des données

L'utilisation de la distance euclidienne, qui somme les contributions de chaque composante, tend à privilégier les attributs fortement étendus. De plus, une telle distance est sensible au facteur d'échelle de l'attribut. Les remèdes usuels consistent à centrer-réduire ou à rapporter à son étendue l'attribut. On propose ici un autre type de prétraitement.

Posons d'abord quelques notations. Notons $\{a_1, ..., a_q\}$ l'ensemble des valeurs du i^{eme} attribut. On note alors $(a_{(1)}, ..., a_{(q)})$ la statistique d'ordre associée. Est ainsi naturellement définie une partition de $\mathbb{R}$ en q intervalles, numérotés de 1 à q. Dès lors, pour tout t réel, on définit $1 \leq r_i(t) \leq q$ comme l'indice de l'intervalle contenant t. Enfin, pour tout $x \in A = \mathbb{R}^d$, on définit $R(x) = (r_1(x_1), ..., r_d(x_d))$.

L'application R ainsi obtenue est appelée opérateur de rang et ne dépend pas du facteur d'échelle de l'attribut. Au final, pour ne pas favoriser les attributs prenant un grand nombre de valeurs, on divise chaque composante par le nombre q de valeurs distinctes. Ce type de prétraitement a l'avantage, contrairement aux deux précédents, d'uniformiser la répartition des valeurs de l'attribut. En apprentissage, les valeurs de l'attribut sont ordonnées, pour un coût en $O(\log n)$, et en test, le calcul de $R(x)$ est en $O(d \log n)$, pour toute instance x.

4 Résultats expérimentaux

4.1 Classification

Le but est d'évaluer la différence de performance de la règle G-NN vis-à-vis de la règle k-NN usuelle. Nous avons suivi pour cela deux directions. La première consiste à comparer la règle G-NN avec la règle 1-NN. Pour chaque base de données, la différence entre le taux de prédiction de G-NN et celui de 1-NN est calculé. Dans un second temps, on a effectué la même comparaison mais avec la règle qu'on appellera ici Bestk-NN : pour chaque jeu de données, on a sélectionné le meilleur taux de prédiction obtenu en faisant varier k dans la règle k-NN ($k \in \{1, 3, 7, 15, 30\}$).

Les différences obtenues sont représentées graphiquement sur la figure 4, par type de graphe. Notons que les bases de données sont ici numérotées de 1 à 10, par taille

croissante, de la base Iris à WaveformNoise.

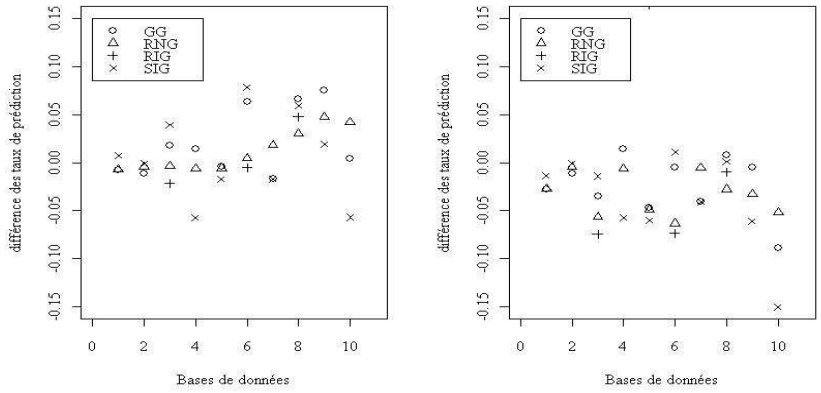

FIG. 4 – Différences de taux de prédiction entre G-NN et 1-NN (à gauche) et entre G-NN et Bestk-NN (à droite)

Comparativement à la règle 1-NN, la classification basée sur GG et RNG aboutit à de meilleurs résultats (+0.02 en moyenne pour GG et +0.01 pour RNG), mais pas de manière significative. Il en est de même si on utilise les sphères d'influence (+0.005).

En ce qui concerne l'application du voisinage d'influence rectangulaire, on se heurte au problème soulevé par l'étude des graphes. Tout le monde ayant tendance à être voisin de tout le monde, la classification se rapproche dangereusement de la prédiction de la classe majoritaire sur l'ensemble d'apprentissage. C'est pourquoi les différences de taux de prédiction sont parfois très élevées et n'apparaissent pas sur les figures.

Lorsqu'on compare la règle G-NN avec Bestk-NN, on constate qu'elle ne la dépasse que rarement et lui est donc inférieure en moyenne (-0.02 pour GG, -0.03 pour RNG et -0.04 pour SIG). Ceci peut laisser penser que les divers types de voisinage proposés ne captent pas de manière optimale la notion de voisinage d'un point.

4.2 Compression

La méthode de compression basée sur le graphe de Delaunay conserve les frontières de décision. Nous avons illustré cette propriété sur le problème XOR (cf figure 5).

Notons tout de même un effet de bord (qui porte ici bien son nom) : les cellules de Voronoi des instances situées sur l'enveloppe convexe du nuage de point sont non bornées, et donnent naissance à des relations d'adjacence peu conformes à l'intuition. L'inconvénient principal de la méthode réside néanmoins dans le coût de calcul de DG exponentiel en la dimension d du problème.

Lorsqu'on remplace DG par un sous-graphe, on perd évidemment la consistance avec les frontières, mais on gagne en taux de réduction. D'ailleurs, le choix du sous-graphe repose sur un compromis entre perte de consistance et augmentation de la

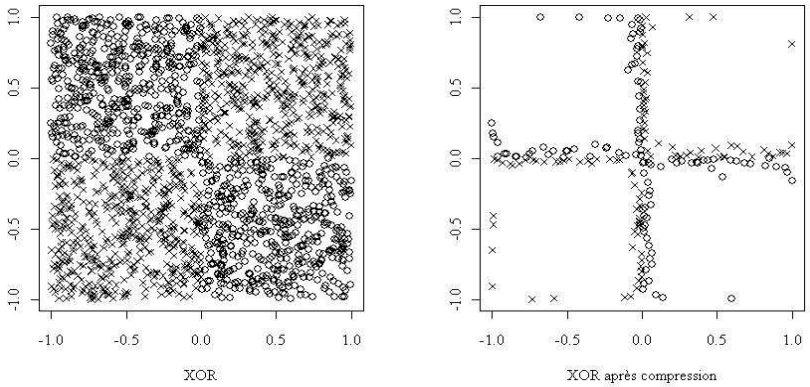

FIG. 5 – Compression du problème XOR avec le graphe de Delaunay

réduction. Du fait que le graphe de voisinage relatif génère beaucoup moins d'arêtes, on est donc a priori moins confiant en son utilisation (cf figure 6).

Le graphe de Gabriel possède donc l'avantage sur RNG d'être plus proche de DG et l'avantage sur DG de nécessiter moins de calculs pour sa construction. De plus, on constate par rapport à DG que l'effet de bord ne se produit pas avec GG. Sachant que plus la dimension augmente, plus l'effet de bord tend à devenir la règle, le comportement de GG apparaît des plus satisfaisant.

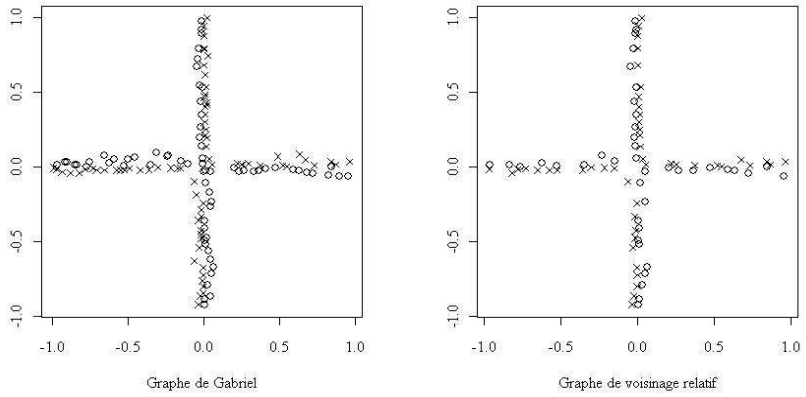

FIG. 6 – Compression du problème XOR avec le graphe de Gabriel et le graphe de voisinage relatif

Ceci se confirme au regard des résultats sur les bases de l'UCI. Le taux de prédiction est sensiblement affecté par la sélection basée sur RNG (-0.04 en moyenne sur les 11

bases de l'UCI), pour un faible taux de conservation (54% en moyenne), alors que la prédiction est quasiment inchangée lorsqu'on utilise *GG* (-0.01). Dans ce cas, le taux de conservation est beaucoup plus élevé (81%).

En lieu et place des graphes *DG*, *GG*, *RNG*, on peut être tenté d'utiliser *SIG* ou *RIG* pour compresser la base de données. Du fait de son goût pour la complétude, RIG est inintéressant : les instances sont quasiment toutes conservées. Pour le graphe d'influence sphérique, sa non connexité peut amener à éliminer tout un amas d'instances identiquement étiquetées. C'est ce qui se passe avec la base Iris, le taux de prédiction de la règle 1-NN passant de 0.96 à 0.63. Intuitivement, si l'on possède un nombre suffisant d'instances, ce problème ne doit pas se poser. D'ailleurs, en dehors de cette base, le taux de prédiction ne subit qu'une diminution de 0.01 en moyenne pour un taux de conservation moyen de 85%, ce qui confère à la méthode un comportement équivalent à celle basée sur le graphe de Gabriel.

4.3 Edition

Pour bien situer les résultats de l'édition basée sur les graphes de proximité, on évalue la différence de comportement avec la règle de Wilson utilisant 3 voisins (notée ici 3-ENN). Les deux critères sélectionnés sont le taux de prédiction et le taux de compression. On a ainsi soustrait le taux obtenu avec 3-ENN du taux obtenu avec l'édition par les graphes, les résultats apparaissant sur la figure 7. Les bases de données sont encore numérotées de 1 à 10.

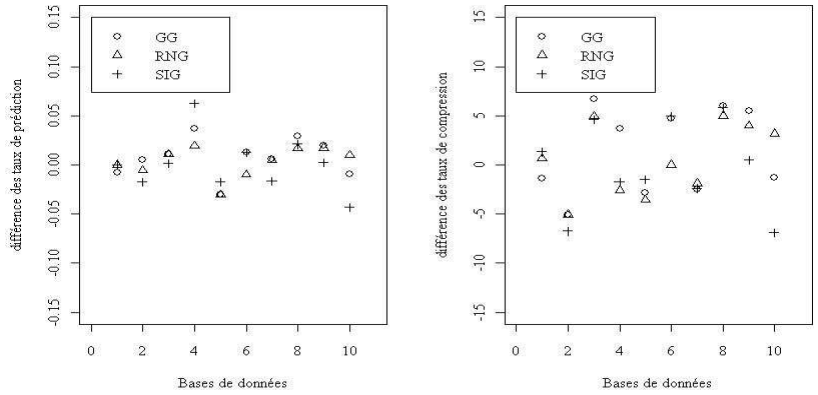

FIG. 7 – Différences de taux de prédiction et de compression entre l'édition basée sur les graphes et 3-ENN

L'édition utilisant le graphe d'influence rectangulaire a été volontairement omise. On constate que les comportements sont globalement similaires. En moyenne, le taux de prédiction augmente peu (0.007 pour *GG*, 0.003 pour *RNG* et 0.0006 pour *SIG*) et le taux de compression reste stable (+0.014 pour *GG*, +0.005 pour *RNG* et -0.002 pour *SIG*).

4.4 Prétraitement

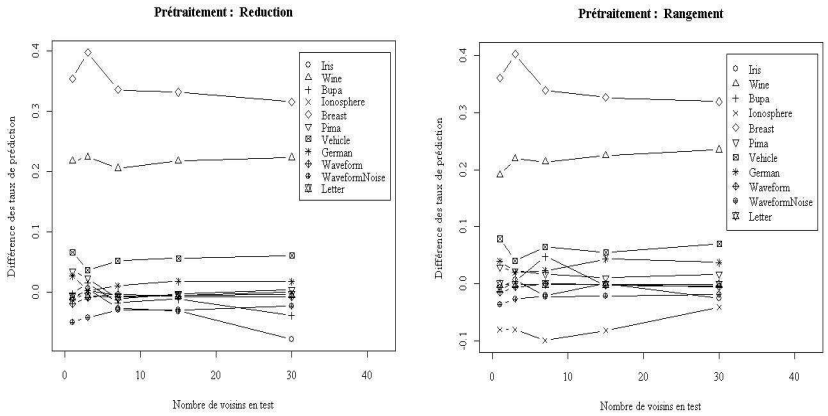

FIG. 8 – Comparaison du taux de prédiction sans et avec réduction (à gauche) et sans et avec rangement (à droite)

Pour illustrer l'évolution du taux de prédiction de la règle k-NN avec le prétraitement, on a, pour $k \in \{1, 3, 7, 15, 30\}$, retranché du taux de prédiction après prétraitement le taux obtenu sans prétraitement, ceci pour chacun des trois prétraitements envisagés. On a ainsi obtenu la figure 8, sur laquelle on peut comparer les variations provoquées par le centrage-réduction ou l'application de l'opérateur de rang sur le taux de prédiction.

Prétraitement	**1-NN**	**3-NN**	**7-NN**	**15-NN**	**30-NN**
Aucun	0.27	0.27	0.27	0.26	0.26
Reduction	0.31	0.32	0.35	0.36	0.38
Normalisation	0.30	0.30	0.31	0.34	0.34
Rangement	0.35	0.36	0.38	0.40	0.42

TAB. 1 – Taux de prédiction de la règle k-NN sur le jeu de données France Télécom

Au regard des résultats obtenus, les trois types de prétraitement ne se distinguent pas de manière évidente. Leurs effets sont globalement identiques sur le taux de prédiction, aspect conforté par le grand nombre et la variété des méthodes étudiées. Une raison essentielle pour laquelle l'application de l'opérateur de rang ne se distingue pas de celle des pondérations est que peu d'attributs possèdent des distributions non uniformes de leurs valeurs, sur les bases de données de l'UCI.

Ainsi, sur un jeu de données France Télécom, contenant des attributs de distribution exponentielle (historique de consommation), la différence est significative (cf tableau 1). La tâche consiste à predire la classe d'âge du client en fonction de sa consommation et des divers services qu'il a souscrit. Les quatre classes cibles sont équilibrées et on

constate d'une part que sans prétraitement, on ne tire rien du jeu de données et, d'autre part, que c'est l'emploi de l'opérateur de rang qui rend la prédiction la plus efficace.

5 Conclusion

Nous avons étudié de manière intensive le changement de critère de voisinage d'une instance. Le graphe d'influence rectangulaire est inutilisable car tombant rapidement dans le "piège" de la dimension.

Pour la classification, le graphe de Gabriel domine en moyenne les graphes d'influence sphérique et de voisinage relatif. Bien que les résultats de ces règles soient supérieurs à ceux de la règle 1-NN classique, elles échouent à obtenir un ajustement optimal du voisinage.

L'édition réalisée à l'aide des graphes améliore le taux de prédiction, le meilleur candidat étant encore le graphe de Gabriel, et conduit à des ensembles d'instances presque de même taille dans les trois cas. On est donc assez confiant en le fait que cette méthode reconnaît les données mal étiquetées, le prix à payer étant une légère modification des frontières de décision.

La compression idéale étant basée sur la triangulation de Delaunay, la considération de sous-graphes est une option intuitivement bien fondée, la préférence allant dès le départ vers le graphe de Gabriel. Ceci a été confirmé par les résultats empiriques. De plus, l'emploi du graphe d'influence sphérique en lieu et place du graphe de Gabriel aboutit à des résultats proches, sa non connexité pouvant cependant poser problème.

En ce qui concerne le prétraitement des données, aucun des trois types mis en balance ici ne se différencie significativement sur l'ensemble des bases de l'UCI. Le remplacement par la statistique d'ordre apparaît néanmoins comme préférable, provoquant une uniformisation de la répartition des valeurs de l'attribut. Une confirmation est apportée par les résultats obtenus sur la base France Télécom, plus discriminante à cet égard.

Références

[Beyer *et al.*, 1999] K. Beyer, J. Goldstein, R. Raghu, et U. Shaft. When is "nearest neighbor" meaningful ? *ICDT Conference proceedings*, 1999.

[Cover et Hart, 1967] T.M. Cover et P.E. Hart. Nearest neighbor pattern classification. *Ins. of electrical and electronics engineers transactions on information theory*, 13 :21–27, 1967.

[Devroye *et al.*, 1996] L. Devroye, L. Gyorfi, et G. Lugosi. *A probabilistic theory of pattern recognition.* Springer Verlag, New-York, 1996.

[Fix et Hodges, 1951] E. Fix et J. Hodges. Discriminatory analysis. nonparametric discrimination : Consistency properties. *Technical Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX*, 1951.

[Sànchez, 2000] J.S. Sànchez. *Aprendizaje y classificación basados en criterios de vecindad. Métodos alternativos y análisis comparativo.* Thèse de doctorat, Universitat Jaume I, 2000.

[Stone, 1977] C. Stone. Consistent non parametric regression. *Annals of statistics*, 5 :595–645, 1977.

[Toussaint *et al.*, 1985] G.T. Toussaint, B.K. Bhattacharya, et R.S. Poulsen. The application of voronoi diagrams to nonparametric decision rules. *Computer Science and Statistics : The Interface*, pages 97–108, 1985.

[Wilson, 1972] D.L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on systems, man and cybernetics*, 2 :408–421, 1972.

Annexe 1

Base	**Instances**	**Attributs**	**Modalités**	**Prédiction majoritaire**
Iris	150	4	3	0,33
Wine	178	13	3	0,40
Bupa	345	6	2	0,58
Ionosphere	351	34	2	0,64
Breast	699	10	2	0,66
Pima	768	8	2	0,65
Vehicle	846	18	4	0,26
German	1000	24	2	0,70
Waveform	5000	21	3	0,34
WaveformNoise	5000	40	3	0,34
Letter	20000	16	26	0,04

TAB. 2 – Description des bases de données UCI utilisées : nombre d'instances, nombre d'attributs, nombre de modalités de la classe cible, taux de prédiction du prédicteur de classe majoritaire

Summary

The Nearest Neighbor Rule is a simple and attractive classification rule, based on a parametric definition of neighborhood. Beside, proximity graphs handle notions of neighborhood far more flexible. The substitution is made here.

The alternatives obtained, not fully tested in the bibliography, were subjected to an intensive experimentation, on UCI and France Télécom data bases. One thus considered various types of pretreatment of the data and several categories of proximity graphs. Moreover, one characterized the effects of the " the curse of dimensionality " on the theoretical behavior of all the graphs presented, an empirical quantification of the phenomenon having been realized.

It comes out from our study that the use of the Gabriel neighborhood causes an improvement on average, and that the better pretreatment is the one based on the rank statistic. Whatever may happen, precautions must be taken in high dimension.