

Classification à base de clustering ou décrire et prédire simultanément

O. Alaoui Ismaili^{1,2}

V. Lemaire¹

A. Cornuéjols²

¹ Orange Labs, 2 av. Pierre Marzin 22307 Lannion, France

² AgroParisTech, 16, rue Claude Bernard 75005 Paris, France

oumaima.alaouiismaili@orange.com

Résumé

Dans certains domaines applicatifs, la compréhension (la description) des résultats issus d'un classifieur est une condition aussi importante que sa performance prédictive. De ce fait, la qualité du classifieur réside donc dans sa capacité à fournir des résultats ayant de bonnes performances en prédiction tout en produisant simultanément des résultats compréhensibles par l'utilisateur. On parle ici du compromis interprétation vs. performance des modèles d'apprentissage automatique. Dans cette thèse, on s'intéresse à traiter cette problématique. L'objectif est donc de proposer un classifieur capable de décrire les instances d'un problème de classification supervisée tout en prédisant leur classe d'appartenance (simultanément).

Mots Clef

Classification, Clustering, Interprétation, Performance, Prédiction, Description

Abstract

In some application areas, the ability to understand the results given by a classifier (description) is as important condition as its predictive performance. In this case, the classifier is considered as important if it can produce comprehensible results with a good predictive performance. This is referred as a trade-off 'interpretation vs. performance'. In our study, we are interested to deal this problematic. The objective is then to find out a model which is able to describe the instances from a supervised classification problem and predict their class membership simultaneously.

Keywords

Classification, Clustering, Interpretation, Performance, Prediction, Description

1 Introduction

De nos jours, les données récoltées par ou pour les entreprises sont devenues un atout important. Les informations présentes, mais à découvrir au sein des grands volumes de données, sont devenues pour ces entreprises un facteur

de compétitivité et d'innovation. Par exemple, les grandes entreprises comme *Orange* et *Amazon* peuvent avoir un aperçu des attentes et des besoins de leurs clients à travers la connaissance de leurs comportements. Ces données permettent aussi de découvrir et d'expliquer certains phénomènes existants ou bien d'extrapoler des nouvelles informations à partir des informations présentes.

Pour pouvoir exploiter ces données et en extraire des connaissances, de nombreuses techniques ont été développées. Par exemple, l'analyse multivariée regroupe les méthodes statistiques qui s'attachent à l'observation et au traitement simultané de plusieurs variables statistiques en vue d'en dégager une information synthétique pertinente. Les deux grandes catégories de méthodes d'analyse statistique multivariées sont, d'une part, les méthodes dites descriptives et, d'autre part, les méthodes dites explicatives.

Les méthodes descriptives ont pour objectif d'aider à structurer et à résumer un ensemble de données issus de plusieurs variables, sans privilégier particulièrement l'une de ces variables. Toutes les variables sont donc prises en compte au même niveau. Les méthodes descriptives d'analyse multivariée les plus utilisées sont l'analyse en composantes principales (ACP) [1], l'analyse factorielle des correspondances (AFC) [1], l'analyse des correspondances multiples (ACM) [1] et les méthodes de clustering [2].

Les méthodes explicatives ont, quant à elles, pour objectif d'expliquer l'une des variables (dite dépendante) à l'aide de deux ou plusieurs variables explicatives (dites indépendantes). Les principales méthodes explicatives utilisées [3] sont la régression multiple, la régression logistique, l'analyse de variance, l'analyse discriminante, les arbres de décision ([4], [5]), les SVM [6], etc.

Si on se place dans le cadre de l'apprentissage automatique [3], on peut placer les méthodes descriptives dans le domaine de l'apprentissage non supervisé et les méthodes explicatives dans le cas de l'apprentissage supervisé.

On se place dans cet article dans le cadre de l'apprentissage supervisé (classification) où il s'agit d'apprendre un concept cible ($X \rightarrow Y$) dans le but de le prédire ultérieurement pour de nouveaux exemples. Dans ce cadre, il existe de nombreux algorithmes que ce soit pour la classi-

fication ou la régression [7]. Pour ce qui est de la classification, certains algorithmes fournissent des résultats difficilement compréhensibles en direct par l'utilisateur : c'est le cas des modèles appelés "boîtes noires" (e.g les réseaux de neurones [8] (ANN) et les séparateurs à vaste marge [6] (SVM)). D'autres sont naturellement plus interprétables : c'est le cas des modèles boîtes blanches (e.g les arbres de décision [4], [5]). Généralement, les modèles boîtes noires sont plus performants que les modèles boîtes blanches. D'où la nécessité d'étudier le compromis interprétation - performance ([9],[10],[11],[12]).

Dans certains domaines appelés "domaines critiques" (e.g la médecine, les services bancaires, ...), la compréhension (la description) des résultats issus d'un modèle d'apprentissage est une condition aussi importante que sa performance prédictive. De ce fait, la qualité de ces modèles réside donc dans leurs capacités à fournir des résultats ayant de bonnes performances en prédiction tout en produisant simultanément des résultats compréhensibles par l'utilisateur. On parle alors dans ce cas de la problématique : compromis *interprétation vs. performance* des modèles d'apprentissage.

Pour tenter de résoudre cette problématique, deux grandes voies existent (voir Figure 1). La première voie consiste à rendre les algorithmes les plus performants (comme les SVM, les ANN, les forêts aléatoires, ...) plus interprétables. Cette voie a fait l'objet de nombreuses recherches dans les années passées et cela dans de nombreuses disciplines. La deuxième voie, elle consiste à améliorer la performance des modèles "naturellement" interprétables". Selon notre connaissance, cette voie a été peu traitée dans la littérature.

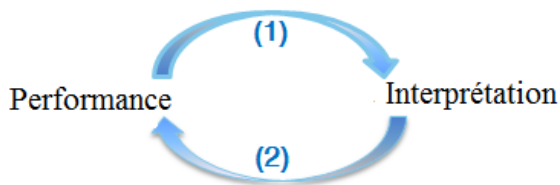


FIGURE 1 – La problématique d'interprétation performance

2 Objectif

Dans l'étude présentée dans cet article et qui fait l'objet d'une thèse en cours, nous nous intéressons à l'étude de la deuxième voie i.e de l'interprétation vers la performance. Cette étude a pour but de découvrir, si elle existe, la structure du concept cible à apprendre puis muni de cette structure de pouvoir prédire l'appartenance au concept cible. L'objectif est donc d'essayer de **décrire et de prédire d'une manière simultanée**. Si le but est atteint, on aura alors à la fois : (1) la prédiction du concept cible et (2) le pourquoi

de la prédiction grâce à la découverte de la structure de ce dernier.

Pour atteindre ce but, au moins pour l'axe "performance-prédiction", on pense alors immédiatement aux arbres de décisions [4, 5] qui figurent parmi les algorithmes de classification les plus interprétables. Ils ont la capacité de fournir à l'utilisateur des résultats compréhensibles (sous formes de règles) et qui semblent donner une structure du concept cible appris. En effet, l'arbre de décision modélise une hiérarchie de tests sur les valeurs d'un ensemble de variables discriminantes. De ce fait, les exemples obtenus suivant un certain chemin (de la racine vers les feuilles terminales) ont normalement la même classe et partagent ainsi les mêmes caractéristiques. Cependant, selon la distribution des données dans l'espace d'entrée, l'arbre de décision crée naturellement des polytopes (fermés et ouverts) à l'aide des règles. La présence des polytopes ouverts empêche l'algorithme de découvrir la structure 'complète' de la variable cible. La figure 2 présente un exemple illustratif du fonctionnement de l'arbre de décision. A partir de ce résultat, on constate que cet algorithme fusionne les deux sous-groupes 1 et 2 (entourés par les cercles bleus) malgré le fait que les exemples du premier sous-groupe ont des caractéristiques différentes de celles du deuxième sous-groupe. Dans le cas extrême, ces deux sous groupes peuvent même être très éloignés et donc être de caractéristiques assez différentes.

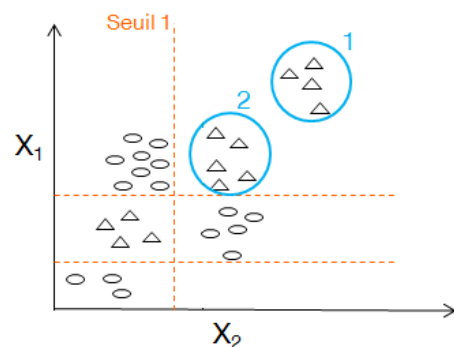


FIGURE 2 – Résolution d'un problème de classification binaire par un arbre de décision

Pour ce qui est de l'axe "interprétation - description", la découverte de la structure globale d'une base de données (non étiquetées) peut être réalisée à l'aide d'un algorithme de clustering. Les résultats fournis par les algorithmes de clustering sont souvent interprétables. L'utilisateur peut facilement identifier les profils moyens des individus appartenant à un groupe mais aussi les variables qui ont un grand impact sur la formation de chaque groupe. Ceci peut être réalisé à l'aide des outils statistiques simples mais variés (e.g le tableau des distances, les représentations graphiques et les indices de qualité). Dans notre cas d'étude, l'utilisation d'un algorithme de clustering s'avère insuffisante. En effet, le concept cible à apprendre n'est pas pris en compte et les prédictions ne peuvent pas être réalisées avec une bonne

performance (par exemple, lorsqu'on ajoute à la partition générée par le clustering le vote majoritaire et un classifieur comme le K-plus proches voisins pour prédire le concept cible des nouveaux exemples). En effet, deux exemples qui partagent les mêmes caractéristiques n'ont pas forcément le même concept cible.

Partant de ce constat, nous nous sommes fixés comme objectif d'essayer de réaliser au sein du même algorithme d'apprentissage la *description* et la *prédiction* du concept cible à apprendre. Pour cela nous proposons d'adapter un algorithme de clustering aux problèmes de classification supervisée. Autrement dit, l'idée est de modifier un l'algorithme de clustering afin qu'il soit un bon classifieur (en termes de prédiction) tout en gardant sa faculté à décrire les données et donc le concept cible à apprendre. On parle alors de la **classification à base de clustering** (ou décrire et prédire simultanément).

L'objectif de la thèse est donc de chercher un modèle qui prend en considération les points suivants :

1. La découverte de la structure interne de la variable cible (la proximité entre les individus).
2. La maximisation de la performance prédictive du modèle.
3. L'interprétation des résultats.
4. La minimisation des connaissances a priori requises de la part de l'utilisateur (i.e pas ou peu de paramètres utilisateur).

Cas d'usage : Le fondement des argumentaires commerciaux par le service marketing est basé sur la connaissance des attentes et des besoins des clients. Généralement, le service dispose d'une base de données clients décrite par un ensemble de variables descriptives et une variable cible (nominale). Par exemple, cette variable peut décrire l'attrition (perte de clientèle) ou l'appétence du client (volonté de souscrire à un service). A partir de ces données, le service marketing cherche à prédire l'appartenance des nouveaux clients à une des *sous-populations* pour laquelle les clients ont à la fois le même concept cible et des caractéristiques proches dans la base de données.

3 Classification à base de clustering

3.1 Définition

Dans notre contexte, la classification à base de clustering est défini comme étant un algorithme de clustering standard soumis à quelques modifications afin qu'il soit adapté aux problèmes de la classification supervisée. L'objectif ici est donc de décrire et prédire d'une manière simultanée (voir Figure 3).

3.2 État de l'art

Dans la littérature, il existe des algorithmes de clustering modifiés dans le but d'atteindre l'objectif désiré. Ces algorithmes sont connus sous le nom de '*supervised*

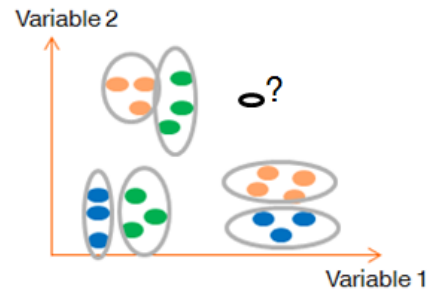


FIGURE 3 – Principe de la classification à base de clustering

clustering'. A titre illustratif, on citera par exemple :

- Al-Harbi et al. [13] proposent des modifications au niveau de l'algorithme des K-moyennes. Ils remplacent la distance euclidienne usuelle par une distance euclidienne pondérée. Le vecteur de poids est choisi de telle sorte que la confiance des partitions générées par l'algorithme des K-moyennes soit maximisée. Cette confiance est défini comme étant le pourcentage d'objets classés correctement par rapport au nombre total d'objets dans le jeu de données. Dans cet algorithme, le nombre de clusters est une entrée.

- Aguilar et al. [14] et Slonim et al. [15] ont proposé des méthodes basées sur l'approche agglomérative ascendante. Dans [14], les auteurs ont proposé un nouvel algorithme de clustering hiérarchique (S-NN) basé sur les techniques du plus proches voisins. Cet algorithme considère chaque entrée comme un cluster distinct. Ensuite, il fusionne successivement les clusters ayant des voisins identiques. Par conséquent, tous les voisins ayant les distances plus courtes que le premier ennemi (i.e l'objet qui n'a pas la même étiquette) seront collectés. Tishby et al. ont introduit dans [16] la méthode 'information bottleneck'. Basée sur cette méthode, ils ont proposé une nouvelle méthode de clustering (agglomérative) [15] qui maximise d'une manière explicite, l'information mutuelle entre les données et la variable cible par cluster.

- Dans [17], Cevikalp et al. ont proposé une méthode qui crée des clusters homogènes, nommée HC. Le nombre et l'emplacement des clusters sont déterminés en fonction de la répartition des clusters ayant des chevauchements entre les classes. L'idée centrale de l'algorithme HC est de partir d'un nombre de clusters égale au nombre de classes puis de diviser les clusters qui se chevauchent en tenant compte l'information supplémentaire donnée par la variable cible.

- Eick et al. [18] proposent quatre algorithmes de clustering supervisés, basés sur des exemples représentatifs. Ce genre d'algorithme a pour but de trouver un sous-ensemble de représentants dans l'ensemble d'entrées de telle sorte

que le clustering généré en utilisant ce dernier minimise une certaine fonction de pertinence. Dans [18], les auteurs utilisent une nouvelle fonction pour mesurer la qualité de ces algorithmes. Cette fonction remplit les deux critères suivants : i) Minimisation de l'impureté de classe dans chaque cluster ii) Minimisation du nombre de clusters.

- SPAM est le premier algorithme proposé par Eick et al. ; Il est une variation de l'algorithme de clustering PAM (Partitioning Around Medoids). Le deuxième algorithme proposé par les auteurs dans [18] est SRIDHCR (Single Representative Insertion/Deletion Steepest Decent Hill Climbing with Randomized Restart). Cet algorithme est un algorithme itératif, il commence par initialiser aléatoirement un certain nombre d'exemples représentatifs. Les clusters sont alors créés en attribuant les exemples au cluster ayant le représentant le plus proche. Par la suite, l'algorithme vise à améliorer la qualité du clustering par l'ajout ou la suppression d'un exemple de l'ensemble des représentants. L'algorithme s'arrête lorsqu'aucune amélioration au niveau de la fonction de pertinence ne peut être réalisée. Le troisième algorithme proposé par ces auteurs est TDS (Top Down Splitting). Cet algorithme suit une approche descendante. Il commence par un seul cluster (i.e le cluster racine qui contient tous les exemples). Ensuite, il divise d'une manière récursive les clusters (si cette division n'entraîne pas une augmentation de la valeur de la fonction de pertinence) en remplaçant le medoid du cluster par deux medoids : Le premier (respectivement le deuxième) medoid correspond au medoid de la classe la plus fréquente (respectivement la deuxième classe fréquente) dans le cluster. Le dernier algorithme proposé par Eick et al. dans [18] est l'algorithme SCEC (Supervised Clustering using Evolutionary Computing). Cet algorithme utilise les techniques évolutionnistes pour trouver l'ensemble des représentants optimal.

- Vilalta et al [12, 19] mais aussi Wu et al. [20] utilisent eux la technique appelée "décomposition des classes. Cette technique contient deux étapes principales : (1) réalisation d'un clustering de type k-moyennes (où k est selon les auteurs une entrée ou une sortie de l'algorithme) par groupe d'exemples qui appartiennent à la même classe j . Le nombre de clusters peut différer par classe. On obtient alors P clusters au total ($P = \sum_j k_j$). (2) Entraînement d'un classifieur sur les P classes résultantes et interprétation des résultats. Cette technique permet aussi l'amélioration des classifieurs linéaires (simples).

Généralement, l'objectif de ces algorithmes est de minimiser le nombre de clusters tout en maximisant la pureté des classes dans ces clusters. La différence entre le clustering standard (non supervisé) et le clustering supervisé est donnée par la figure 4.

Comme dans le cas des arbres de décision, ces algorithmes

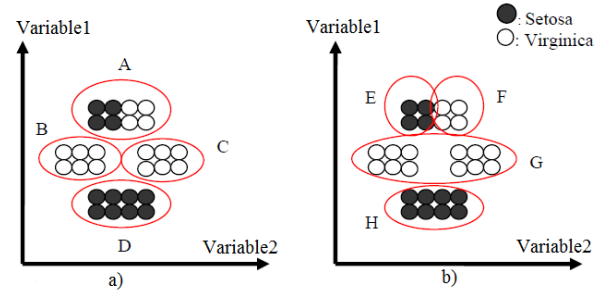


FIGURE 4 – La différence entre le clustering standard a) et le clustering supervisé b)

ne permettent pas une découverte complète de la structure interne de la variable cible (voir figure 4 b)) : L'algorithme fusionne les deux groupes B et C malgré qu'ils ont des caractéristiques différentes.

3.3 Positionnement

L'objectif de la thèse est de chercher un modèle qui prend en considération les points énumérés en fin de Section 2 :

1. La découverte de la structure interne de la variable cible (la proximité entre les individus).
2. La maximisation de la performance prédictive du modèle.
3. L'interprétation des résultats.
4. la minimisation des connaissances a priori requises de la part de l'utilisateur (i.e peu de paramètres d'entrée).

Comme une première piste de travail, nous nous intéressons à modifier l'algorithme des K-moyennes. Cet algorithme figure parmi les algorithmes de clustering les plus répandus et le plus efficace en rapport temps de calcul et qualité [21]. Les différentes étapes de l'algorithme des K-moyennes modifié sont présentées dans l'algorithme 1.

-
- 1) Prétraitement des données
 - 2) Initialisation des centres
 - 3) Répéter un certain nombre de fois (R) jusqu'à convergence
 - 3.1 Coeur de l'algorithme
 - 4) Choix de la meilleure convergence
 - 5) Mesure d'importance des variables (après la convergence et sans réapprendre le modèle)
 - 6) Prédiction la classe des nouveaux exemples.
-

Algorithme 1 : K-moyennes.

3.4 Travaux réalisés

Etape 1 des k-moyennes (voir Algorithme 1) : Généralement, la tâche de clustering nécessite une étape de prétraitement non supervisé afin de fournir des clusters

intéressants (pour l'algorithmes des K-moyennes voir par exemple [22] et [23]). Cette étape de prétraitement peut empêcher certaines variables de dominer lors du calcul des distances. En s'inspirant de ce résultat, nous avons montré dans [24] (à travers l'évaluation de deux approches : conditional Info et Binarization) que l'utilisation d'un prétraitement supervisé peut aider l'algorithme des K-moyennes standard à atteindre une bonne performance prédictive. La prédiction de la classe étant basée sur l'appartenance au cluster le plus proche après la fin de convergence puis sur un vote majoritaire dans le cluster concerné.

Etape 5 des k-moyennes (voir Algorithme 1) : Dans [25] nous avons proposé une méthode supervisée pour mesurer l'importance des variables après la convergence du modèle. L'importance d'une variable est mesurée par son pouvoir prédictive à prédire l'ID-cluster. Autrement dit, une variable est importante si elle peut prédire la partition générée par le modèle de classification à base de clustering. Dans cette étude, nous n'avons pas considéré les interactions qui peuvent existées entre les variables (e.g une variable n'est importante qu'en présence des autres). Ceci, fait l'objet des futurs travaux (parmi d'autres).

3.5 Travaux en cours

Etape 2 des k-moyennes (voir Algorithme 1) : Dans [26] nous avons proposé une nouvelle méthode d'initialisation des k-moyennes dans le cas supervisé : le cas où la valeur de K correspond au nombre de classes à prédire (C). Cette méthode est basée sur l'idée de la décomposition des classes après avoir prétraité les données à l'aide de la méthode proposée dans [24]. A l'avenir nous comptons étendre cette méthode lorsque $K \neq C$.

3.6 Travaux à venir

Pour la deuxième partie de la thèse en cours (la thèse est à sa moitié exactement) on s'intéressera à traiter :

Etape 4 des k-moyennes (voir Algorithme 1) : L'algorithme des k-moyennes n'assurant pas de trouver un minimum global. De ce fait, il est souvent exécuté plusieurs fois (on parle de "réplicates") et la meilleure solution en terme d'erreur est alors choisie. Dans le cadre de la classification à base de clustering, le critère utilisé pour choisir la meilleure réplique doit prendre en considération le compromis homogénéité - pureté des clusters. Dans les travaux à venir, on cherchera à définir un nouveau critère pour le choix de la meilleure "réplicates" dans le cas supervisé.

Etape 5 des k-moyennes (voir Algorithme 1) : La fonction objective utilisée dans l'algorithme des k-moyennes standard consiste à minimiser l'inertie intra et par conséquent maximiser l'inertie inter cluster. Dans les travaux à venir, on cherchera un nouveau critère à optimiser pour pouvoir

atteindre l'objectif d'un algorithme à base de clustering.

Etape 6 des k-moyennes (voir Algorithme 1) : par défaut la classe prédite est la classe majoritaire présente dans le cluster. On cherchera à améliorer ce point comme cela existe déjà pour les arbres de décisions.

Si l'ensemble de ces travaux aboutissent on aura alors un algorithme complet de "classification à base de clustering".

Références

- [1] Pages, J.P., Cailliez, F., Escoufier, Y. Analyse factorielle : un peu d'histoire et de géométrie. *Revue de Statistique Appliquée, Vol XXVII*, pp. 5-28, 1979.
- [2] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering : A review. *ACM Comput. Surv.*, 31(3) :264-323, September 1999.
- [3] Antoine Cornuéjols, Laurent Miclet. Apprentissage artificiel : concepts et algorithmes, Eyrolles 2010.
- [4] John Ross Quinlan. C4. 5 : programs for machine learning, volume 1. Morgan kaufmann, 1993.
- [5] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. Classification and regression trees. *CRC press*, 1984.
- [6] Vladimir N. Vapnik. The Nature of Statistical Learning Theory. *Springer-Verlag New York, Inc., New York, NY, USA*, 1995.
- [7] S. B. Kotsiantis. Supervised machine learning : A review of classification techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering : Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3-24, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press.
- [8] Simon Haykin. Neural Networks : A Comprehensive Foundation. *Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition*, 1998.
- [9] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3) pp. 389-422, 2002.
- [10] Monirul Kabir, Md Monirul Islam, and Kazuyuki Murase. A new wrapper feature selection approach using neural network. *Neurocomputing*, 73(16) pp. 3273-3283, 2010.
- [11] Glenn Fung, Sathyakama Sandilya, and R. Bharat Rao. Rule extraction from linear support vector machines. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 32-40. ACM, 2005.
- [12] Ricardo Vilalta, and Irina Rish. A Decomposition of Classes via Clustering to Explain and Improve Naive

- Bayes. *ECML, volume 2837 of Lecture Notes in Computer Science*, pp. 444-455. Springer, 2003
- [13] Sami H Al-Harbi and Victor J Rayward-Smith. Adapting k-means for supervised clustering. *Applied Intelligence*, 24(3), pp. 219-226, 2006.
- [14] Aguilar J., Roberto Ruiz, José C Riquelme, and Raúl Giráldez. Snn : A supervised clustering algorithm. *In Engineering of Intelligent Systems*, pp. 207-216. Springer, 2001.
- [15] Noam Slonim and Naftali Tishby. Agglomerative information bottleneck. *MIT Press*, pp 617-623, 1999.
- [16] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 1999.
- [17] Hakan Cevikalp, Diane Larlus, and Frederic Jurie. A supervised clustering algorithm for the initialization of rbf neural network classifiers. *In Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th, pages 1â€š4. IEEE, 2007.*
- [18] Christoph F Eick, Nidal Zeidat, and Zhenghong Zhao. Supervised clustering algorithms and benefits. *In Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*, pp. 774-776. IEEE, 2004.
- [19] Francisco Ocegueda-Hernandez and Ricardo Vilalta. An Empirical Study of the Suitability of Class Decomposition for Linear Models : When Does It Work Well ? *In SIAM* 2013.
- [20] Junjie Wu, Hui Xiong and Jian Chen. COG : local decomposition for rare class analysis *In Data Min Knowl Disc (DMKD)*, pp. 191-220. Springer, 2009.
- [21] Jain, Anil K. Data Clustering : 50 Years Beyond K-means. *Pattern Recogn. Lett, Elsevier Science Inc*, pp. 651-666, 2009
- [22] Milligan, G., Cooper, M. A study of standardization of variables in cluster analysis. *In : Journal of Classification, Springer-Verlag.*, 5(2) pp.181-204,1988
- [23] Celebi E. M., Hassan A. Kingravi, Patricio A. Vela. A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm. *Journal of Expert Systems with Applications*, 40(1) pp.200-210, 2013
- [24] Alaoui Ismaili O., Lemaire V., Cornu ejols A. Supervised pre-processings are useful for supervised clustering. *Springer Series Studies in Classification, Data Analysis, and Knowledge Organization*, 2015
- [25] Alaoui Ismaili O., Lemaire V., and Cornu ejols A. A supervised methodology to measure the variables contribution to a clustering. *In International Conference on Neural Information Processing (ICONIP), Kuching, Sarawak, Malaysia*, 2014
- [26] Vincent Lemaire, Oumaima Alaoui Ismaili, Antoine Cornu ejols "An Initialization Scheme for Supervized K-means", *to appear in International Joint Conference on Neural Networks (IJCNN), IEEE, Ireland*, 2015