

Une approche Bayésienne pour l'apprentissage de règles séquentielles de classification

Elias Egho¹, Dominique Gay¹, Marc Boullé¹, Nicolas Voisine¹, Fabrice Clérot¹

ORANGE LABS, Lannion, France
pre.nom@orange.com

Résumé : Nous proposons un nouveau cadre pour extraire des règles séquentielles de classification. Après avoir formalisé l'espace de règles à explorer, nous suggérons un nouveau critère Bayésien pour évaluer l'intérêt des règles séquentielles et un algorithme d'extraction efficace et libre de tout paramètre utilisateur. Les expériences montrent que (i) le nouveau critère identifie les règles intéressantes et robustes ; (ii) l'utilisation des règles extraites comme nouveaux descripteurs de base de données permet d'atteindre de meilleures performances prédictives que les méthodes concurrentes basées sur les règles séquentielles de classification.

Mots-clés : Fouille de données séquentielles, Règles de Classification

1 Introduction

La classification supervisée de données séquentielles (Xing *et al.*, 2010) est pertinente pour un large spectre de domaines d'applications, pour la biologie (Deshpande & Karypis, 2002), l'analyse de textes (Sebastiani, 2002) ou encore l'analyse de données web (Tan & Kumar, 2002). La classification supervisée basée sur la fouille de règles séquentielles est très populaire en raison de leur interprétabilité. Une règle séquentielle est une expression de la forme $\pi : s \rightarrow c_i$ où s est une séquence, prémisse de la règle, et c_i une valeur d'attribut classe. π peut s'interpréter comme : "lorsqu'un objet respecte la séquence d'évènements s alors c'est souvent un objet de classe c_i ". Ainsi, un nouvel objet entrant qui supporte π se verra affecté la classe c_i .

La plupart des Classifieurs Basés sur les Règles Séquentielles (CBRS) adoptent une stratégie similaire : d'abord, un ensemble de règles est extrait en fonction d'une mesure d'intérêt, puis, soit un classifieur dédié, e.g., une liste de décisions, est construit à partir d'un sous-ensemble de règles (Tseng & Lee, 2005; Zhou *et al.*, 2013), soit les règles sont directement utilisées comme nouveaux descripteurs dans un processus de classification supervisée (Lesh *et al.*, 1999; Deng & Zaïane, 2010; Lam *et al.*, 2012). Si les approches existantes ont généralement de bonnes performances prédictives, on peut toutefois mettre en évidence deux de leurs faiblesses :

L'instabilité des mesures d'intérêt. Considérons deux mesures communément utilisées pour l'évaluation de règles de classification, la confiance (Lesh *et al.*, 1999; Tseng & Lee, 2005) et le taux d'accroissement (Growth Rate : GR) (Ji *et al.*, 2005; Deng & Zaïane, 2010). Sur la figure 1, nous montrons que les règles extraites en fonction de ces mesures ne sont pas individuellement

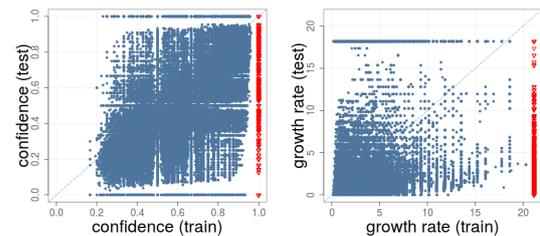


FIGURE 1 – Comparaison des valeurs de confiance et GR des règles séquentielles de classification dans une expérience 50%-apprentissage/ 50%-test pour les données skater. ∇ pour le top-1000 des règles en fonction de leur valeur de confiance, resp. GR, sur les données d'apprentissage.

robustes : ici, pour la base de données skater (Mörchen & Ultsch, 2007), chaque point représente une règle, son abscisse est sa valeur pour la mesure sur la base de données d'apprentissage et son ordonnée sa valeur en test. Nous observons un nuage de points très diffus dans le repère apprentissage/test, ce qui indique que les valeurs de confiance et de GR sont très instables des données d'apprentissage aux données test.

La malédiction du paramétrage. la plupart des approches existantes nécessitent l'ajustement de paramètres : un seuil de fréquence minimum et/ou un seuil pour la mesure d'intérêt lors de la phase d'extraction ; puis le nombre de règles à utiliser dans l'ensemble final utilisé. Malheureusement, le paramétrage est crucial puisque les performances des classifieurs en est fortement dépendant (Coenen & Leng, 2007) mais il n'est pas tâche facile – en particulier chaque domaine d'application peut nécessiter son propre paramétrage.

Dans cet article, nous proposons une solution à ces deux problèmes en suggérant (i) un nouveau critère robuste pour l'évaluation des règles, (ii) un algorithme d'extraction de règles efficace et libre de tout paramètre utilisateur. L'évaluation empirique en section 4 démontre la pertinence de notre approche : le critère permet d'identifier les règles intéressantes et robustes, et l'utilisation des règles extraites comme nouveaux descripteurs nous permet d'obtenir de meilleures performances prédictives que l'état de l'art des CBRS.

2 Contexte & Définitions

Soit $\mathcal{I} = \{e_1, e_2, \dots, e_m\}$ un ensemble fini de m items distincts. Une **séquence** s sur \mathcal{I} est une liste ordonnée $s = \langle s_1, \dots, s_{\ell_s} \rangle$ où $s_i \in \mathcal{I}$; ($1 \leq i \leq \ell_s, \ell_s \in \mathbb{N}$). Une séquence $s' = \langle s'_1 \dots s'_{\ell_{s'}} \rangle$ est une **sous-séquence** de $s = \langle s_1 \dots s_{\ell_s} \rangle$, qu'on note $s' \preceq s$, si il existe des indices $1 \leq i_1 < i_2 < \dots < i_{\ell_{s'}} \leq \ell_s$ tels que $s'_z = s_{i_z}$ pour tout $z = 1 \dots \ell_{s'}$ et $\ell_{s'} \leq \ell_s$. On dit que s est une **super-séquence** de s' . $\mathbb{T}(\mathcal{I})$ représente l'ensemble de toutes les séquences possibles sur \mathcal{I} . Soit $\mathcal{C} = \{c_1, \dots, c_j\}$ un ensemble fini fini de j classes distinctes. Une **base de données séquentielles labellisées** \mathcal{D} est un ensemble fini de n triplets (sid, s, c) , où sid est un identifiant de séquence, s une séquence ($s \in \mathbb{T}(\mathcal{I})$) et c une valeur de classe ($c \in \mathcal{C}$). L'ensemble $\mathcal{D}_{c_i} \subseteq \mathcal{D}$ contient les n_{c_i} séquences de même classe c_i . Le **support** ou la **fréquence** d'une séquence s dans \mathcal{D} , noté $f(s)$, est défini par $f(s) = |\{(sid', s', c') \in \mathcal{D} | s \preceq s'\}|$. Nous dénotons la valeur $n - f(s)$ par $\bar{f}(s)$; et le support de s dans \mathcal{D}_c est noté $f_c(s)$ alors que $\bar{f}_c(s)$ vaut $n_c - f_c(s)$.

Definition 1 (Modèle de règle séquentielle de classification (MRSC))

Soit \mathcal{D} une base de données séquentielles labellisées à j classes. Une règle séquentielle de classification π est une expression de la forme : $\pi : s \rightarrow f_{c_1}(s), f_{c_2}(s), \dots, f_{c_j}(s)$ où s est une séquence, appelée prémisses de la règle et $f_{c_i}(s)$ est le support de s dans chaque \mathcal{D}_{c_i} , $i = 1 \dots j$. L'espace des modèles à explorer est l'ensemble de tous les MRSCs ainsi définies. Chacune d'entre elles est caractérisée par la hiérarchie de paramètres suivante : les items qui constituent la prémisses de la règle, l'ordre d'occurrence des items dans la prémisses et la distribution des classes dans la prémisses (dans le corps) et hors de la prémisses (hors du corps).

Un cadre Bayésien pour l'évaluation de motifs séquentiels. Considérant le cadre Bayésien introduit par Boullé (2006), le problème de fouille de règles peut être reformulé comme un problème de sélection de modèles. Afin de choisir le "meilleur" modèle de règle dans l'espace

de modèle, nous utilisons une approche Bayésienne, dite Maximum A Posteriori (MAP) : nous cherchons à maximiser $p(\pi|\mathcal{D})$, la probabilité a posteriori d'une règle π connaissant les données \mathcal{D} . Considérant la règle Bayes, $p(\pi|\mathcal{D})$ prend la forme suivante :

$$p(\pi|\mathcal{D}) = \frac{p(\pi, \mathcal{D})}{p(\mathcal{D})} = \frac{p(\pi) \times p(\mathcal{D}|\pi)}{p(\mathcal{D})}$$

Puisque $p(\mathcal{D})$ est constant dans le problème d'optimisation courant, cela revient à maximiser l'expression $p(\pi) \times p(\mathcal{D}|\pi)$. Le critère d'évaluation, appelé *cost*, est défini par le logarithme négatif de $p(\pi|\mathcal{D})$ et s'exprime comme suit :

$$cost(\pi) = -\log(\underbrace{p(\pi)}_{\text{prior}} \times \underbrace{p(\mathcal{D} | \pi)}_{\text{vraisemblance}}) \propto -\log(\underbrace{p(\pi | \mathcal{D})}_{\text{posterior}})$$

En considérant (i) un a priori hiérarchique sur les paramètres d'une règle (comme indiqué dans sa définition) pour $p(\pi)$ et (ii) une distribution multinomiale pour la probabilité des données connaissant le modèle de règle ($p(\mathcal{D} | \pi)$), il est possible d'obtenir une expression exacte de la fonction *cost*. Le manque de place ne nous permet pas de détailler chaque terme de la fonction *cost*, toutefois, notons que ce critère s'obtient de manière similaire pour les règles de classification dans les données transactionnelles (Gay & Boullé, 2012). L'expression complète et exacte de la fonction de *cost* pour une règle est alors :

$$\begin{aligned} cost(\pi) = & \log(m + 1) + \log(\ell_{max} + 1) + \log\left(\frac{m^{k_s}}{k_s!}\right) + \log(k_s^{\ell_s}) \\ & + \log\left(\binom{f(s) + j - 1}{j - 1}\right) + \log\left(\binom{\overline{f(s)} + j - 1}{j - 1}\right) \\ & + \log(f(s)!) - \sum_{i=1}^j \log(f_{c_i}(s)!) + \log(\overline{f(s)}!) - \sum_{i=1}^j \log(\overline{f_{c_i}(s)}!) \end{aligned}$$

où ℓ_s est le nombre d'items dans la séquence s , k_s le nombre d'items distincts de la séquence s , ($k_s \leq \ell_s$) et ℓ_{max} le nombre items de la plus longue séquence de \mathcal{D} .

Comme l'amplitude des valeurs de *cost* dépend du nombre n de séquences et du nombre m d'items dans les données, nous suggérons un critère normalisé, appelé *level*, qui sera notre mesure d'intérêt pour évaluer et comparer des règles.

Definition 2 (Level)

Le *level* d'une règle π est défini par : $level(\pi) = 1 - \frac{cost(\pi)}{cost(\pi_\emptyset)}$ où π_\emptyset est le modèle de règle nul (i.e., la règle par défaut avec une prémisse vide).

Le *level* permet d'identifier la frontière dans l'espace de règles entre les règles pertinentes et celles qui ne le sont pas. En effet, les règles π telles $level(\pi) \leq 0$ sont moins probables que la règle nulle π_\emptyset . Leur utilisation pour caractériser les classes est plus coûteux, en terme de *cost*, que π_\emptyset : de telles règles sont considérées comme non-pertinentes. Au contraire, les règles telles que $0 < level(\pi) \leq 1$ constituent les motifs intéressants. En fait, les règles à fort *level* (ou faible *cost*) sont les plus probables connaissant les données et indiquent qu'il y a une corrélation entre la prémisse de la règle et l'attribut classe.

3 Stratégie et algorithme d'extraction de règles séquentielles de classification

L'extraction de motifs séquentiels (Agrawal & Srikant, 1995) est un problème NP-difficile. Bien que les propriétés du seuil minimum de fréquence et des représentations condensées (Manila & Toivonen, 1996) permettent de réduire le temps de calcul, le problème reste difficile pour les grandes bases de données. Comme le *level* ne possède pas de propriétés (comme celles de la fréquence) pour faciliter l'extraction de motifs, une exploration exhaustive pour obtenir les règles à *level* positif est inenvisageable. En effet, la taille de l'espace de recherche est exponentiel en m , le nombre d'items : $\sum_{i=1}^{\ell_{max}} m^i \equiv O(m^{\ell_{max}})$. Pour cette première approche, nous préférons donc une formulation plus simple et réaliste du problème, à savoir "Extraire avec diversité un sous-ensemble de règles à *level* positifs".

Notre algorithme d'extraction de règles, appelé **MiSeRe** (pour *Mining Sequential Classification Rules*), est libre de tout paramètre utilisateur, il emploie une stratégie aléatoire de recherche basée sur les objets séquences qui promeut la diversité et enfin il est dit "any-time" – plus on lui accorde de temps de calcul, plus il apprend. La principale difficulté est de calculer efficacement les distributions de classes $f_{c_1}(s), \dots, f_{c_j}(s)$. Pour ce faire, nous utilisons la représentation en BitSet des séquences et des opérations Booléennes présentées par Ayres *et al.* (2002). Ainsi, extraire une règle est effectué en $O(\ell_s \times n \times \log(n))$.

Algorithme 1 : MiSeRe

```

input :  $\mathcal{D}$ , une base de données séquentielles labellisées
output :  $\mathcal{R}$ , un ensemble de règles à level positif
1  $S = \{s = \langle s_1 \rangle; s_1 \in \mathcal{I}\}$ ;
2  $\mathcal{R} = \{\pi : s \rightarrow f_{c_1}(s), \dots, f_{c_j}(s); s \in S \wedge \text{level}(\pi) > 0\}$ ;
3 while  $\neg$  TempsLimite do
4    $s =$  ChoisirSéquenceAléatoire( $\mathcal{D}$ );
5    $ads =$  CalculerNombreSousSéquences( $s$ );
6   for  $i = 1$  to  $\log(ads)$  do
7      $s' =$  GénérerSousSéquencesAléatoires( $s$ );
8      $\pi : s' \rightarrow f_{c_1}(s'), \dots, f_{c_j}(s')$ ;
9     if  $\text{level}(\pi) > 0 \wedge \pi \notin \mathcal{R}$  then
10       $\mathcal{R} = \mathcal{R} \cup \{\pi\}$ 
11 return  $\mathcal{R}$ ;
```

Procédure de classification. Nous suggérons d'utiliser les règles extraites par MiSeRe comme nouveaux descripteurs pour recoder la base de données séquentielles initiale \mathcal{D} en une base de données transactionnelles binaires. Un nouvel attribut binaire est créé pour chaque règle π et prend la valeur 1 pour un objet (sid, s, c) si s est une super-séquence de la prémisse de π et 0 sinon. Cette procédure présente deux avantages : (i), il existe un grand nombre de classifieurs disponibles applicables à ces données recodées ; (ii), dans certains cas d'étude, les séquences ne sont qu'une partie de la description des objets à classer en plus d'attributs numériques ou catégoriels classiques : ainsi, remplacer la partie séquentielle des données par des attributs binaires pertinents enrichit les données avant l'utilisation d'un classifieur.

4 Validation expérimentale

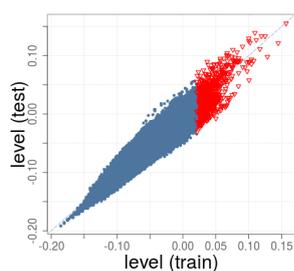
Nous présentons un résumé des résultats de l'évaluation empirique de notre approche. Une visualisation détaillée et interactive des résultats ainsi que le code JAVA de MiSeRe sont disponibles sur le site qui accompagne ce travail (Egho *et al.*, 2015) : <http://misere.co.nf>. Les expériences sont destinées à nourrir la discussion autour des points suivants : la stabilité du *level*, les performances prédictives de notre procédure de classification, la comparaison des performances prédictives de MiSeRe avec l'état de l'art des CBRS.

Pour l'évaluation empirique, nous utilisons 11 bases de données réelles communément utilisées par la littérature en classification de séquences. Ces données présentent une large variété en termes de nombre de séquences, d'items, de longueur de séquences et de classes et sont issues de divers domaines d'applications (voir Table 1).

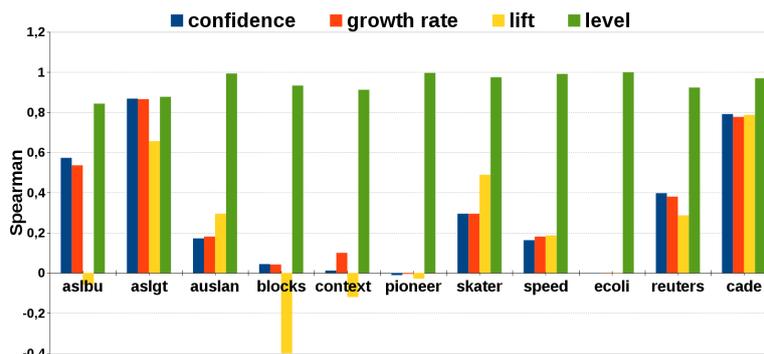
Stabilité du critère *level*. Pour évaluer la stabilité d'une mesure d'intérêt, nous menons des expériences apprentissage / test : chaque base est divisé en deux parties égales, 50% pour l'apprentissage en fonction de la mesure d'intérêt et 50% pour le test pour évaluer la mesure. Pour chaque règle extraite les valeurs de mesure en apprentissage et en test sont comparées. Nous extrayons les motifs séquentiels fréquents de la base d'apprentissage avec CSPADE (Zaki, 2000) (seuil de fréquence minimum de 2% et gap maximum entre items de 2). Des règles sont construites à partir de ces motifs et nous calculons les valeurs de *level*, confiance, GR et lift de ces règles en apprentissage et en test. Remarquons qu'en figure 1, relatif à l'exemple d'introduction, les valeurs de *level* des règles extraites sont sensiblement plus stables que la confiance et GR. Les mêmes observations tiennent pour les autres bases de données. Pour une vue globale de la stabilité des mesures étudiées, nous étudions la concordance des rangs des valeurs des mesures dans une expérience apprentissage/test. Pour chaque base, nous ordonnons les règles en fonction de leur valeur de mesure en apprentissage puis en test. Ensuite, la concordance des rangs d'apprentissage et de test est analysée au moyen du coefficient de corrélation de Spearman (Myers & Well, 2003). La figure 2b montre que les rangs de valeurs *level* en apprentissage et en test sont très corrélés (valeur de coefficient proche de 1) ; le *level* est donc stable alors que les autres mesures sont très instables (faible corrélation des rangs en apprentissage/test).

Données	# Séquences	# Items	Longueur Max	# Classes
<i>aslbu</i>	441	140	58	7
<i>aslgt</i>	3493	47	197	40
<i>auslan</i>	200	12	24	10
<i>blocks</i>	210	8	24	8
<i>context</i>	240	54	272	5
<i>pioneer</i>	160	92	127	3
<i>skater</i>	530	41	260	6
<i>speed</i>	530	41	260	7
<i>ecoli</i>	106	16	28	2
<i>reuters</i>	5459	14529	533	8
<i>cade</i>	15000	111766	19763	12

TABLE 1 – Description des données expérimentales.



(a) Résultat de l'expérience apprentissage/test pour le *level* sur les données de la figure 1 en introduction.



(b) Analyse de la concordance des rangs par le coefficient de Spearman dans une expérience apprentissage/test.

FIGURE 2 – Etude comparative de la stabilité des critères *level*, confiance, GR et lift.

Pour confirmer la robustesse du *level*, nous menons l'expérience suivante : pour chaque base de données, nous assignons une classe $c \in \mathcal{C}$ de manière aléatoire à chaque objet, puis nous

laissons tourner MiSeRe pendant 30 minutes : aucune règle n’est extraite puisque qu’elles ont toutes un *level* négatif. Au contraire, pour la moitié des bases disponibles, certaines règles à fort GR, confiance ou lift sont extraites alors qu’il n’y a rien à découvrir. Ainsi, le *level* est bien une mesure robuste, évite les motifs non-pertinents et le sur-apprentissage.

Performance prédictive de notre approche. Pour évaluer les performances prédictives de notre approche, nous utilisons plusieurs classifieurs standards sur les données recodées avec les règles obtenues avec MiSeRe : le Bayésien naïf (NB), la forêt aléatoire (RF), l’arbre de décision (C4.5 alias J48), une machine à vecteurs supports (SVM), deux approches de k plus proches voisins k Star et IBk accessibles avec WEKA (Hall *et al.*, 2009) – que nous utilisons avec les paramètres par défaut. Nous utilisons aussi le Bayésien naïf sélectif¹ (SNB) de Boullé (2007). Les résultats de précision sont obtenus par validation croisée stratifiée en 10 parties : bien sûr MiSeRe n’exploite que la partie apprentissage. Bien que MiSeRe soit de type “anytime”, pour ces expériences, nous faisons le choix (que nous motivons par la suite) d’extraire 2^{10} , i.e., 1024 règles. Les résultats détaillés de précision moyenne sont disponibles dans (Egho *et al.*, 2015). Pour comparer les performances de plusieurs classifieurs sur plusieurs bases de données nous utilisons le test de Friedman couplé au test post-hoc de Nemenyi (au seuil de significativité $\alpha = 0.05$), suggérés par Demšar (2006). L’hypothèse nulle est rejetée, indiquant que les classifieurs comparés en sont pas équivalents en termes de précision. Le résultat de cette expérience est résumé par le diagramme de différence critique de performance en figure 3, où l’on peut observer le rang moyen de chaque classifieur sur les 11 bases de données. Bien qu’aucun classifieur ne soit meilleur que tous les autres, nous pouvons identifier deux groupes : {SVM, J48, k Star, IBk , NB} au sein duquel il n’y a pas de différence significative de performance, et {SNB, RF}, performances sont meilleures que les autres bien qu’on ne puisse pas dire qu’ils sont significativement meilleurs que SVM. Notre recommandation est donc de combiner MiSeRe avec soit le SNB ou la RF. Puisque le SNB est libre de tout paramètre utilisateur, c’est celui que nous utiliserons par la suite.

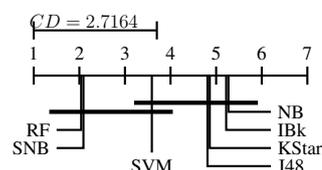


FIGURE 3 – Diagramme de différence critique de performance entre divers classifieurs sur des données recodées par les SCRMs extraites par MiSeRe.

Efficacité de MiSeRe. Notre méthode d’extraction, *MiSeRe*, est contrôlée par une contrainte de temps de calcul pendant lequel un certain nombre de règles est extrait. Nous étudions l’évolution de la performance prédictive de *SNB-MiSeRe* en terme de précision en fonction du nombre de règles extraites (pour un nombre de règles $\rho = 2^\alpha$; $\alpha \in [0; 14]$, voir Figure 4). Nous observons que la performance augmente avec le nombre de règles, puis se stabilise après quelques dizaines de règles pour les petites bases et quelques centaines (voire 1024) pour les plus grandes bases. De manière générale, après quelques centaines de règles, le système *SNB-MiSeRe* atteint ces performances maximales. Il en est de même pour les autres classifieurs couplés à MiSeRe (Egho *et al.*, 2015).

Nous étudions aussi l’efficacité de notre algorithme *MiSeRe*. La figure 4b présente le temps de calcul en fonction du nombre de règles extraites. Pour la plupart des bases de données, l’extraction d’un millier de règles est effectuée en moins de 80 secondes. Pour la base *ecoli*,

1. <http://www.khiops.com>

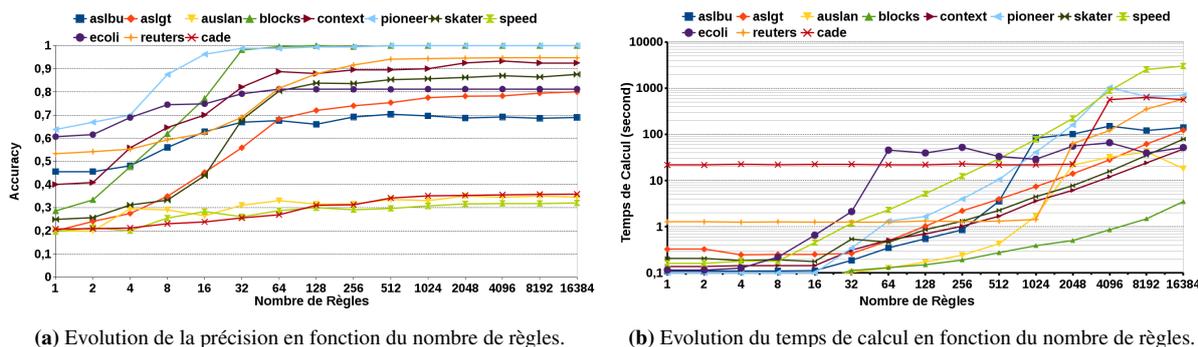


FIGURE 4 – Evolution des performances de MiSeRe.

seulement 53 règles à *level* positif sont extraites, c’est pourquoi *MiSeRe* a un temps d’exécution fixe ($\simeq 60s$ après 53 règles). Pour les données *cade*, de type texte, le temps d’exécution est d’environ 20s jusqu’à 2048 règles. En effet, la base contient 111776 mots distincts, donc les 2048 règles extraites ont une prémisse singleton. Ce temps de calcul correspond à la première étape (Lignes 1-2) de l’algorithme 1.

MiSeRe et l’état de l’art. Nous comparons les performances prédictives de *SNB-MiSeRe* avec quatre classifieurs à base de règles séquentielles de l’état de l’art : SCII Match, SCII CBA (Zhou *et al.*, 2013), BayesFM (Lesh *et al.*, 1999) et CBS (Tseng & Lee, 2005). Puisque les concurrents ne fournissent que des résultats de précision, ce sera notre critère de comparaison de performance. Les expériences sont menées avec les paramètres utilisés dans les articles originaux et les résultats de précision moyenne sont obtenus par validation croisée stratifiée en 10 parties. Les résultats détaillés sont accessibles dans (Egho *et al.*, 2015). A nouveau, nous appliquons le test de Friedman couplé au test post-hoc de Nemenyi. Le résultat de ces tests est présenté en figure 5. Le diagramme de différence critique de performance est sans appel : il y a bien une différence significative de performance entre *SNB-MiSeRe* qui est classé premier et ses concurrents. On observe les mêmes résultats pour *MiSeRe* couplé avec les autres classifieurs standards utilisés précédemment (Egho *et al.*, 2015).

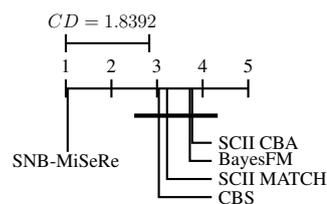


FIGURE 5 – Diagramme de différence critique de performance entre *SNB-MiSeRe* et ses concurrents.

5 Conclusion

Nous proposons une solution complète pour la classification supervisée à base de règles. Dans un cadre Bayésien, nous définissons un nouveau critère d’évaluation pour les règles séquentielles ; nous présentons aussi *MiSeRe*, un algorithme efficace pour extraire des règles pertinentes et robustes. L’usage de ces règles combiné à des classifieurs standards démontre de meilleures performances prédictives que les méthodes de l’état de l’art basées sur les règles. Notons que le critère suggéré est indépendant de l’algorithme *MiSeRe* et peut être donc appliqué aux règles séquentielles extraites par un autre algorithme. De plus, la généralité de la méthode employée nous permet d’envisager des extensions du nouveau critère pour des données séquentielles plus complexes, comme des séquences multi-dimensionnelles.

Références

- AGRAWAL R. & SRIKANT R. (1995). Mining sequential patterns. In *ICDE'95*, p. 3–14.
- AYRES J., FLANNICK J., GEHRKE J. & YIU T. (2002). Sequential pattern mining using a bitmap representation. In *KDD'02*, p. 429–435 : ACM.
- BOULLÉ M. (2006). MODL : A bayes optimal discretization method for continuous attributes. *Machine Learning*, **65**(1), 131–165.
- BOULLÉ M. (2007). Compression-based averaging of selective naive bayes classifiers. *Journal of Machine Learning Research*, **8**, 1659–1685.
- COENEN F. & LENG P. H. (2007). The effect of threshold values on association rule based classification accuracy. *Data & Knowledge Engineering*, **60**(2), 345–360.
- DEMŠAR J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, **7**, 1–30.
- DENG K. & ZAĀANE O. R. (2010). An occurrence based approach to mine emerging sequences. In *DaWaK'10*, p. 275–284.
- DESHPANDE M. & KARYPIS G. (2002). Evaluation of techniques for classifying biological sequences. In *PAKDD'02*, p. 417–431.
- EGHO E., GAY D., BOULLÉ M., VOISINE N. & CLÉROT F. (2015). MiSeRe : Mining Sequential Classification Rules. <http://misere.co.nf>.
- GAY D. & BOULLÉ M. (2012). A bayesian approach for classification rule mining in quantitative databases. In *ECML/PKDD'12*, p. 243–259.
- HALL M. A., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P. & WITTEN I. H. (2009). The WEKA data mining software : an update. *SIGKDD Explorations*, **11**(1), 10–18.
- JI X., BAILEY J. & DONG G. (2005). Mining minimal distinguishing subsequence patterns with gap constraints. In *IEEE ICDM'05*, p. 194–201.
- LAM H. T., MOERCHEN F., FRADKIN D. & CALDERS T. (2012). Mining compressing sequential patterns. In *SDM'12*, p. 319–330.
- LESH N., ZAKI M. J. & OGIHARA M. (1999). Mining features for sequence classification. In *ACM SIGKDD'99*, p. 342–346.
- MANNILA H. & TOIVONEN H. (1996). Multiple uses of frequent sets and condensed representations (extended abstract). In *KDD'96*, p. 189–194.
- MÖRCHEN F. & ULTSCH A. (2007). Efficient mining of understandable patterns from multivariate interval time series. *Data Mining and Knowledge Discovery*, **15**(2), 181–215.
- MYERS J. L. & WELL A. D. (2003). *Research Design and Statistical Analysis*. New Jersey : Lawrence Erlbaum Associates.
- SEBASTIANI F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, **34**(1), 1–47.
- TAN P. & KUMAR V. (2002). Discovery of web robot sessions based on their navigational patterns. *Data Mining and Knowledge Discovery*, **6**(1), 9–35.
- TSENG V. S. & LEE C. (2005). CBS : A new classification method by using sequential patterns. In *SDM'05*, p. 596–600.
- XING Z., PEI J. & KEOGH E. J. (2010). A brief survey on sequence classification. *SIGKDD Explorations*, **12**(1), 40–48.
- ZAKI M. J. (2000). Sequence mining in categorical domains : Incorporating constraints. In *CIKM'00*, p. 422–429 : ACM.
- ZHOU C., CULE B. & GOETHALS B. (2013). Itemset based sequence classification. In *ECML/PKDD'13*, p. 353–368.