

Exploration efficace du treillis disjonctif pour l'extraction de représentations concises des motifs fréquents

I. Denden¹, T. Hamrouni^{1,2}, S. Ben Yahia¹

¹ Faculté des Sciences de Tunis, Département des Sciences de l'Informatique, Campus Universitaire 1060, Tunis, Tunisie,

{tarek.hamrouni, sadok.benyahia}@fst.rnu.tn.

² Centre de Recherche en Informatique de Lens, IUT de Lens, Rue de l'Université SP 16, 62307 Lens Cedex, France,

{hamrouni}@cril.univ-artois.fr.

RÉSUMÉ. Dans plusieurs contextes d'extraction relatifs à des applications réelles, le nombre de motifs fréquents pouvant être extraits est très élevé. Un tel phénomène rend l'exploitation et la manipulation de ces informations une tâche quasi-impossible. Dans le but de pallier ce problème, la notion de représentation concise exacte des motifs fréquents a été introduite afin d'offrir aux utilisateurs un ensemble d'éléments, de taille réduite, à partir duquel la régénération sans perte d'information de tous les motifs fréquents est possible. Dans un travail récent, nous avons introduit une nouvelle représentation concise exacte basée sur des motifs particuliers, appelés *fermés disjonctifs*, caractérisés grâce à leurs supports disjonctifs. Nous avons prouvé que dans plusieurs contextes d'extraction la taille de cette représentation est plus réduite que celles des principales représentations concises exactes proposées dans la littérature. Dans cet article, nous introduisons un nouvel algorithme, nommé DCPR_MINER, dédié à l'extraction de la représentation concise basée sur les fermés disjonctifs. L'étude empirique menée sur des contextes "benchmark" prouve que DCPR_MINER est significativement plus rapide que l'algorithme MEP, permettant l'extraction de la représentation basée sur les motifs essentiels fréquents – l'unique autre utilisant le support disjonctif.

ABSTRACT. In many real-life contexts, the number of frequent patterns is often huge. This phenomenon makes the exploitation and the handling of such amount of information by experts very difficult. To palliate this problem, the notion of exact concise representation of frequent patterns was introduced in order to offer a compact set of elements from which we can derive all frequent patterns without information loss. In a recent work, we have introduced a new exact concise representation based on particular elements called *disjunctive closed patterns*, characterized thanks to their disjunctive supports. We have proved that for several contexts its size is more compact than the main concise representations of the literature. In this paper, we introduce a new algorithm, called DCPR_MINER, dedicated to the extraction of the disjunctive closed patterns-based concise representation. The empirical study carried out on benchmark contexts proves that DCPR_MINER significantly outperforms the algorithm MEP, allowing the extraction of the frequent essential patterns-based representation – the unique other one using the disjunctive support.

MOTS-CLÉS : motif fréquent, représentation concise, support disjonctif, motif fermé disjonctif, motif essentiel.

KEYWORDS : frequent pattern, concise representation, disjunctive support, disjunctive closed pattern, essential pattern.

1. Introduction et motivations

Étant donné un ensemble d'items et un ensemble d'objets, le problème d'extraction des motifs fréquents consiste à dégager, à partir d'un contexte d'extraction, les motifs ayant une fréquence d'apparition (*i.e.*, un support conjonctif ou, simplement, support) supérieure à un seuil minimal défini par l'utilisateur [1]. Depuis son introduction, ce problème a sollicité l'intérêt de plusieurs chercheurs vu que les motifs fréquents constituent une mine d'informations permettant aux experts humains de comprendre les corrélations existantes entre les items du contexte. Les résultats expérimentaux ont prouvé que lorsque cette dernière présente des données fortement corrélées ou que le seuil minimal du support est très faible, le nombre de motifs fréquents est très élevé. Ainsi, dans de tels scénarios, la gestion et l'exploitation de cette masse d'information deviennent très difficiles par les experts humains. À cet effet, plusieurs chercheurs se sont penchés sur la définition d'ensembles compacts à partir desquels nous pouvons générer de manière exacte l'ensemble des motifs fréquents munis de leurs supports exacts. De tels ensembles sont nommés *représentations concises exactes des motifs fréquents*.

Dans la littérature, une multitude de représentations concises exactes a été proposée (*e.g.*, [2, 3, 4, 5, 6]). Néanmoins, quelques unes semblent se distinguer selon une caractérisation originale des motifs les constituant. Outre le support conjonctif, ces représentations concises exactes tiennent également compte d'une autre métrique, nommée *support disjonctif*, afin de réduire davantage le nombre de motifs les constituant tout en offrant la possibilité de dériver directement les supports disjonctifs et négatifs des motifs fréquents. Cette classe de représentations regroupe celle basée sur les motifs essentiels fréquents [5] et celle basée sur les fermés disjonctifs [6]. D'un point de vue structurel, les motifs essentiels (*resp.* fermés disjonctifs) sont les éléments minimaux (*resp.* maximaux) des classes d'équivalence associées dans le treillis disjonctif. Les représentations qui en découlent ont été montrées comme étant en relation avec d'autres classes importantes de motifs telles que les règles d'association généralisées, les expressions booléennes, etc [6]. Par ailleurs, une étude empirique menée dans [7] a montré que la représentation basée sur les fermés disjonctifs est souvent plus compacte que les principales représentations de la littérature. Toutefois, les travaux réalisés jusqu'à maintenant [5, 6, 7] n'ont pas évoqué un facteur important : le temps d'extraction de ces représentations. De plus, aucun algorithme n'a été proposé afin d'extraire la représentation concise basée sur les fermés disjonctifs. Afin de combler ces points, nous nous proposons, dans un premier temps, d'introduire un nouvel algorithme dédié à l'extraction de la représentation basée sur les fermés disjonctifs. Dans un second temps, nous menons une série d'expérimentations sur des contextes "benchmark", associées à des applications réelles, dont l'objectif est double : (i) Comptabiliser les temps consommés par les algorithmes d'extraction des représentations basées respectivement sur les motifs essentiels fréquents et les fermés disjonctifs, (ii) Prouver que l'extraction de la représentation basée sur les fermés disjonctifs est plus rapide que celle de la représentation basée sur les motifs essentiels fréquents.

Le reste du papier est organisé comme suit. La section 2 présente les notions de base utilisées tout le long de ce papier. Dans la section 3, nous présentons une synthèse des représentations basées sur l'exploration du treillis disjonctif. La section 4 introduit un nouvel algorithme, nommé DCPR_MINER, permettant l'extraction de la représentation basée sur les fermés disjonctifs. Dans la section 5, nous menons une étude empirique qui prouve l'utilité de l'algorithme proposé. La conclusion et les perspectives font l'objet de la section 6.

2. Notions de base

Dans cette section, nous présentons brièvement l'ensemble des notions de base qui seront utiles dans le reste de cet article.

Définition 1 (Contexte d'extraction) Un contexte d'extraction est un triplet $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$ dans lequel $\mathcal{O}$ et $\mathcal{I}$ sont, respectivement, des ensembles finis d'objets et d'items, et $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{I}$ est une relation binaire entre les objets et les items. Un couple $(o, i) \in \mathcal{R}$ dénote le fait que l'objet $o \in \mathcal{O}$ contient l'item $i \in \mathcal{I}$.

Exemple 1 Le tableau 1 illustre un contexte d'extraction dans lequel l'ensemble des objets est $\mathcal{O} = \{1, 2, 3, 4, 5\}$ et l'ensemble des items est $\mathcal{I} = \{A, B, C, D, E, F\}$.

	A	B	C	D	E	F
1	x	x	x	x		
2	x		x	x	x	
3	x		x		x	x
4	x	x		x	x	x
5		x	x	x		x

Tableau 1. Un exemple d'un contexte d'extraction.

La définition suivante présente les différents supports d'un motif.

Définition 2 [6] (Supports d'un motif) Soient $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$ un contexte d'extraction et I un motif. Nous distinguons trois types de supports correspondants à un motif I :

$$Supp(I) = |\{o \in \mathcal{O} \mid (\forall i \in I, (o, i) \in \mathcal{R})\}|$$

$$Supp(\vee I) = |\{o \in \mathcal{O} \mid (\exists i \in I, (o, i) \in \mathcal{R})\}|$$

$$Supp(\bar{I}) = |\{o \in \mathcal{O} \mid (\forall i \in I, (o, i) \notin \mathcal{R})\}|$$

$Supp(I)$ (resp. $Supp(\vee I)$ et $Supp(\bar{I})$) est nommé support conjonctif (resp. disjonctif et négatif) de I . Dans la suite, le support conjonctif sera simplement noté support.

Exemple 2 Considérons le contexte illustré par le tableau 1. Alors, nous avons $Supp(BC) = |\{1, 5\}| = 2$, $Supp(\vee BC) = |\{1, 2, 3, 4, 5\}| = 5$ et $Supp(\overline{BC}) = |\{\emptyset\}| = 0$.

Définition 3 présente la notion de motif essentiel fréquent.

Définition 3 (motif fréquent / motif essentiel / motif essentiel fréquent) Pour un seuil de support minsup défini par l'utilisateur, un motif I est qualifié de fréquent si $Supp(I) \geq \text{minsup}$. I est qualifié d'essentiel si $Supp(\vee I) > \max\{Supp(\vee I \setminus i) \mid i \in I\}$ [5]. I est un motif essentiel fréquent s'il est à la fois essentiel et fréquent.

Nous notons par $\mathcal{MF}$ (resp. $\mathcal{MEF}$) l'ensemble des motifs fréquents (resp. motifs essentiels fréquents).

Exemple 3 Soit le contexte du Tableau 1 pour minsup = 2. AC est un motif essentiel fréquent. En effet, AC est un motif fréquent puisque $Supp(AC)$, à savoir 3, est supérieur à minsup. De plus, AC est un motif essentiel car $Supp(\vee AC)$, à savoir 5, est strictement supérieur que $\max\{Supp(\vee A), Supp(\vee C)\}$, à savoir 4.

Dans [5], il a été prouvé que l'ensemble des motifs essentiels fréquents est un idéal d'ordre [9]. En effet, si I est un motif essentiel fréquent alors tous ses sous-ensembles sont des motifs essentiels fréquents. Par contre, si I est un motif infréquent ou non-essentiel alors tous ses sur-ensembles sont des motifs non-essentiels ou infréquents. Dans le reste du papier, nous supposons donc que le contexte ne contient que des items fréquents, i.e. que tous les items infréquents ont été élagués. La définition suivante présente la fermeture disjonctive d'un motif.

Définition 4 (Fermeture disjonctive d'un motif) La fermeture disjonctive d'un motif I , notée $h(I)$, est égale à $\max_{\subseteq} \{I' \subseteq \mathcal{I} \mid (I \subseteq I') \wedge (\text{Supp}(\vee I') = \text{Supp}(\vee I))\}$.

$h(I)$ est ainsi le motif maximal, selon la taille, contenant I et ayant le même support disjonctif que I .

Exemple 4 Considérons le contexte illustré par Tableau 1. Nous avons $h(D) = BD$. En effet, BD est le motif maximal contenant D et tel que son support disjonctif est égal à celui de D , à savoir 4.

3. Représentations concises définies à travers le support disjonctif

Dans cette section, nous visons à présenter brièvement les représentations concises exactes basées sur des motifs caractérisés à travers leurs supports disjonctifs.

3.1. Représentation concise basée sur les essentiels fréquents

Introduite par Casali *et al.* [5], cette représentation concise exacte s'est basée sur la notion de motif essentiel fréquent. Néanmoins, cette représentation utilise aussi l'ensemble des motifs fréquents maximaux, par rapport à l'inclusion ensembliste, afin de connaître le statut de fréquence d'un motif (*i.e.*, fréquent ou pas).

Théorème 5 [5] (Représentation concise basée sur les essentiels fréquents) Soit $\mathcal{BD}^+(\mathcal{MF})$ l'ensemble des motifs fréquents maximaux munis de leurs supports conjonctifs. La représentation $\mathcal{RM}\mathcal{EF}$ formée par l'ensemble $\mathcal{MEF}$ des motifs essentiels fréquents, munis de leurs supports disjonctifs, augmenté par l'ensemble $\mathcal{BD}^+(\mathcal{MF})$ est une représentation concise exacte de l'ensemble des motifs fréquents $\mathcal{MF}$.

Dans la littérature, seul l'algorithme MEP [5] a été proposé afin d'extraire cette représentation concise exacte. Afin d'extraire l'ensemble $\mathcal{BD}^+(\mathcal{MF})$, MEP fait appel à un algorithme externe d'extraction des motifs maximaux fréquents. L'extraction des motifs essentiels fréquents s'effectue en parcourant l'espace de recherche en largeur d'abord afin de ne retenir que les motifs essentiels fréquents, tels qu'ils ont été caractérisés à travers la définition 3.

3.2. Représentation concise basée sur les fermés disjonctifs

La représentation concise basée sur les fermés disjonctifs a été introduite par Hamrouni *et al.* [6] afin de pallier les inconvénients de la représentation $\mathcal{RM}\mathcal{EF}$ (*e.g.*, l'hétérogénéité de la nature du support des éléments retenus – disjonctif vs. conjonctif – et donc de l'espace de recherche associé) tout en assurant une dérivation directe des supports disjonctifs et négatifs des motifs fréquents. Dans cette sous-section, nous allons nous restreindre aux notions essentielles relatives à cette représentation. Pour plus de détails, prière de se référer à [6].

Définition 6 (Ensemble $\mathcal{MFDE}$) L'ensemble $\mathcal{MFDE}$ regroupe les fermetures disjonctives des motifs essentiels fréquents : $\mathcal{MFDE} = \{h(I) \mid I \in \mathcal{MEF}\}$.

Définition 7 (Ensembles $\mathcal{MFDA}$ et $\mathcal{MFDAC}$) Soit $\mathcal{MEII}$ l'ensemble des motifs essentiels inférieurs de tailles impaires, tels que tous leurs sous-ensembles stricts sont des motifs essentiels fréquents. $\mathcal{MFDA}$ est l'ensemble des fermés disjonctifs des éléments de $\mathcal{MEII}$: $\mathcal{MFDA} = \{h(I) \mid I \in \mathcal{MEII}\}$. En outre, l'ensemble $\mathcal{MFDAC}$ est défini comme étant l'ensemble des motifs de $\mathcal{MFDA}$ strictement couverts par au moins un fermé disjonctif de $\mathcal{MFDE}$: $\mathcal{MFDAC} = \{I \in \mathcal{MFDA} \setminus \mathcal{MFDE} \mid \exists I_1 \in \mathcal{MFDE}, I \subset I_1\}$.

Théorème 8 [6] (Représentation concise basée sur les motifs fermés disjonctifs) *L'ensemble des fermés disjonctifs $\mathcal{RMFD} = \mathcal{MFDE} \cup \mathcal{MFDA}$ munis de leurs supports disjonctifs est une représentation concise exacte de l'ensemble des motifs fréquents $\mathcal{MF}$.*

4. DCPR_Miner

Dans cette section, nous introduisons un nouvel algorithme de type *Générer-et-tester*, nommé DCPR_MINER, dédié à l'extraction de la représentation concise basée sur les fermés disjonctifs. Le calcul des fermetures disjonctives des motifs de même taille s'effectue en une seule passe sur le contexte d'extraction grâce au lemme suivant.

Lemme 1 [6] (Caractérisation structurelle de la fermeture disjonctive) *La fermeture disjonctive $h(I)$ d'un motif I est l'ensemble maximal des items qui sont apparus **uniquement** dans les objets contenant au moins un item de I .*

Ainsi, pour un motif I , $h(I)$ peut être calculé à partir d'un contexte d'extraction en deux étapes. Tout d'abord, nous devons dégager l'ensemble des items ayant apparus dans les objets ne contenant aucun item de I . Ensuite, en calculant le complémentaire par rapport à $\mathcal{I}$ de l'ensemble dégagé, le résultat obtenu est égal à $h(I)$.

Par définition, les motifs essentiels fréquents sont les plus petits éléments dans les classes d'équivalence disjonctives engendrées par l'opérateur de fermeture h (cf. Définition 3). Ils sont donc les premiers éléments à donner lieu aux fermetures disjonctives suite à un parcours par niveau (par taille ascendante des candidats). Le repérage des motifs essentiels fréquents peut être efficacement effectué grâce aux fermetures disjonctives. Ceci est réalisé dans DCPR_MINER moyennant le lemme suivant.

Lemme 2 (Caractérisation des motifs essentiels à travers les fermés disjonctifs) *Soit $I \subseteq \mathcal{I}$. Nous avons :*

$$[\text{Iest un motif essentiel}] \Leftrightarrow [\forall I' \subseteq \mathcal{I} \mid (I' \subset I) \wedge (|I'| = |I| - 1), I \not\subseteq h(I')].$$

Autrement dit, I ne peut être un motif essentiel que s'il n'est inclus dans aucune fermeture disjonctive de l'un de ses sous-ensembles directs. Cette nouvelle caractérisation des motifs essentiels, utilisée par DCPR_MINER, permet de connaître si un motif est essentiel ou non sans pour autant calculer son support disjonctif. En effet, il nous suffit d'avoir les fermetures disjonctives de ses sous-ensembles directs pour décider si un motif est essentiel ou pas. Le pseudo-code de DCPR_MINER est donné par Algorithme 1. Les notations utilisées par DCPR_MINER sont décrites dans le tableau 2.

Notation	Description
$\mathcal{C}_i$ (resp. $\mathcal{L}_i$)	: Ensemble des motifs essentiels candidats (resp. fréquents) de taille i .
X_i	: Motif de taille i .
$X_i.h$	: Fermeture disjonctive de X_i .
$X_i.\bar{h}$	: Complémentaire du fermé disjonctif de X_i .
$X_i.\text{Supp_Conj}$ (resp. $X_i.\text{Supp_Disj}$)	: Support conjonctif (resp. disjonctif) de X_i .

Tableau 2. Notations adoptées dans DCPR_MINER.

L'extraction des fermés disjonctifs des ensembles $\mathcal{MFDE}$ et $\mathcal{MFDA}$ munis de leurs supports disjonctifs est effectuée par la procédure CALCUL_SUPPORTS_FERMETURE (cf. ligne 4). Ces ensembles sont le résultat du calcul des fermetures disjonctives des motifs essentiels candidats. La génération des candidats de taille $i + 1$ est effectuée par la fonction APRIORI-GEN [1] appliquée sur les motifs essentiels maintenus de taille i (cf. ligne 5). Cette fonction s'assure de ne maintenir dans l'ensemble $\mathcal{C}_{i+1}$ que les candidats ayant tous leurs sous-ensembles directs des essentiels fréquents. L'élagage des motifs

Algorithme 1 : DCPR_MINER

Données : Un contexte d'extraction $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$, et le seuil minimal du support $minsup$.

Résultat : $\mathcal{RMFD} = \mathcal{MFDE} \cup \mathcal{MFDA}$.

1 Début

2 $\mathcal{MFDE} := \emptyset; \mathcal{MFDA} := \emptyset; i := 1; C_1 := \mathcal{I};$

3 **tant que** $(C_i \neq \emptyset)$ **faire**

4 $\text{CALCUL_SUPPORTS_FERMETURE}(\mathcal{K}, minsup, C_i, \mathcal{L}_i, \mathcal{MFDE}, \mathcal{MFDA});$

5 $C_{i+1} := \text{APRIORI-GEN}(\mathcal{L}_i);$

6 $C_{i+1} := \{X_{i+1} \in C_{i+1} \mid \forall Y_i \subset X_{i+1}, X_{i+1} \not\subseteq h(Y_i)\};$

7 $i := i + 1;$

8 $\mathcal{MFDA} := \{X \in \mathcal{MFDA} \setminus \mathcal{MFDE} \mid \exists Y \in \mathcal{MFDE}, X \subset Y\};$

9 **retourner** $\mathcal{MFDE} \cup \mathcal{MFDA};$

10 **Fin**

11

12 **Procédure** $\text{CALCUL_SUPPORTS_FERMETURE}(\mathcal{K}, minsup, C_i, \mathcal{L}_i, \mathcal{MFDE}, \mathcal{MFDA})$

13 **Début**

14 $\mathcal{L}_i := \emptyset;$

15 **pour chaque** $(O \in \mathcal{O})$ **faire**

16 **pour chaque** $(X_i \in C_i)$ **faire**

17 $\Omega := X_i \cap I / *I$ dénote les items associés à l'objet $O^*/;$

18 **si** $(\Omega = \emptyset)$ **alors**

19 $X_i.h := X_i.\bar{h} \cup I;$

20 **sinon**

21 $X_i.Supp_Disj := X_i.Supp_Disj + 1;$

22 **si** $(\Omega = X_{i+1})$ **alors**

23 $X_i.Supp_Conj := X_i.Supp_Conj + 1;$

24 **pour chaque** $(X_i \in C_i)$ **faire**

25 **si** $(X_i.Supp_Conj \geq minsup)$ **alors**

26 $\mathcal{L}_i := \mathcal{L}_i \cup \{X_i\};$

27 $X_i.h := \mathcal{I} \setminus X_i.\bar{h};$

28 $\mathcal{MFDE} := \mathcal{MFDE} \cup \{(X_i.h, X_i.Supp_Disj)\};$

29 **sinon**

30 **si** $(i \text{ est impair})$ **alors**

31 $X_i.h := \mathcal{I} \setminus X_i.\bar{h};$

32 $\mathcal{MFDA} := \mathcal{MFDA} \cup \{(X_i.h, X_i.Supp_Disj)\};$

33 **Fin**

non-essentiels de l'ensemble C_{i+1} est effectué grâce à l'instruction 6 basée sur le lemme 2. Enfin, la construction de l'ensemble $\mathcal{MFDA}$ à partir de $\mathcal{MFDA}$ est effectuée grâce à l'instruction de la ligne 8.

Dans ce qui suit, nous introduisons les propositions prouvant la correction, la complétude et la terminaison de DCPR_MINER, ainsi que sa complexité au pire des cas.

Proposition 9 (Propriétés de correction et de complétude) *L'algorithme DCPR_MINER est correct et complet. Il permet d'extraire avec exactitude tous les éléments de l'ensemble $\mathcal{RMFD}$ munis de leurs supports disjonctifs exacts.*

Proposition 10 (Propriété de terminaison) *L'algorithme DCPR_MINER termine.*

Proposition 11 (Complexité des phases de DCPR_MINER) *Soient $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$ un contexte d'extraction, $m = |\mathcal{O}|$ et $n = |\mathcal{I}|$. Alors, la complexité au pire des cas de l'extraction des ensembles $\mathcal{MFDE}$ et $\mathcal{MFDA}$ est de l'ordre de $O((n^2 + m \times n) \times 2^n)$. Tandis que la complexité au pire des cas de la dérivation de l'ensemble $\mathcal{MFDA}$ à partir de $\mathcal{MFDA}$ est de l'ordre de $O(n \times \binom{n}{\lceil \frac{n}{2} \rceil} \times 2^n)$.*

Les preuves de ces propositions sont omises faute d'espace.

5. Étude expérimentale

Dans cette section, nous présentons les résultats de l'étude expérimentale que nous avons menée, permettant de comparer les performances des algorithmes d'extraction des représentations concises définies à travers le support disjonctif, à savoir DCPR_MINER et MEP. Dans cette étude, nous nous focalisons sur les cas les plus intéressants, à savoir les contextes d'extraction où les items sont fortement corrélés. En effet, dans de tels contextes, les cardinalités des représentations concises exactes sont largement inférieures à celle des motifs fréquents [7].

Les expérimentations ont été effectuées sur un PC muni d'un processeur Intel doté d'une horloge de fréquence 1,73 GHz avec 2 Go de mémoire centrale. Le système d'exploitation utilisé est Fedora Core 7, avec 2 Go d'espace d'échange. Les algorithmes DCPR_MINER et MEP⁽¹⁾ sont implémentés en C++, et le compilateur utilisé pour la génération des exécutables qui leurs sont relatifs est gcc 4.1.2. Les caractéristiques des contextes "benchmark" sur lesquelles les tests ont été effectués sont étalés dans le tableau 3⁽²⁾. Les performances de DCPR_MINER et de MEP sont illustrées par la figure 1.

Nom du contexte	Nombre d'objets	Nombre d'items	Taille moyenne des objets
CONNECT	67,557	129	43
PUMSB	49,046	2,113	74
MUSHROOM	8,124	119	23
CHESS	3,196	75	37

Tableau 3. Caractéristiques des contextes.

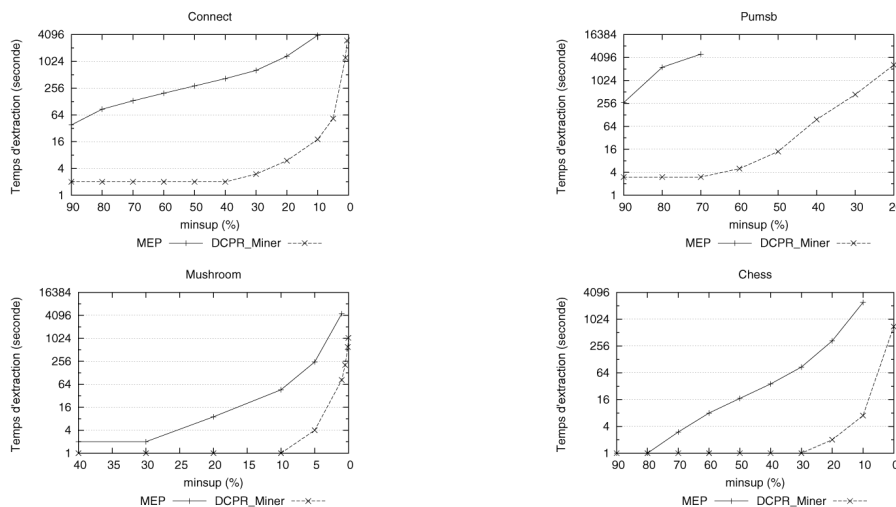


Figure 1. Performances de DCPR_MINER vs. celles de MEP.

À la vue des résultats collectés, nous pouvons constater les assertions suivantes :

- 1. Efficacité de l'extraction de la représentation concise $\mathcal{RMFD}$:** En effet, à l'exception des seuils *minsup* très faibles (e.g., *minsup* = 20% pour PUMSB), DCPR_MINER permet l'extraction de la représentation $\mathcal{RMFD}$ dans un temps très réduit. Pour des seuils très faibles de *minsup*, il garantit également un temps d'extraction acceptable.
- 2. Coût considérable d'extraction de la représentation concise $\mathcal{RMF}$:** Même pour

1. Les auteurs souhaitent remercier A. Casali et S. Lopes pour avoir fourni le code source de MEP.

2. Ces contextes sont disponibles à l'adresse suivante : <http://fimi.cs.helsinki.fi/data>.

des seuils de *minsup* élevés, les temps d'extraction consommés par MEP sont élevés. De tels résultats ont été causés par deux facteurs à savoir l'utilisation d'un algorithme externe afin d'extraire l'ensemble $\mathcal{BD}^+(\mathcal{MF})$, et l'extraction des motifs essentiels fréquents à travers leurs définition classique. En effet, une telle définition oblige MEP à calculer les supports disjonctifs des motifs candidats afin de déterminer s'ils sont essentiels ou non.

3. DCPR_MINER est plus efficace que MEP : Pour l'ensemble des contextes considérés, les temps consommés par DCPR_MINER sont nettement inférieurs à ceux de MEP. En effet, DCPR_MINER profite des fermetures disjonctives pour éviter le calcul des supports disjonctifs des candidats non-essentiels en les élaguant grâce au lemme 2. Le nombre de ces derniers augmente considérablement en faisant décroître la valeur de *minsup*.

6. Conclusion

Dans ce papier, nous avons fait un survol de l'état de l'art relatif aux représentations concises des motifs fréquents définies à travers le support disjonctif. Par la suite, nous avons introduit un nouvel algorithme, nommé DCPR_MINER, permettant d'extraire celle basée sur les fermés disjonctifs. Les expérimentations menées sur des contextes "benchmark" montrent que : (1) L'algorithme introduit permet l'extraction efficace de la représentation désirée ; (2) DCPR_MINER est plus efficace que l'algorithme MEP d'extraction de la représentation basée sur les essentiels fréquents. Les perspectives futures de recherche concernent la conception, l'implémentation et l'évaluation d'algorithmes efficaces de régénération de l'ensemble des motifs fréquents à partir de ces représentations.

7. Bibliographie

- [1] R. AGRAWAL, R. SRIKANT, « Fast algorithms for mining association rules », *Proceedings of the 20th International Conference on Very Large Databases (VLDB 1994)*, Santiago, Chile, pages 478–499, 1994.
- [2] N. PASQUIER, Y. BASTIDE, R. TAOUIL, G. STUMME, L. LAKHAL, « Generating a condensed representation for association rules », *Journal of Intelligent Information Systems*, vol. 24, n° 1, pages 25–60, 2005.
- [3] T. CALDERS, B. GOETHALS, « Non-derivable itemset mining », *Journal of Data Mining and Knowledge Discovery*, vol. 14, n° 1, pages 171–206, 2007.
- [4] J. MUHONEN, H. TOIVONEN, « Closed non-derivable itemsets », *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006)*, Springer-Verlag, LNAI, vol. 4213, Berlin, Germany, pages 601–608, 2006.
- [5] A. CASALI, R. CICHETTI, L. LAKHAL, « Essential patterns : A perfect cover of frequent patterns », *Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2005)*, Springer-Verlag, LNCS, vol. 3589, Copenhagen, Denmark, pages 428–437, 2005.
- [6] T. HAMROUNI, I. DENDEN, S. BEN YAHIA, E. MEPHU NGUIFO, « A new concise representation of frequent patterns through disjunctive search space », *Proceedings of the 5th International Conference on Concept Lattices and their Applications (CLA 2007)*, Montpellier, France, pages 50–61, 2007.
- [7] I. DENDEN, T. HAMROUNI, S. BEN YAHIA, E. MEPHU NGUIFO, « Étude expérimentale des représentations concises des itemsets fréquents », Rapport technique, CRIL-CNRS de Lens, France, 2007.
- [8] J. GALAMBOS, I. SIMONELLI, « Bonferroni-type inequalities with applications », Springer, 2000.
- [9] B. GANTER, R. WILLE, « Formal Concept Analysis », Springer, 1999.