Approches déclaratives pour la fouille de données orientée motifs

Séminaire HEUDIASYC

Samir Loudni

15 Mai 2018

Principales thématiques de recherche

Programmation par contraintes et recherche locale

- recherches locales à grand voisinages (VNS)
- contraintes globales relaxées
- décomposition de graphes et contraintes

Contraintes pour la fouille

- fouille de données ensemblistes et séquentielles sous contraintes
- introduction de la souplesse en fouille de données sous contraintes

Méthodes d'optimisation pour la fouille d'ensemble de motifs

- clustering conceptuel sous contraintes en PLNE
- PLNE pour le problème de tuilage
- recherche de solutions équitables en optimisation multi-agents

Plan de la présentation

1. Introduction
2. Contraintes globales pour la fouille de motifs séquentiels
3. La PPC pour l'extraction de Skypatterns (souples)

Programmation par Contraintes (PPC)

"Un cadre générique pour la résolution de problèmes combinatoires"

- Description déclarative du problème par un triplet (X,D,C) où
  - $X = \{x_1, \ldots, x_n\}$ est l'ensemble fini de variables
  - $D = \{D_1, \ldots, D_n\}$ est l'ensemble de domaines finis des variables
  - $C = \{c_1, \ldots, c_e\}$ est l'ensemble de contraintes restreignant les valeurs des variables x_i

- **Résolution** = Énumération + Filtrage
  - méthodes de résolution génériques
  - algorithmes de filtrage des contraintes individuelles

solution $\equiv$ une affectation complète des variables de X satisfaisant toutes les contraintes de C

Processus de résolution

Filtrage

- réduction de domaines : suppression des valeurs des domaines de variables ne pouvant mener à aucune solution
- propagation : transmission entre les contraintes des évènements de suppression de valeurs

Recherche

- phase d'énumération : stratégies pour affecter des valeurs aux variables + Backtrack

Processus de résolution

Filtrage

- réduction de domaines : suppression des valeurs des domaines de variables ne pouvant mener à aucune solution
  - exemple : $x_1 > x_2$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$
    ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
- propagation : transmission entre les contraintes des évènements de suppression de valeurs

Recherche

- phase d'énumération : stratégies pour affecter des valeurs aux variables + Backtrack

Processus de résolution

Filtrage

- réduction de domaines : suppression des valeurs des domaines de variables ne pouvant mener à aucune solution
  - exemple : $x_1 > x_2$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$
    ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
- propagation : transmission entre les contraintes des évènements de suppression de valeurs
  - exemple : $x_1 > x_2$, $x_1 = x_3$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$, $D_3 = \{1, 3\}$
    $(x_1 > x_2)$ ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$

Recherche

- phase d'énumération : stratégies pour affecter des valeurs aux variables + Backtrack

Processus de résolution

Filtrage

- réduction de domaines : suppression des valeurs des domaines de variables ne pouvant mener à aucune solution
  - exemple : $x_1 > x_2$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$
    ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
- propagation : transmission entre les contraintes des évènements de suppression de valeurs
  - exemple : $x_1 > x_2$, $x_1 = x_3$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$, $D_3 = \{1, 3\}$
    $(x_1 > x_2)$ ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
    $(x_1 = x_3)$ ➠ $D_1 = \{1, \cancel{2}\}$, $D_3 = \{1, \cancel{3}\}$

Recherche

- phase d'énumération : stratégies pour affecter des valeurs aux variables + Backtrack

Processus de résolution

Filtrage

- réduction de domaines : suppression des valeurs des domaines de variables ne pouvant mener à aucune solution
  - exemple : $x_1 > x_2$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$
    ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
- propagation : transmission entre les contraintes des évènements de suppression de valeurs
  - exemple : $x_1 > x_2$, $x_1 = x_3$ et $D_1 = \{1, 2\}$, $D_2 = \{0, 1, 2, 3\}$, $D_3 = \{1, 3\}$
    $(x_1 > x_2)$ ➠ $D_1 = \{1, 2\}$, $D_2 = \{0, 1, \cancel{2}, \cancel{3}\}$
    $(x_1 = x_3)$ ➠ $D_1 = \{1, \cancel{2}\}$, $D_3 = \{1, \cancel{3}\}$
    $(x_1 = 1) \wedge (x_1 > x_2)$ ➠ $D_2 = \{0, \cancel{1}\}$

Recherche

- phase d'énumération : stratégies pour affecter des valeurs aux variables + Backtrack

Contraintes globales

Deux apports :

1. expressivité : utile pour exprimer des relations complexes impliquant un nombre quelconque de variables
   - `AllDifferent`$(\{x_1, \ldots, x_n\}) \Leftrightarrow \bigwedge_{1 \leq i < j \leq n} (x_i \neq x_j)$
     ➠ `AllDifferent` : deux cours ne doivent pas avoir lieu en même temps

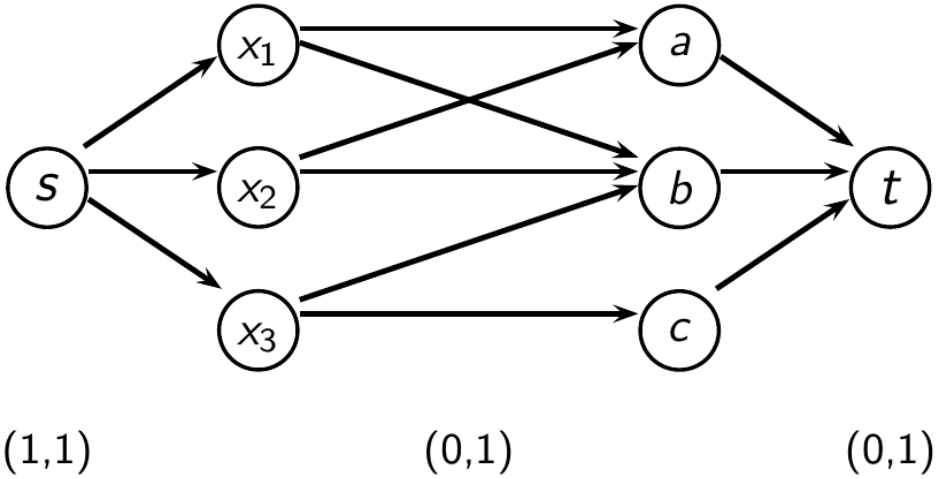
2. meilleur filtrage :
   - Un niveau cohérence au moins aussi élevé que celui que l'on pourrait maintenir sur un réseau des contraintes élémentaires.
   - filtrage mis en œuvre à l'aide d'autres outils
     - algorithmes de graphes/d'automates,
     - réseau de flot,
     - ...

➠ Les contraintes globales sont une spécificité de la PPC

Exemple : `AllDifferent`

$x_1 \in \{a, b\}$, $x_2 \in \{a, b\}$, $x_3 \in \{b, c\}$

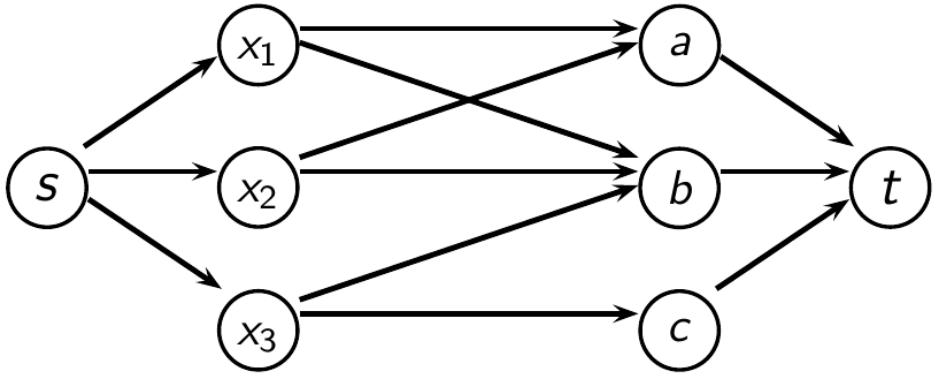
`AllDifferent`(x_1, x_2, x_3)



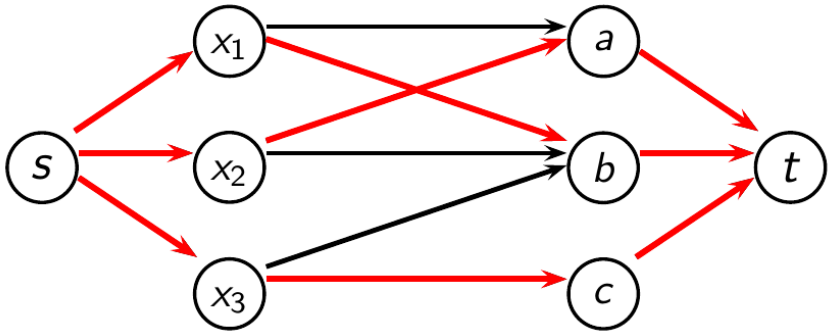
Exemple : `AllDifferent`

$x_1 \in \{a, b\}$, $x_2 \in \{a, b\}$, $x_3 \in \{b, c\}$

`AllDifferent`(x_1, x_2, x_3)



(1,1) (0,1) (0,1)



(1,1) (0,1) (0,1)

Cohérence globale

Définition (Cohérence globale)

Une contrainte globale $\mathrm{GC}(X)$ portant sur les variables $X = \{x_1, \ldots, x_n\}$ de domaines respectifs $D_1, \ldots, D_n$ est dite globalement cohérente si pour chaque variable x_i $(i = 1, ...n)$ et chaque valeur $d \in D_i$ il existe un tuple $(d_1, \ldots, d_{i-1}, d, d_{i+1}, \ldots, d_n)$ vérifiant la contrainte GC où $d_j \in D_j$.

➠ tuple autorisé = solution d'une GC

Contraintes globales pour la fouille de motifs séquentiels

Contexte : Fouille de données

Fouille de données

- but : découvrir des informations nouvelles dans les BD [Agrawal 93]
- processus : recherche de régularités (ou corrélations) :
  - motifs
  - règles d'association
  - clustering

	g_1	g_2	g_3	g_4
s_1	x			x
s_2	x	x	x	
s_3		x		x
s_4	x	x	x	
s_5	x	x		x

motif fréquent : $g_1 g_2$

règle d'assoc. : $g_1 g_2 \rightarrow g_3$

Fouille de motifs séquentiels

- prise en compte de la dimension séquentielle

Extraction de motifs séquentiels : préliminaires

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

- item : littéral, ex : $A, B, C, D, \ldots$
- séquence : liste ordonnée d'items, ex : $\langle ABCBC \rangle$
- base de séquences : ensemble de tuples (sid, s)
- motif : sous-séquence d'items, ex : $\langle AC \rangle$ est sous-séquence de $\langle ABCBC \rangle$
- support d'un motif
- extraction de motifs séquentiels

Extraction de motifs séquentiels (EMS)

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

$\texttt{cover}(\langle AC \rangle) = \{(1, s_1), (2, s_2)\}$ ➠ $\texttt{sup}(\langle AC \rangle) = 2$

- séquence
- motif
- support d'un motif : nombres de séquences dans lequel apparaît le motif
- extraction de motifs séquentiels

Extraction de motifs séquentiels (EMS)

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

avec $\theta = 2$ ($\sup(p) \geq 2$)

$FS = \{\langle A \rangle, \langle B \rangle, \langle C \rangle, \langle AB \rangle, \langle AC \rangle, \langle BB \rangle, \langle BC \rangle, \langle ABC \rangle, \langle \mathbf{BBC} \rangle\}$ ➠ 9 motifs !

- motif
- support d'un motif
- extraction de motifs fréquents : extraction de TOUS les motifs $>$ à un seuil (θ)

Extraction de motifs séquentiels (EMS)

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

avec $\theta = 2$ ($\sup(p) \geq 2$)

$FS = \{\langle A \rangle, \langle B \rangle, \langle C \rangle, \langle AB \rangle, \langle AC \rangle, \langle BB \rangle, \langle BC \rangle, \langle ABC \rangle, \langle \mathbf{BBC} \rangle\}$ ➠ 9 motifs !

- motif
- support d'un motif
- extraction de motifs fréquents : extraction de TOUS les motifs $>$ à un seuil (θ)

L'anti-monotonie de la contrainte de support minimal

Propriété d'anti-monotonie (Agrawal & Sirkant, 94)

Une contrainte c est anti-monotone, si et seulement si, pour tout motif satisfaisant c, tous ses sous-motifs satisfont également c.

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

Avec un seuil de support minimal $\theta = 2$

$\langle BA \rangle$ n'est pas fréquent $\rightarrow$ aucun sur-motif de $\langle BA \rangle$ ne pourra être fréquent

EMS sous contraintes : contrainte de longueur minimale

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

Définition

$minSize(p, \ell_{min})$ impose que les motifs extraits p doivent contenir au moins ℓ_{min} items.

En imposant $\sup(p) \geq 2 \wedge minSize(p, 3)$

seuls 2 motifs sont extraits : $\langle ABC \rangle$ and $\langle BBC \rangle$

EMS sous contraintes : contrainte d'item

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

Définition

$reg(p, exp)$ assure que le motif p doit être reconnu par un automate d'états finis associé à l'expression régulière exp.

EMS sous contraintes : contrainte d'expression régulière

sid	Séquence
1	$\langle ABCBC \rangle$
2	$\langle BABC \rangle$
3	$\langle AB \rangle$
4	$\langle BCD \rangle$

En imposant $\mathrm{sup}(p) \geq 2 \wedge reg(p, exp)$ où $exp = B\{BC|D\}$

le motif $\langle BBC \rangle$ est extrait de la base

Fouille de motifs sous contraintes

Pourquoi contraindre les motifs ?

- afin de **réduire** le nombre de motifs extraits
- cibler les motifs potentiellement intéressants
- intégrer les connaissances du domaine dans la fouille

Approches existantes :

Contraintes	GSP	SPADE	cSPADE	SPAM	SPIRIT	BIDE	GAP-BIDE	PrefixSpan	TSP	TKS
Fréquence minimale	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Item										
Taille			✓						✓	✓
Gap			✓				✓			
Expressions régulières					✓					
Fermeture						✓	✓			
top-k									✓	✓

➠ Des méthodes dédiées selon les contraintes

La PPC pour la fouille séquentielle

Utiliser la PPC et ses solveurs génériques !

- contraintes complexes
- grande généricité
- ajout/suppression de contraintes (utilisateur)
- exhaustivité, optimalité

➠ Passage à l'échelle : enjeu majeur pour les approches PPC existantes

Principales limites de l'existant

Encodage de la base de séquences avec des contraintes réifiées
m variables booléennes S_s sont utilisées t.q.

$$(S_s = 1) \Leftrightarrow (\text{p est sous-séquence de s})$$

- Pro :
  - Encodage de la contrainte de fréquence immédiat
    ➠ $sup_{SDB}(p) = \Sigma_{s \in SDB} S_s$
- Cons :
  - nécessite ($m = |SDB|$) contraintes réifiées pour encoder toute la base
    ➠ Prohibitif même pour les bases de taille moyenne

Une contrainte globale pour l'extraction de motifs séquentiels [Kemmar et al. (CP'15)]

Une nouvelle contrainte globale PREFIX-PROJECTION pour l'extraction de motifs séquentiels

➠ Peut être combinée avec d'autres contraintes

Contraintes	PP	SPADE	cSPADE	SPAM	SPIRIT	BIDE	GAP-BIDE	PrefixSpan
Fréquence minimale	✓	✓	✓	✓	✓	✓	✓	✓
Item	✓							
Taille	✓		✓					
Gap			✓				✓	
Expressions régulières	✓				✓			
Fermeture						✓	✓	
top-k	✓							

➠ Ne requiert ni contraintes ni variables supplémentaires pour encoder la relation de sous-séquence

Prefix-Projection($[x_1, \ldots, x_\ell], \theta$) (1/2)

- $x = <x_1, \ldots, x_\ell>$ doit être un motif séquentiel t.q. $\sup(x) \geq \theta$
- Chaque variable x_i a pour domaine $D_i = \mathcal{I} \cup \{\square\}$
  - $\mathcal{I}$: ensemble de n items
  - $\square$: le symbole vide ($\square \notin \mathcal{I}$) indiquant la fin de la séquence
- Deux règles sur les domaines :
  - le premier item de x doit être non vide ($\square \notin D_1$)
  - $\forall i \in [2..(\ell - 1)], (x_i = \square) \rightarrow (x_{i+1} = \square)$

Prefix-Projection($[x_1, \ldots, x_n], \theta$) (2/2)

Exemple :

- ensemble d'items $\mathcal{I} = \{A, B, C, D\}$, $\theta = 2$ et $x = < x_1, x_2, x_3 >$

sid	Séquence
1	$ABCDBC$
2	$BABC$
3	AB
4	BCD

- soit $x_1 = A$
- projection préfixée selon A :
  $SDB|_A = \{(1, \langle BCDBC \rangle), (2, \langle BC \rangle), (3, \langle B \rangle)\}$
- seuls les suffixes s débutant par B ou C pourront satisfaire $\mathtt{sup}(s) \geq 2$
- donc $x_2 \neq A$, $x_2 \neq D$ et idem pour x_3

  ➠ $D(x_2) = D(x_3) = \{B, C, \square\}$

Résultats expérimentaux

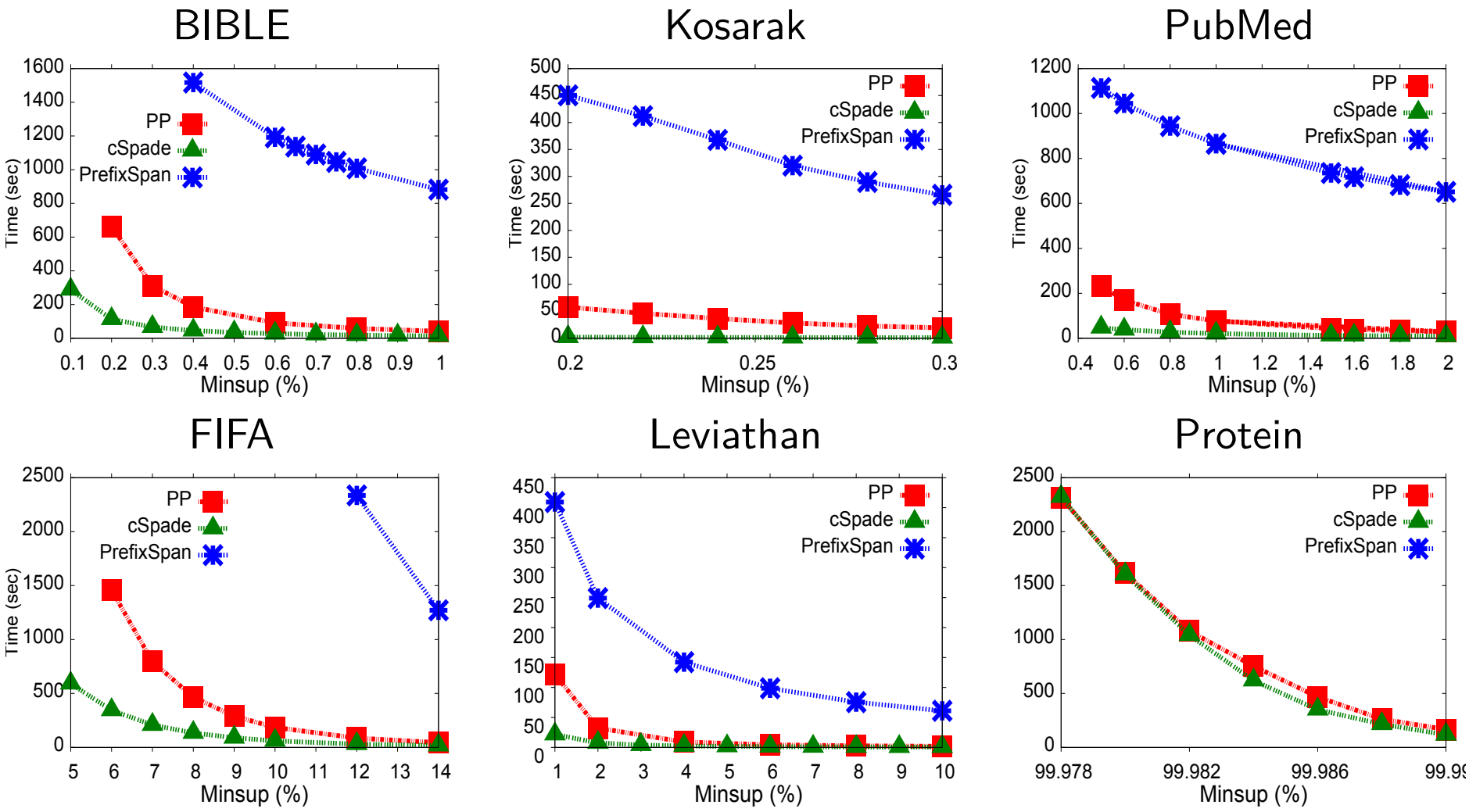
- Datasets :

dataset	$\|SDB\|$	$\|\mathcal{I}\|$	avg $(\|s\|)$	$\max_{s \in SDB}(\|s\|)$	type de données
Leviathen	5834	9025	33.81	100	book
FIFA	20450	2990	34.74	100	web click stream
BIBLE	36369	13905	21.64	100	bible
Kosarak	69999	21144	7.97	796	web click stream
Protein	103120	24	482	600	protein sequences
PubMed	17527	19931	29	198	bio-medical text
data-200K	200000	20	50	86	synthetic dataset

- Implémentation : `Gecode`, timeout = `3600 secondes`
  ➥ `https://sites.google.com/site/prefixprojection4cp/`

Résultats expérimentaux

Comparaison avec les méthodes ad hoc



Bilan

Point de vue PPC

- plusieurs contraintes globales pour la fouille de séquences et d'itemsets
- filtrage exploitant des propriétés bien connues en fouille

Contraintes	PP	SPADE	cSPADE	SPAM	SPIRIT	BIDE	GAP-BIDE	PrefixSpan
Fréquence minimale	✓	✓	✓	✓	✓	✓	✓	✓
Item	✓							
Taille	✓		✓					
Gap	✓		✓				✓	
Expressions régulières	✓				✓			
Fermeture	✗					✓	✓	
top-k	✓							

Point de vue modélisation et résolution

- cadre unifié pour "composer" plusieurs contraintes
- premières approches PPC qui passent à l'échelle

La PPC pour l'extraction de Skypatterns (souples)

Skypatterns : motivations

- Offrir à l'utilisateur une nouvelle façon d'exprimer ses préférences :

  ➥ plusieurs types de mesures peuvent être combinées

- Éviter le problème de choix de seuils (bien connu en FD) :
  - quelle est la valeur souhaitable pour le seuil de fréquence minimale ?
  - combiner plusieurs mesures : comment fixer simultanément plusieurs seuils ?

- Découverte des **motifs** satisfaisant une forme d'intérêt global

  ➥ relation de Pareto dominance

Skypatterns : un exemple

Tid	Items					
t_1	A				E	F
t_2		B	C	D	E	
t_3		B	C	D	E	F
t_4	A	B	C	D	E	
t_5		B	C	D		
t_6		B			E	F
t_7	A	B	C	D	E	F

$\{\texttt{freq}, \texttt{aire}\}$

Skypatterns : un exemple

Tid	Items					
t_1	A				E	F
t_2		B	C	D	E	
t_3		B	C	D	E	F
t_4	A	B	C	D	E	
t_5		B	C	D		
t_6		B			E	F
t_7	A	B	C	D	E	F

Motif	`freq`	`aire`
AB	2	4
$BCDE$	4	16
BCD	5	15
DEF	2	6
B	6	6
AEF	2	6
CF	2	4
E	6	6

$\vdots$

$\{\texttt{freq}, \texttt{aire}\}$

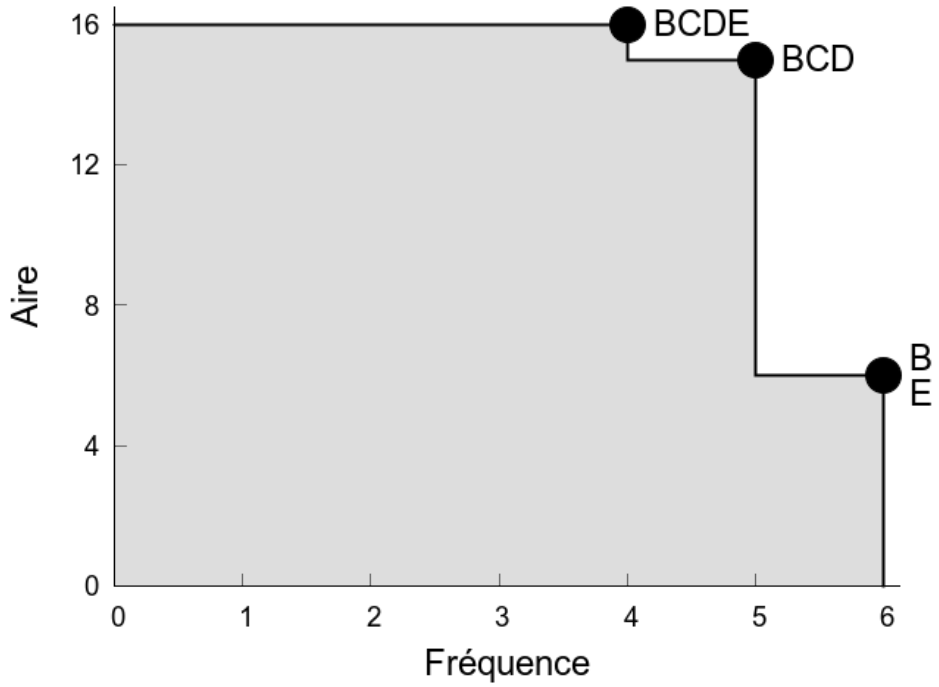
Skypatterns : un exemple

Tid	Items					
t_1	A				E	F
t_2		B	C	D	E	
t_3		B	C	D	E	F
t_4	A	B	C	D	E	
t_5		B	C	D		
t_6		B			E	F
t_7	A	B	C	D	E	F

Motif	`freq`	`aire`
~~AB~~	2	4
$BCDE$	4	16
BCD	5	15
~~DEF~~	2	6
B	6	6
~~AEF~~	2	6
~~CF~~	2	4
E	6	6

⋮

$Skypattern(\{\texttt{freq}, \texttt{aire}\}) = \{BCDE, BCD, B, E\}$



$|\mathcal{L}_{\mathcal{I}}| = 2^6$, mais juste 4 skypatterns

Skypatterns : plus formellement

➠ notion de skylines dans l'extraction de motifs [Soulet et al. (ICDM'11)]

M : ensemble de mesures $\mathcal{I}$: ensemble d'items $\mathcal{L}_{\mathcal{I}} = 2^{\mathcal{I}}$: ensemble de motifs

(Pareto-)dominance entre motifs : Un motif p_i domine un autre motif p_j par rapport M noté $p_i \succ_M p_j$ ssi $\forall m \in M, m(p_i) \geq m(p_j) \wedge \exists m' \in M, m'(p_i) > m'(p_j)$

➥ un skypattern par rapport à M est un *motif non-dominé par rapport à M*

L'**opérateur skypattern** $Sky(\mathcal{L}_{\mathcal{I}}, M)$ calcule tous les skypatterns par rapport à M :

$$Sky(\mathcal{L}_{\mathcal{I}}, M) = \{p \in \mathcal{L}_{\mathcal{I}} \mid \nexists\, y \in \mathcal{L}_{\mathcal{I}} : y \succ_M p\}$$

Extraction des skypatterns

Notre proposition [Ugarte et al. (CPAIOR'14, AIJ'17)]

Énumérer naïvement tous les motifs candidats ($\mathcal{L}_{\mathcal{I}}$) et les comparer deux à deux n'est pas faisable

Deux principes clés :

- tirer profit de la représentation condensée des mesures condensables de M [Soulet et al. (DMKD'08)]
  - identification d'un petit ensemble de mesures M' ($M' \subseteq M$)
    ➠ principe de skylineabilité
- utilisation de CSP dynamiques pour réduire la zone de dominance en faisant des comparaisons binaires (deux à deux) entre les motifs

Extraction des skypatterns en utilisant la PPC (1/2)

1. **Principe :**
   - Extraire un motif initial s_1 fermé par rapport à M'
   - Chercher un motif s_2 qui n'est pas dominé par s_1
   - Chercher un motif s_3 qui n'est pas dominé par s_1 ni par s_2

     $\vdots$
   - Jusqu'à ce qu'il ne soit plus possible d'étendre la zone de dominance

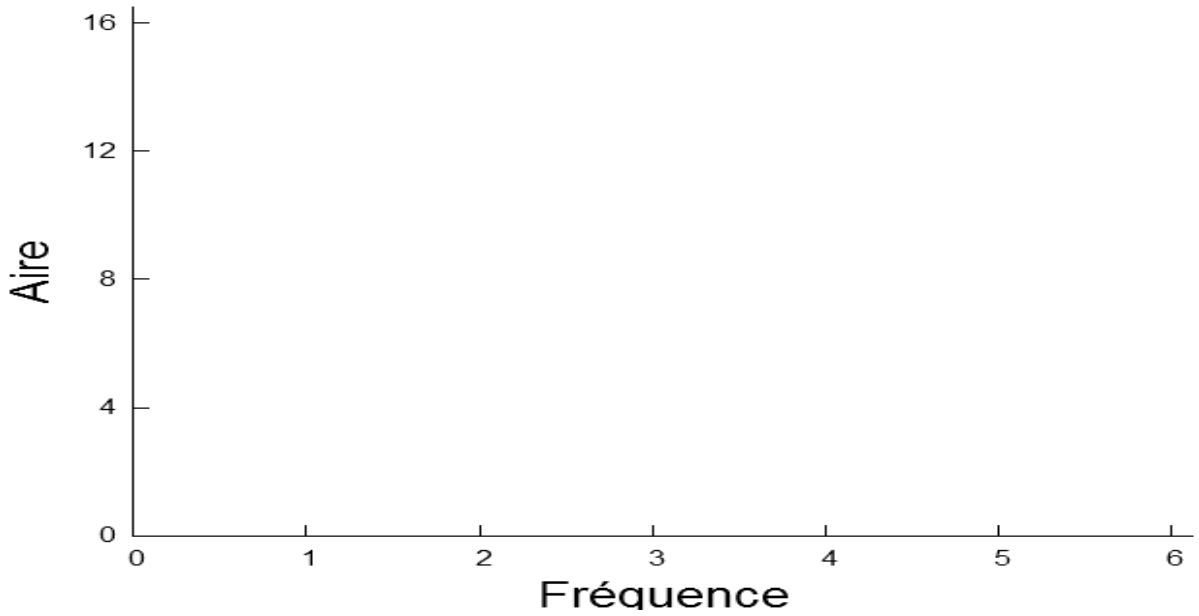
2. **Résolution :**
   - Les contraintes sont postées dynamiquement pendant l'étape d'extraction (pour chaque candidat s_i, ajouter la contrainte $s_i \not\succ_M p$)

➥ La zone de dominance est réduite de façon incrémentale grâce au filtrage

$$(s_i \not\succ_M p) \equiv \bigvee_{m \in M} m(s_i) < m(p) \vee \bigwedge_{m \in M} m(s_i) = m(p)$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



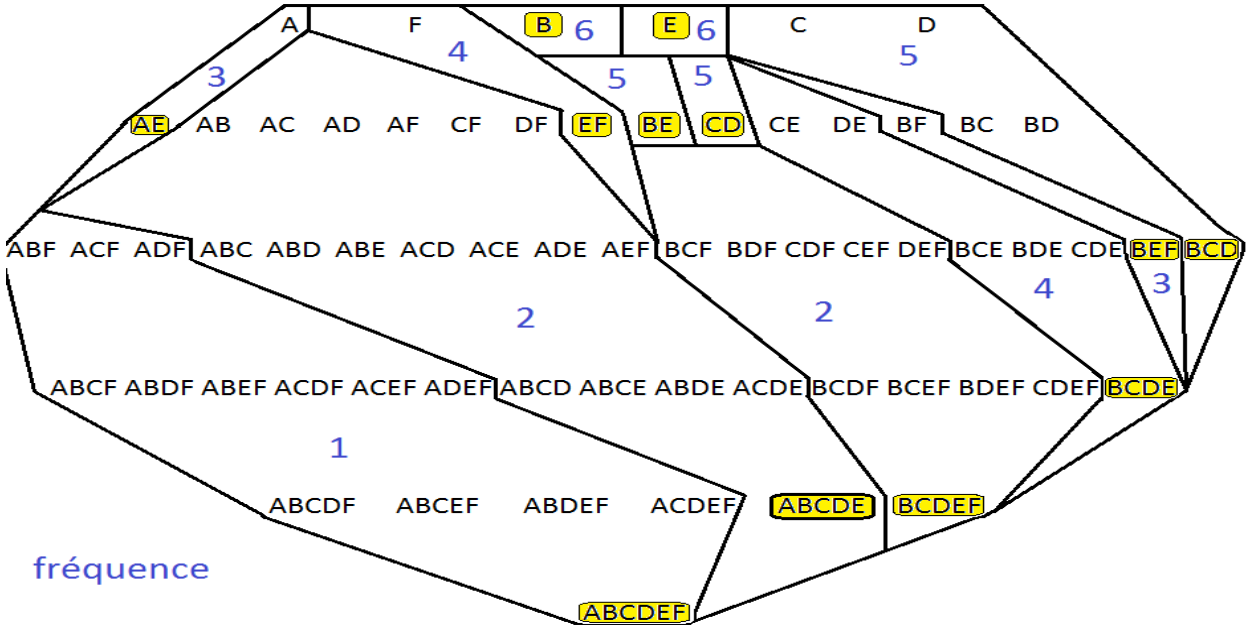
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv$$

$$Candidats =$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



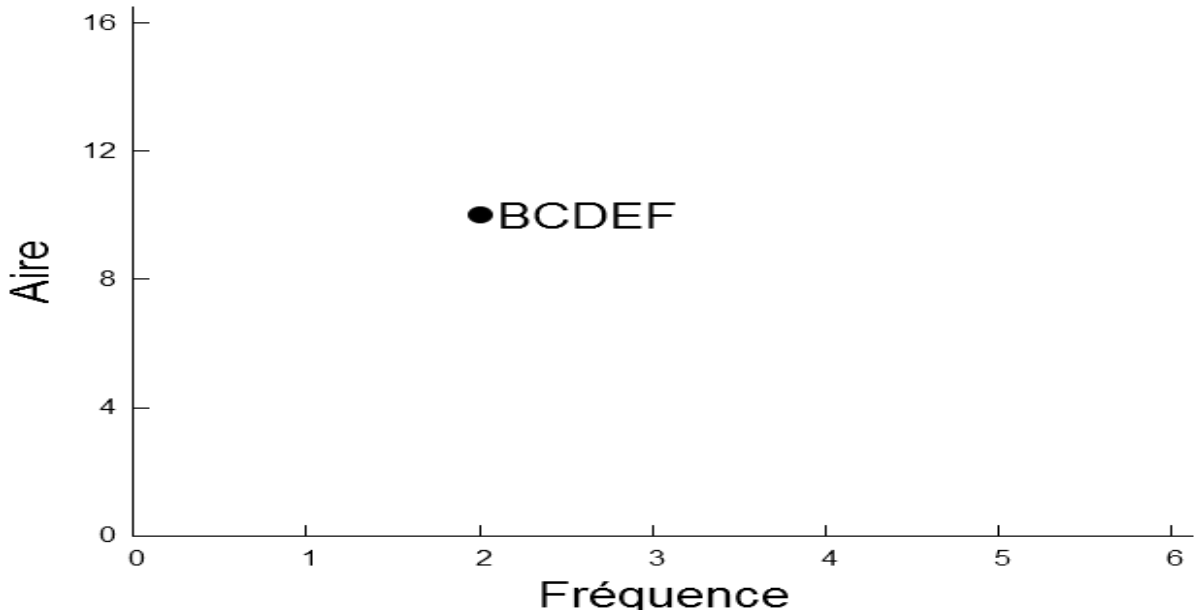
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p)$$

$$Candidats =$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



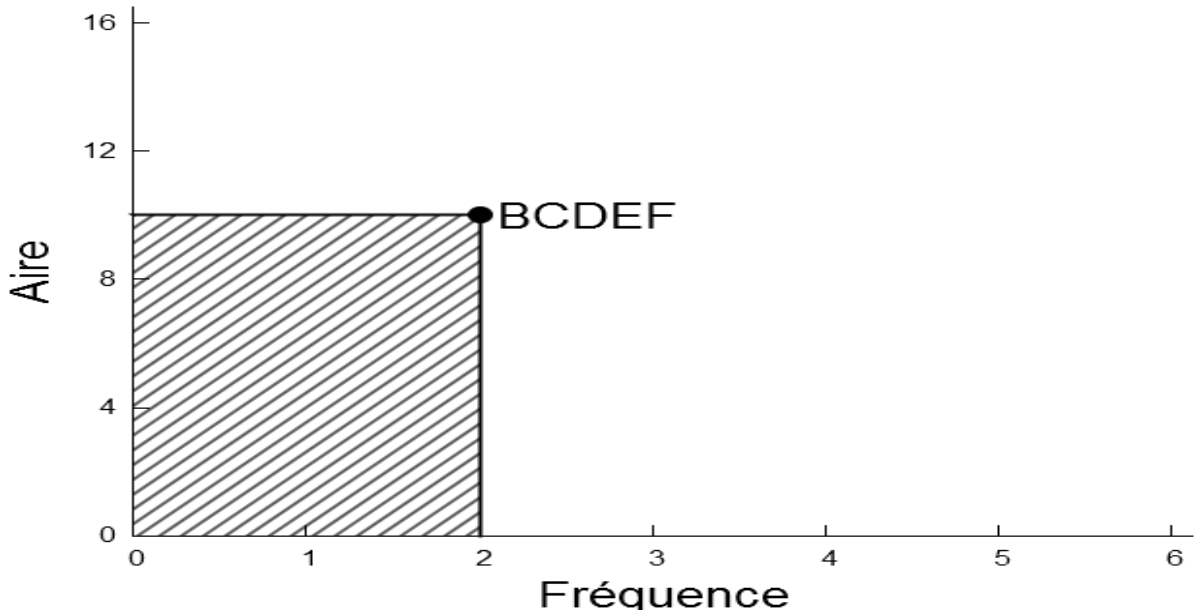
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



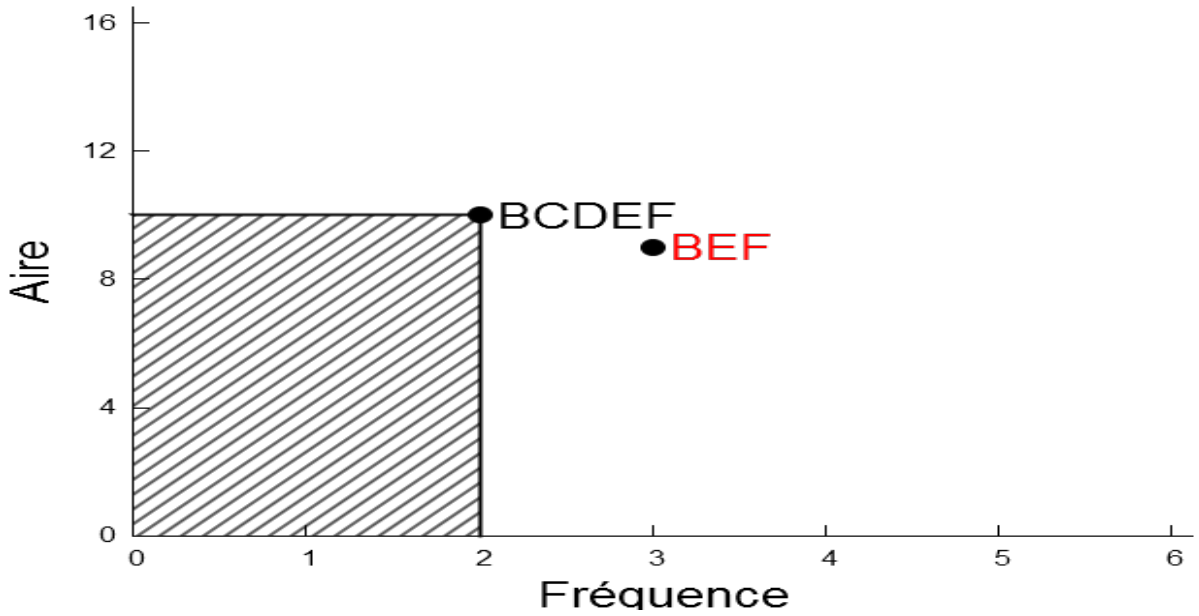
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



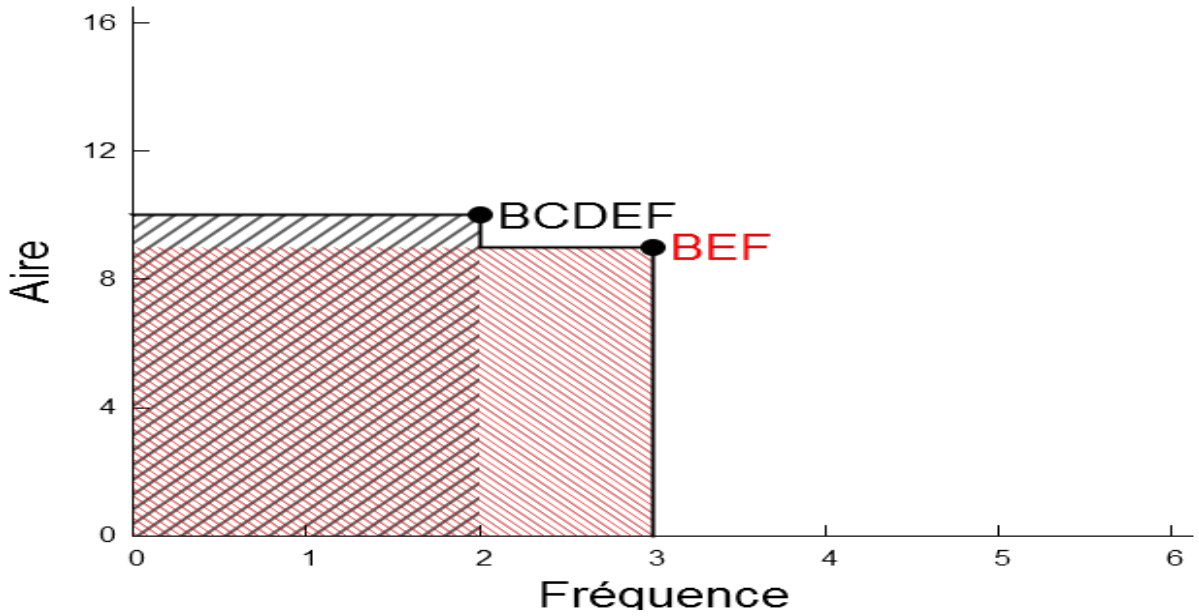
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1}, \underbrace{\text{BEF}}_{s_2},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



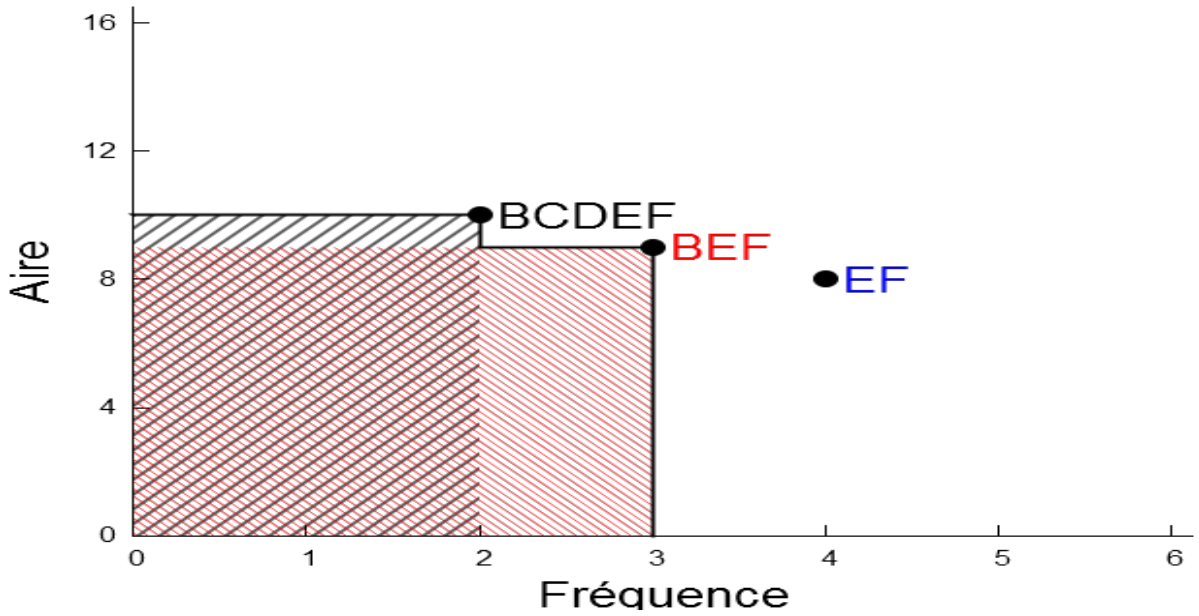
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p) \wedge (s_2 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1}, \underbrace{\text{BEF}}_{s_2},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



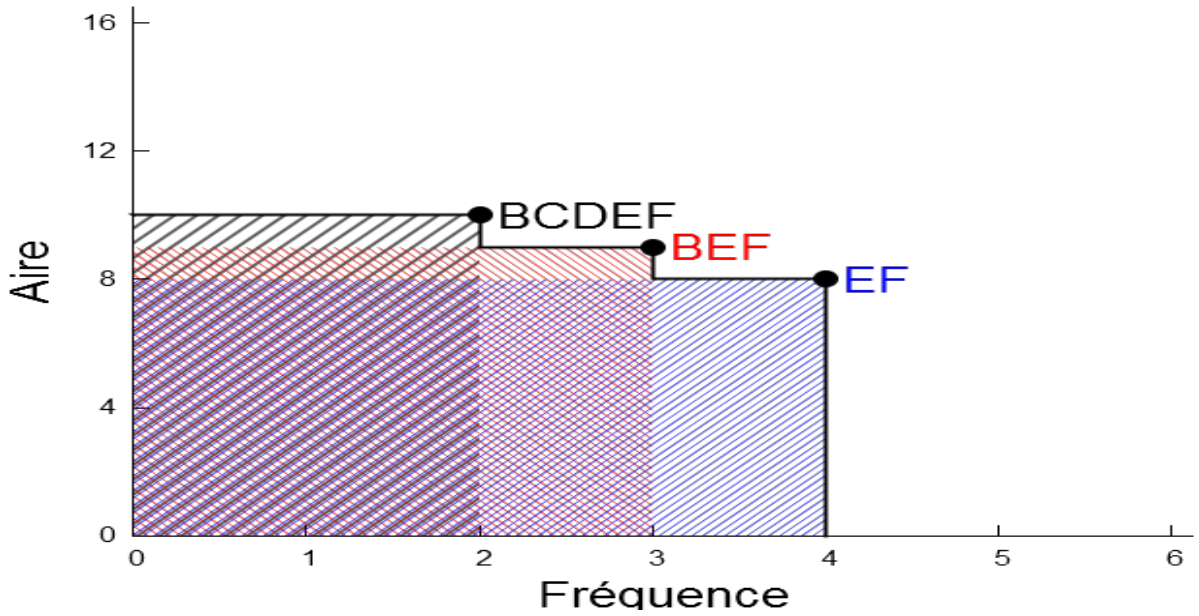
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p) \wedge (s_2 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1}, \underbrace{\text{BEF}}_{s_2}, \underbrace{\text{EF}}_{s_3},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



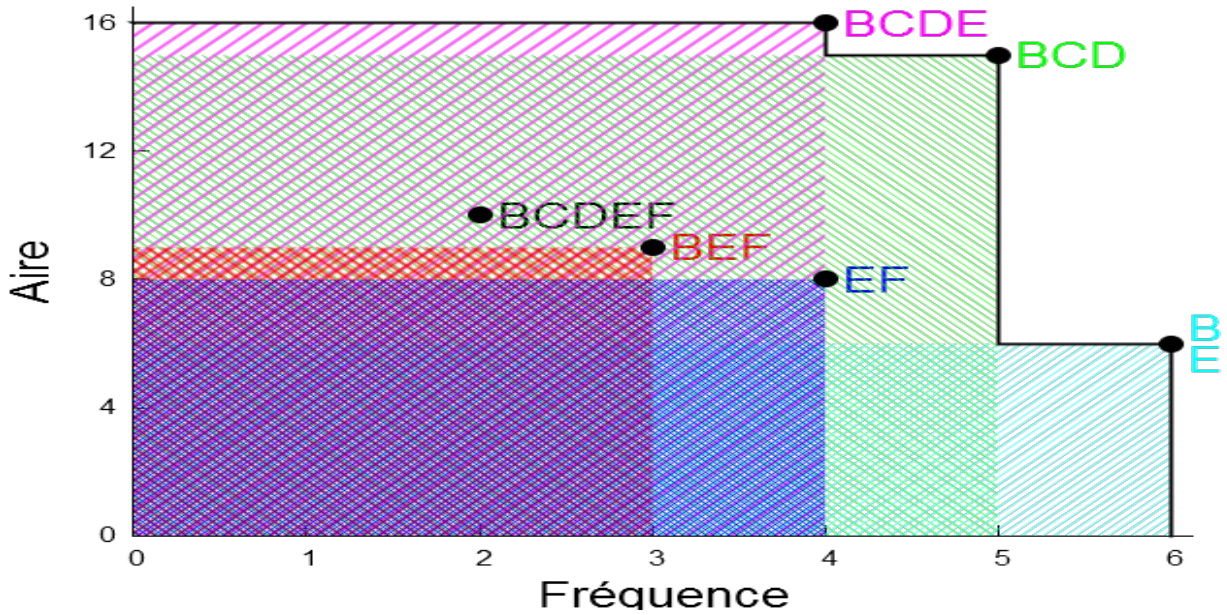
$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p) \wedge (s_2 \not\succ_M p) \wedge (s_3 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1}, \underbrace{\text{BEF}}_{s_2}, \underbrace{\text{EF}}_{s_3},$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F



$$M = \{\texttt{freq}, \texttt{aire}\}$$

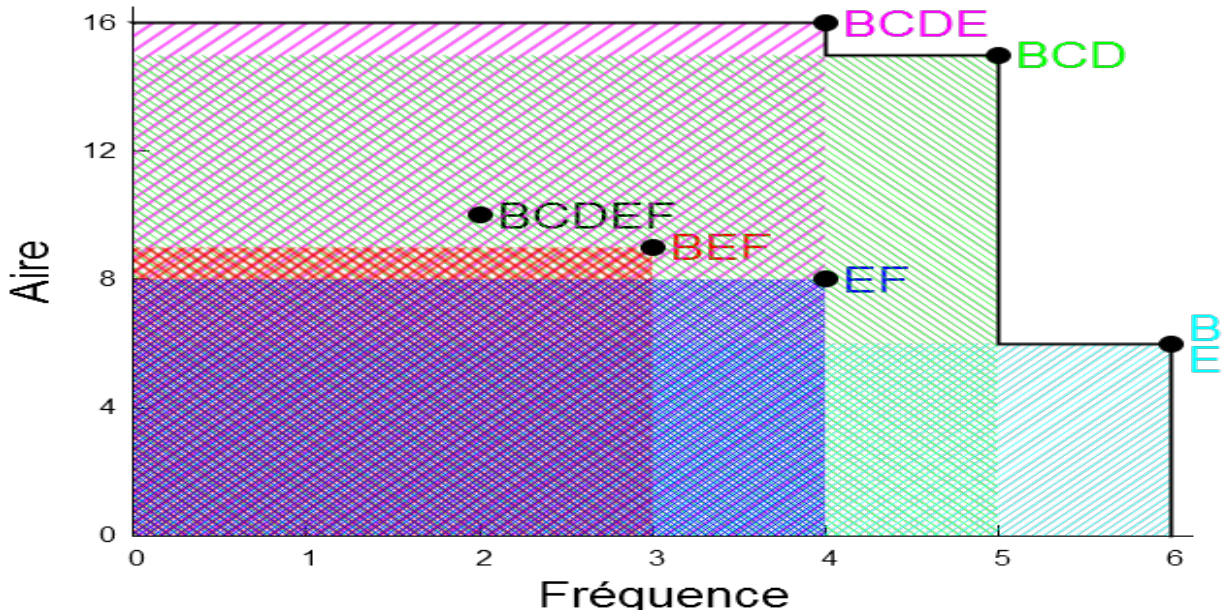
$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p) \wedge (s_2 \not\succ_M p) \wedge (s_3 \not\succ_M p) \wedge (s_4 \not\succ_M p) \wedge (s_5 \not\succ_M p) \wedge (s_6 \not\succ_M p)$$

$$Candidats = \{\underbrace{\text{BCDEF}}_{s_1}, \underbrace{\text{BEF}}_{s_2}, \underbrace{\text{EF}}_{s_3}, \underbrace{\text{BCDE}}_{s_4}, \underbrace{\text{BCD}}_{s_5}, \underbrace{\text{B}}_{s_6}, \underbrace{\text{E}}_{s_7}\}$$

Extraction des skypatterns en utilisant la PPC (2/2)

Trans.	Items					
t_1		B			E	F
t_2		B	C	D		
t_3	A				E	F
t_4	A	B	C	D	E	
t_5		B	C	D	E	
t_6		B	C	D	E	F
t_7	A	B	C	D	E	F

$|\mathcal{L}_{\mathcal{I}}| = 2^6 = 64$ motifs
4 skypatterns



$$M = \{\texttt{freq}, \texttt{aire}\}$$

$$q(p) \equiv \texttt{fermé}_{M'}(p) \wedge (s_1 \not\succ_M p) \wedge (s_2 \not\succ_M p) \wedge (s_3 \not\succ_M p) \wedge (s_4 \not\succ_M p) \wedge (s_5 \not\succ_M p) \wedge (s_6 \not\succ_M p)$$

$$Candidats = \{\underbrace{\cancel{BCDEF}}_{s_1}, \underbrace{\cancel{BEF}}_{s_2}, \underbrace{\cancel{EF}}_{s_3}, \underbrace{\underbrace{BCDE}_{s_4}, \underbrace{BCD}_{s_5}, \underbrace{B}_{s_6}, \underbrace{E}_{s_7}}_{Sky(Candidats, M)}\}$$