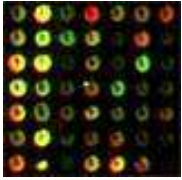


Extraction de concepts formels et d'autres types de bi-ensembles dans des données denses et bruitées

Jérémy Besson Jean-François Boulicaut Céline Robardet





Introduction

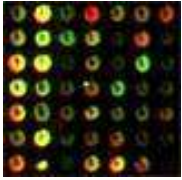
- ❖ Les données d'expression
- ❖ Différents types de motifs locaux
- ❖ Bases de données inductives
- ❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion

Introduction



Les données d'expression

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

- Mesure du niveau d'expression de gènes par différentes expériences utilisant des méthodes haut-débit
 - Enrichissement des données par des connaissances du domaine
- ⇒ Tableau booléen encodant plusieurs relations relatives à un même phénomène.

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

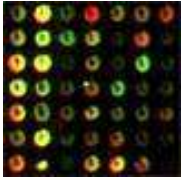
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Les données d'expression

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

G_4 et G_5 varient significativement dans les expériences biologiques E_3 et E_4 et les facteurs de transcription TF_1 et TF_2 peuvent expliquer cette co-expression.

Hypothèses biologiques :

- fonction identique
- même voie de régulation
- gènes co-régulés

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

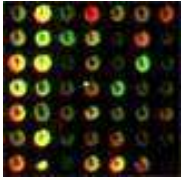
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de motifs locaux

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

- 1-RECTANGLE : $\{\{E_2E_3\}, \{G_1G_2\}\}$ est un 1-rectangle.

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

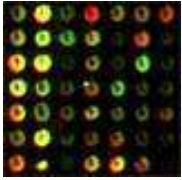
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de motifs locaux

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

- 1-RECTANGLE : $\{\{E_2E_3\}, \{G_1G_2\}\}$ est un 1-rectangle.
- ITEMSET : $\{\{E_2, E_3, E_4\}, \{G_1G_2\}\}$ est un itemset de fréquence 3.

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

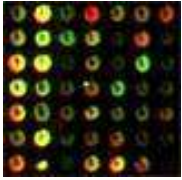
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de motifs locaux

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

- 1-RECTANGLE : $\{\{E_2E_3\}, \{G_1G_2\}\}$ est un 1-rectangle.
- ITEMSET : $\{\{E_2, E_3, E_4\}, \{G_1G_2\}\}$ est un itemset de fréquence 3.
- OBJECTSET : $\{\{E_2E_3\}, \{G_1, G_2, G_3, G_4\}\}$ est un objectset dont la fréquence est 4.

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

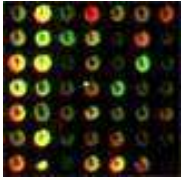
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de motifs locaux

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

- 1-RECTANGLE : $\{\{E_2E_3\}, \{G_1G_2\}\}$ est un 1-rectangle.
- ITEMSET : $\{\{E_2, E_3, E_4\}, \{G_1G_2\}\}$ est un itemset de fréquence 3.
- OBJECTSET : $\{\{E_2E_3\}, \{G_1, G_2, G_3, G_4\}\}$ est un objectset dont la fréquence est 4.
- CONCEPT FORMEL : $\{\{E_2E_3E_4\}, \{G_1G_2G_3G_4\}\}$ est un concept.

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

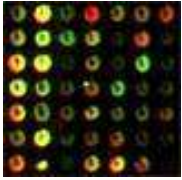
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de motifs locaux

Plus formellement...

- Données booléennes encodant $\mathbf{r} \subseteq \mathcal{O} \times \mathcal{I}$

- Correspondance de Galois

$$(X, Y) \in 2^{\mathcal{O}} \times 2^{\mathcal{I}}$$

$$\phi(X) = \{i \in \mathcal{I} \mid \forall x \in X, (x, i) \in \mathbf{r}\}$$

$$\psi(Y) = \{o \in \mathcal{O} \mid \forall y \in Y, (o, y) \in \mathbf{r}\}$$

- (X, Y) est un ITEMSET si $\psi(Y) = X$
- (X, Y) est un OBJECTSET si $\phi(X) = Y$
- (X, Y) est un CONCEPT FORMEL si $\psi(Y) = X$ et $\phi(X) = Y$

Introduction

❖ Les données d'expression

❖ Différents types de motifs locaux

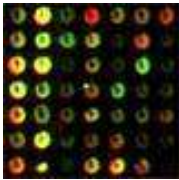
❖ Bases de données inductives

❖ Différents types de contraintes

Extraction de concepts formels

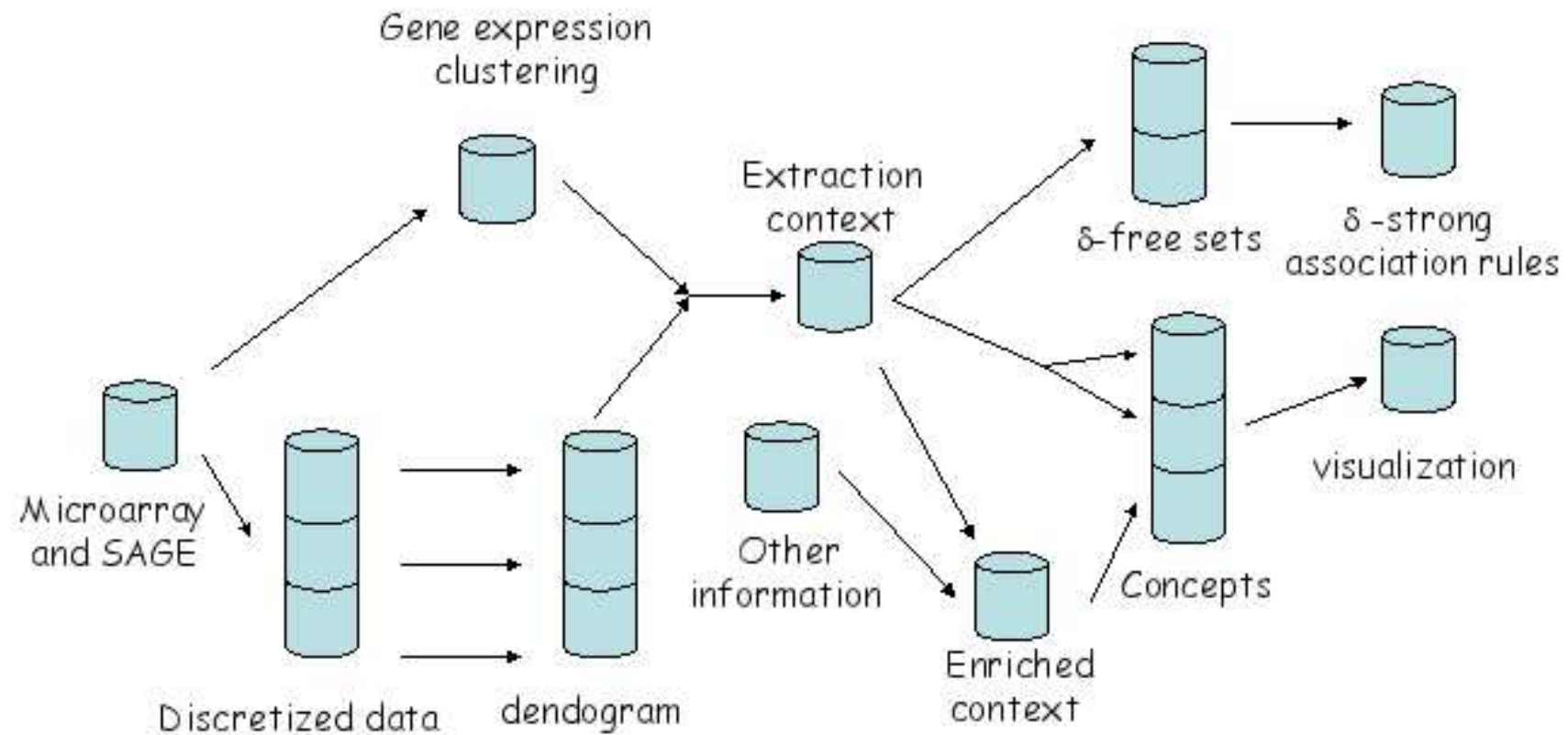
Motifs locaux robustes au bruit

Conclusion



Bases de données inductives

Requêtes inductives qui retournent **tous** les motifs (locaux) satisfaisant une certaine contrainte.



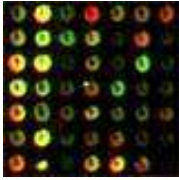
Introduction

- ❖ Les données d'expression
- ❖ Différents types de motifs locaux
- ❖ Bases de données inductives
- ❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Différents types de contraintes

Permettent de sélectionner des motifs pertinents

- Contraintes de fréquence minimale ou de taille minimale sur les objets ou items
- Présence ou absence d'objets ou d'items
- Contraintes de surface minimale

Permettent de rendre les extractions faisables

Enumération de type APRIORI ; (X, Y) un motif :

- ♦ $|X| > \tau_1$
- ♦ $x \in X$

Dû à la décroissance de X

- ♦ $|Y| < \tau_2$
- ♦ $y \notin Y$

Dû à la croissance de Y

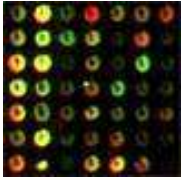
Introduction

- ❖ Les données d'expression
- ❖ Différents types de motifs locaux
- ❖ Bases de données inductives
- ❖ Différents types de contraintes

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion



Introduction

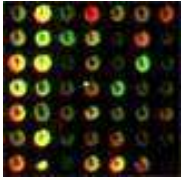
Extraction de concepts formels

- ❖ Pourquoi un nouvel algorithme ?
- ❖ D-MINER
- ❖ Evaluation expérimentale

Motifs locaux robustes au bruit

Conclusion

Extraction de concepts formels



Pourquoi un nouvel algorithme ?

Extraction de concepts dans des contextes booléens :

- denses
- dont les deux dimensions sont grandes

On souhaite pousser les contraintes de type

- $|X| > \tau_1$
- $|Y| > \tau_2$
- $|X| \times |Y| > \tau$

⇒ Enumération décroissante de X et de Y

Introduction

Extraction de
concepts formels

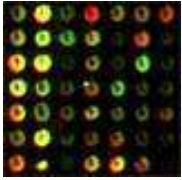
❖ Pourquoi un
nouvel
algorithme ?

❖ D-MINER

❖ Evaluation
expérimentale

Motifs locaux
robustes au bruit

Conclusion



D-Miner

Constraint-based mining of formal concepts in transactional data, PaKDD'04.

Constraint-based concept mining and its application to microarray data analysis, IDA'05.

	G_1	G_2	G_3	G_4
E_1	1	1	0	0
E_2	1	1	1	0
E_3	0	1	1	1

(E_1, G_3G_4)

Si $E_1 \notin X$ alors $G_3 \in Y$ ou $G_4 \in Y$ Si $E_1 \in X$, alors $G_3 \notin Y$ et $G_4 \notin Y$

Introduction

Extraction de concepts formels

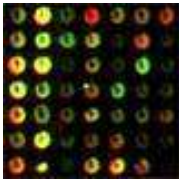
❖ Pourquoi un nouvel algorithme ?

❖ D-MINER

❖ Evaluation expérimentale

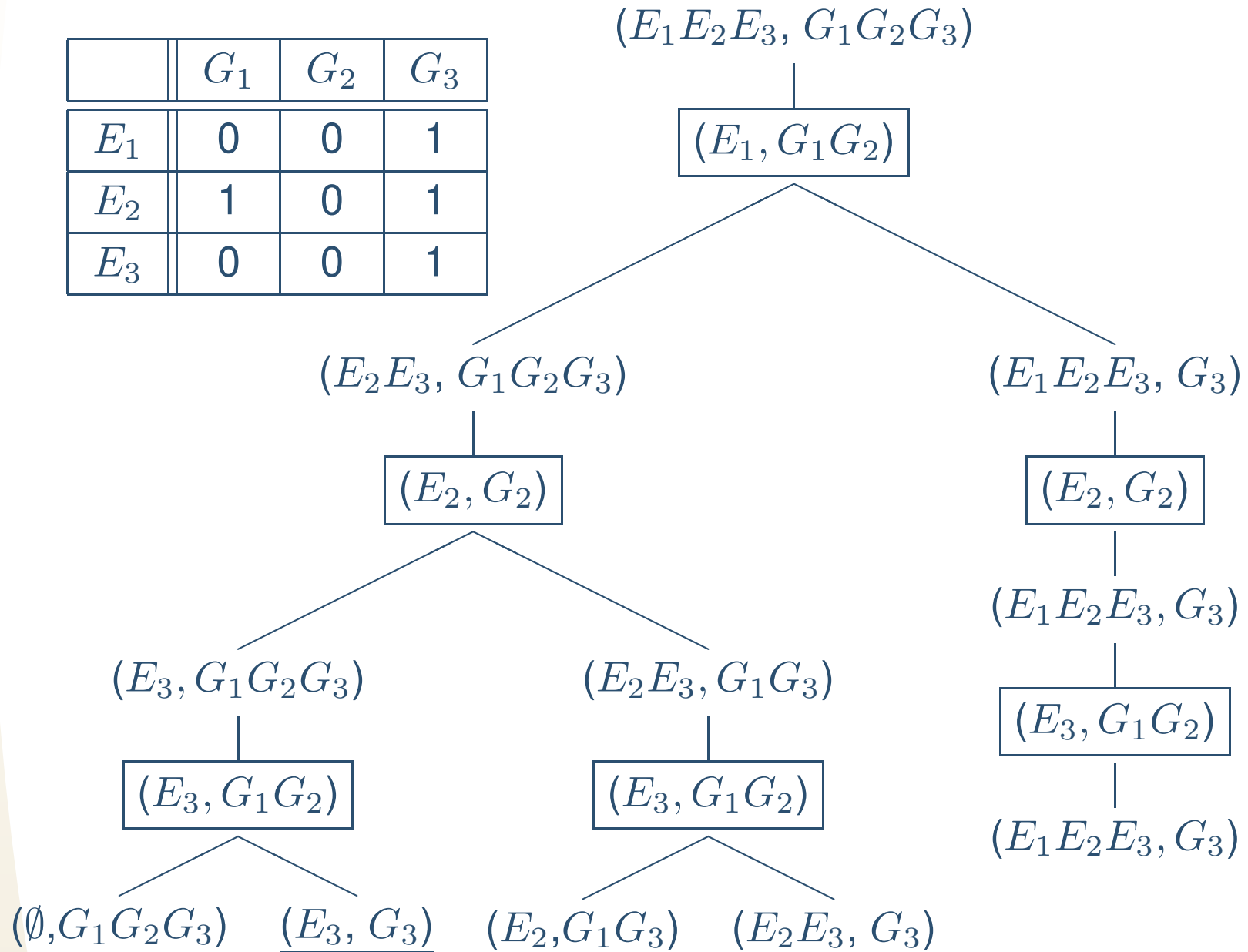
Motifs locaux robustes au bruit

Conclusion



D-Miner

	G_1	G_2	G_3
E_1	0	0	1
E_2	1	0	1
E_3	0	0	1



Introduction

Extraction de concepts formels

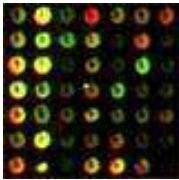
❖ Pourquoi un nouvel algorithme ?

❖ D-MINER

❖ Evaluation expérimentale

Motifs locaux robustes au bruit

Conclusion



D-Miner

Introduction

Extraction de concepts formels

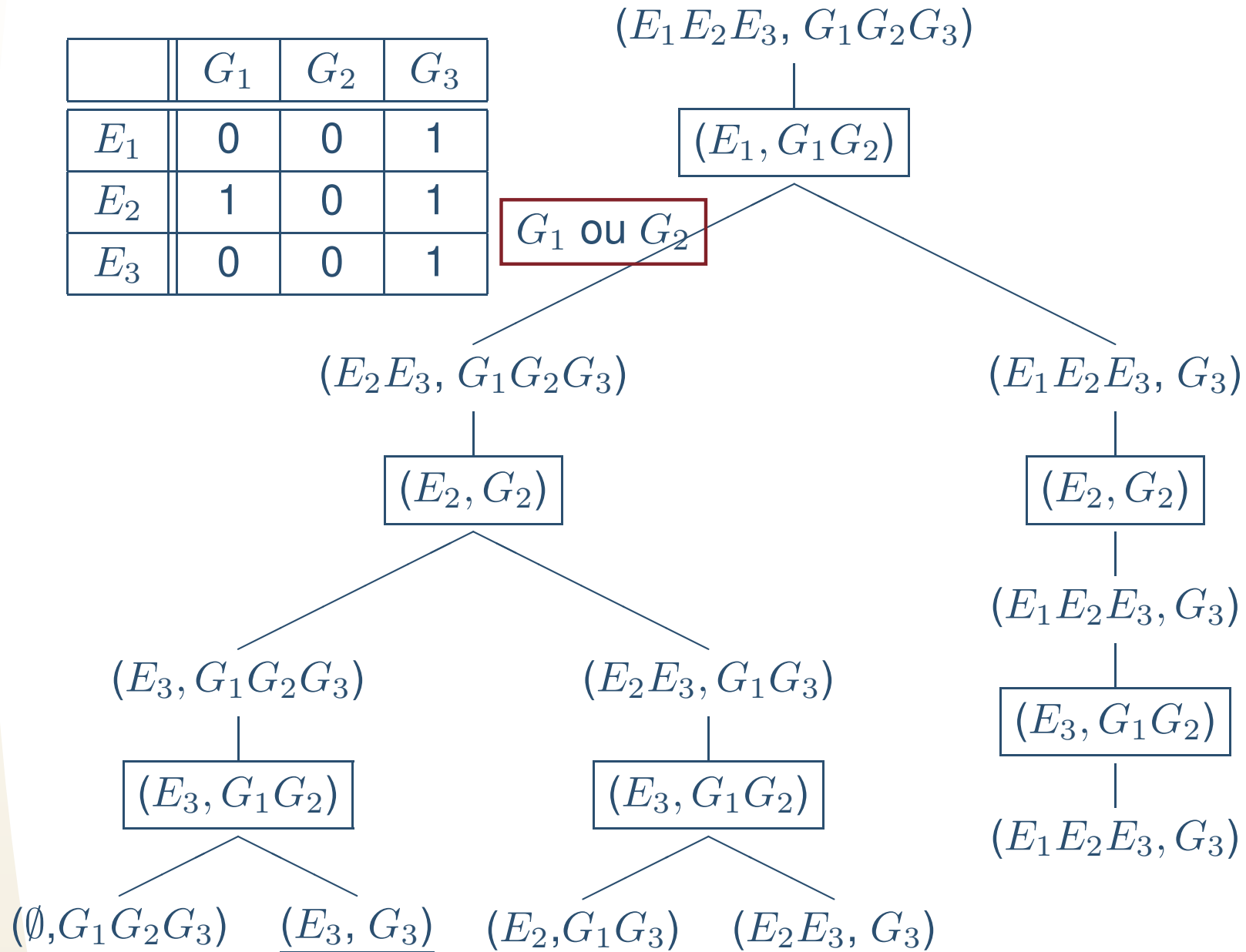
❖ Pourquoi un nouvel algorithme ?

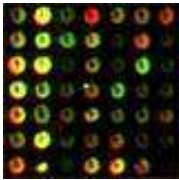
❖ D-MINER

❖ Evaluation expérimentale

Motifs locaux robustes au bruit

Conclusion





D-Miner

Introduction

Extraction de concepts formels

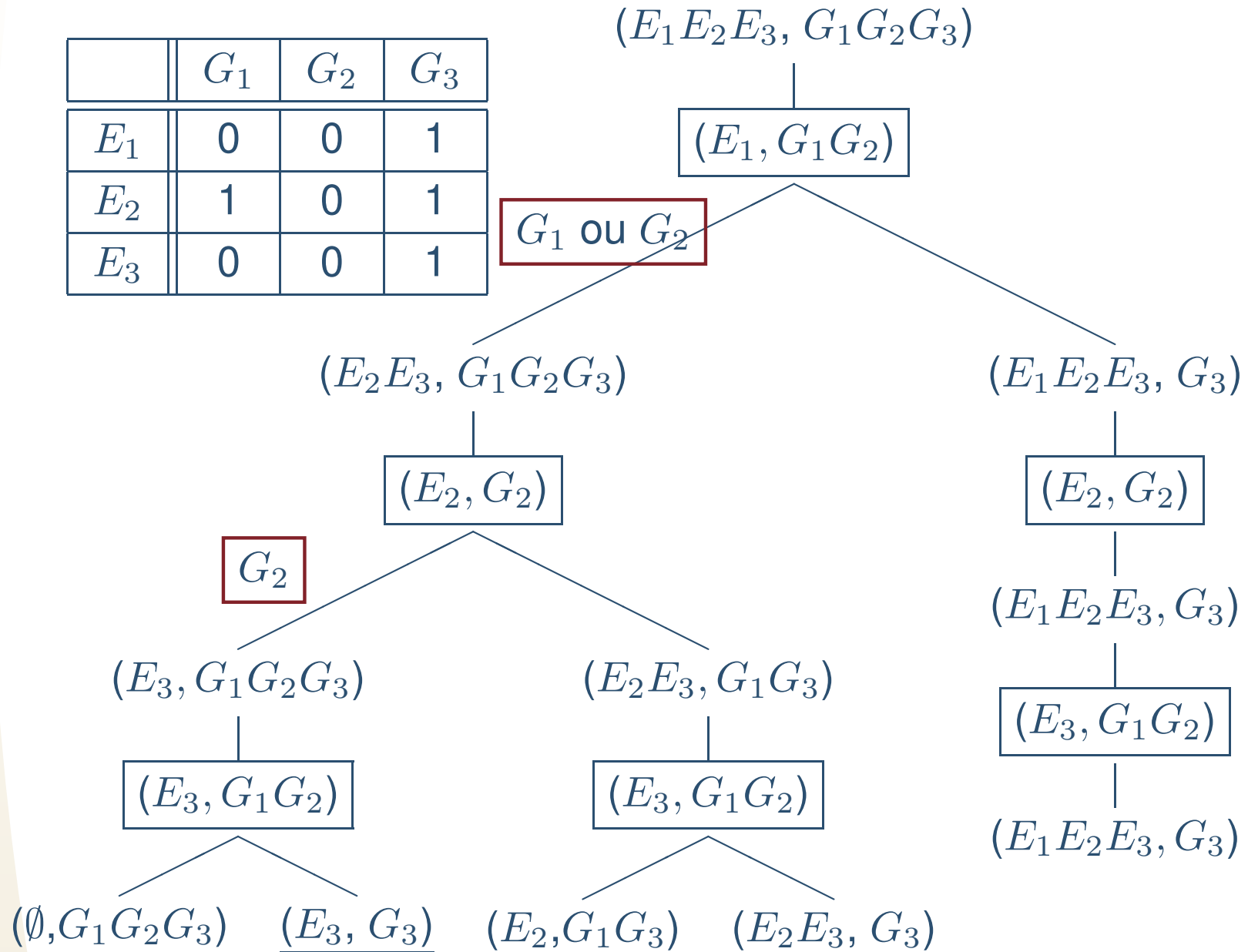
❖ Pourquoi un nouvel algorithme ?

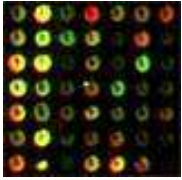
❖ D-MINER

❖ Evaluation expérimentale

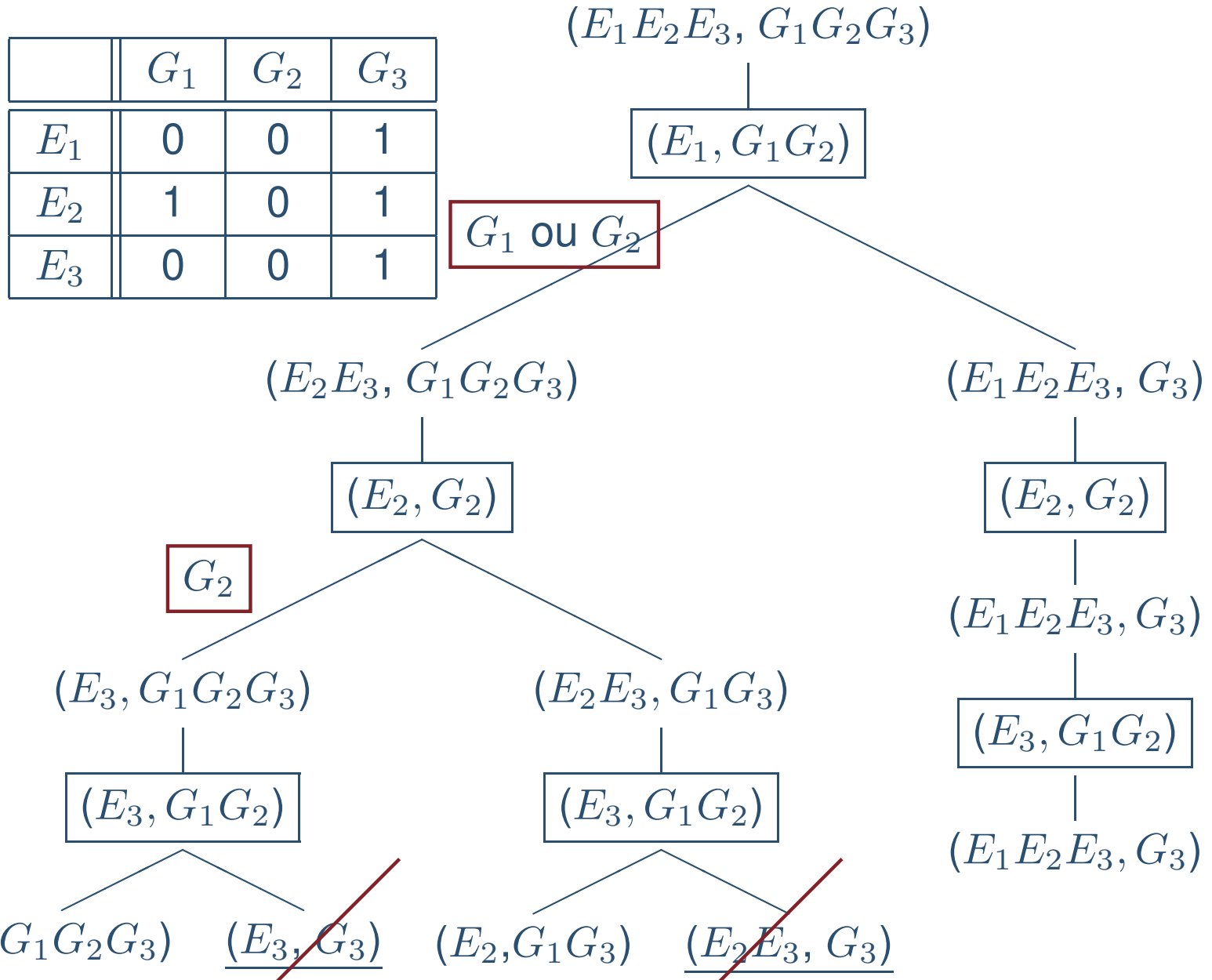
Motifs locaux robustes au bruit

Conclusion





D-Miner



Introduction

Extraction de concepts formels

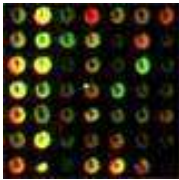
❖ Pourquoi un nouvel algorithme ?

❖ D-MINER

❖ Evaluation expérimentale

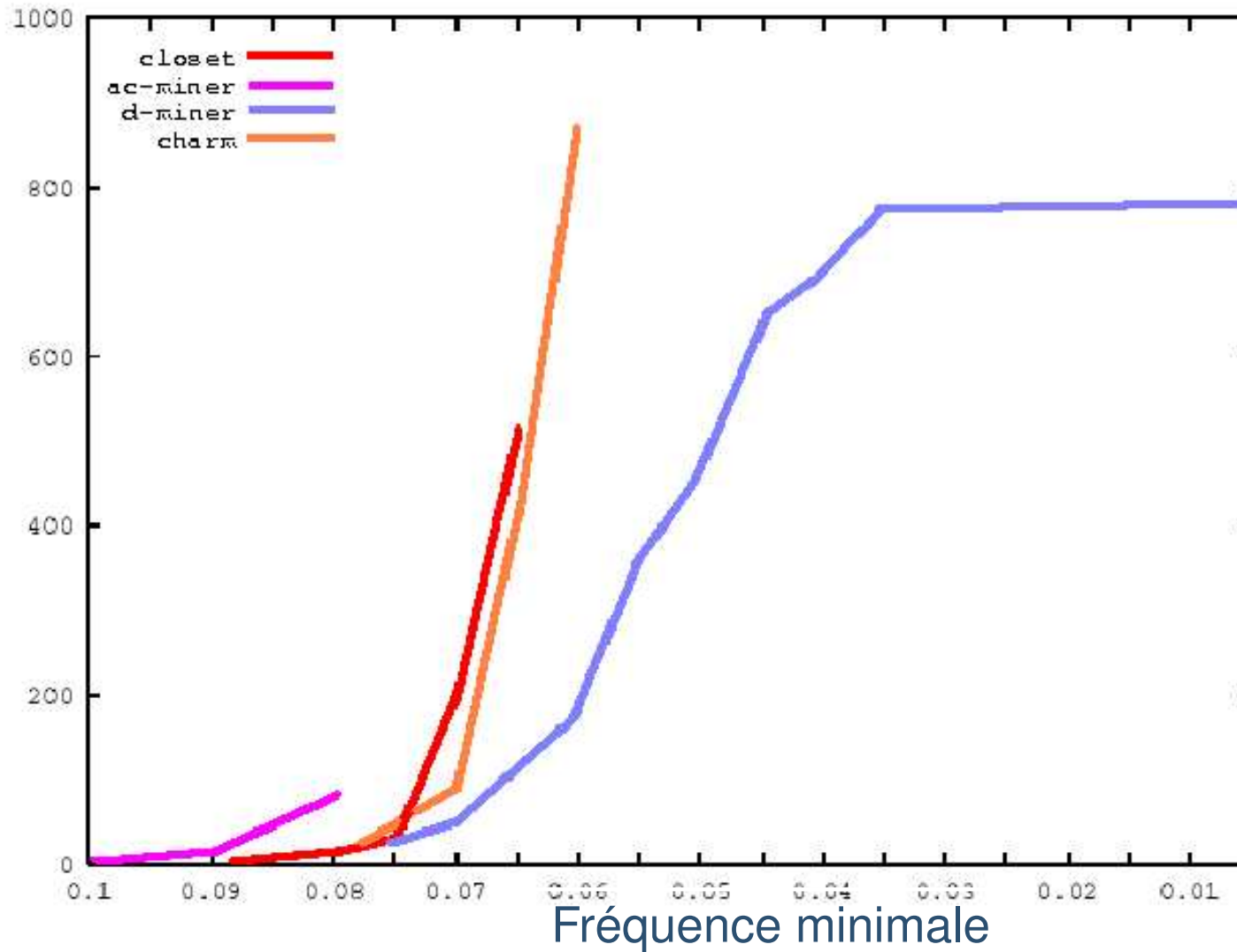
Motifs locaux robustes au bruit

Conclusion



Evaluation expérimentale

Temps



Introduction

Extraction de concepts formels

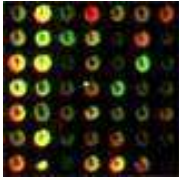
❖ Pourquoi un nouvel algorithme ?

❖ D-MINER

❖ Evaluation expérimentale

Motifs locaux robustes au bruit

Conclusion



Introduction

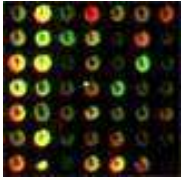
Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion

Motifs locaux robustes au bruit



Motivations

- Concepts formels : ensembles maximaux d'objets et d'items qui sont toujours en relation
- Contrainte souvent trop forte
- « Concepts » avec quelques exceptions devraient être utiles

	g1	g2	g3	g4	g5	g6	g7
s1	1	1	1	1	1	0	0
s2	1	1	1	1	1	0	0
s3	1	1	1	1	1	0	0
s4	1	1	1	1	1	0	0
s5	0	0	0	0	0	1	1
s6	0	0	0	0	0	1	1
s7	0	0	0	0	0	0	0

$(\{s1, s2, s3, s4\}, \{g1, g2, g3, g4, g5\})$

$(\{s5, s6\}, \{g6, g7\})$

Introduction

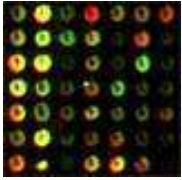
Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Motivations

- Concepts formels : ensembles maximaux d'objets et d'items qui sont toujours en relation
- Contrainte souvent trop forte
- « Concepts » avec quelques exceptions devraient être utiles

	g1	g2	g3	g4	g5	g6	g7
s1	1	0	1	1	1	0	0
s2	1	1	1	1	0	0	0
s3	1	1	1	1	1	0	0
s4	1	1	1	1	1	0	0
s5	0	0	0	0	0	1	1
s6	0	0	0	0	1	1	0
s7	1	0	0	0	0	0	0

- $(\{s1, s2, s3, s4, s7\}, \{g1\})$
- $(\{s1, s2, s3, s4\}, \{g1, g3, g4\})$
- $(\{s2, s3, s4\}, \{g1, g2, g3, g4\})$
- $(\{s3, s4\}, \{g1, g2, g3, g4, g5\})$
- $(\{s1, s3, s4\}, \{g1, g3, g4, g5\})$
- $(\{s1, s3, s4, s6\}, \{g5\})$
- $(\{s5, s6\}, \{g6\})$
- $(\{s5\}, \{g6, g7\})$
- $(\{s6\}, \{g5, g6\})$

Introduction

Extraction de concepts formels

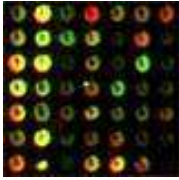
Motifs locaux robustes au bruit

❖ Motivations

- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion

10% de bruit



Motivations

Propriétés souhaitées :

- Complétude des extractions
- Nombre borné de valeurs 0 par ligne (α) et par colonne (β) - en absolu, ou relativement à la taille du bi-ensemble
- Maximalité dans les deux dimensions
- Pertinence des motifs : il n'existe pas de ligne à l'extérieur du motif qui soit identique à une ligne dans celui-ci (resp. colonne).
- Existence de deux fonctions $\phi : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{I}}$ et $\psi : 2^{\mathcal{I}} \rightarrow 2^{\mathcal{O}}$
- Correspondance de Galois
 1. $\forall v, w \in \mathcal{O}$, si $v \leq_{\mathcal{O}} w$ alors $\phi(w) \leq_{\mathcal{I}} \phi(v)$,
 2. $\forall i, j \in \mathcal{I}$, si $i \leq_{\mathcal{I}} j$ alors $\psi(j) \leq_{\mathcal{O}} \psi(i)$,
 3. $\forall v \in \mathcal{O}, \forall i \in \mathcal{I}, v \leq_{\mathcal{O}} \psi(\phi(v))$ et $i \leq_{\mathcal{I}} \phi(\psi(i))$
- Faisabilité du calcul en pratique sur des données réelles

Introduction

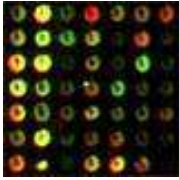
Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Motivations

Nos propositions :

- **Classification hiérarchique de concepts** : *Using classification and visualization on pattern databases for gene expression data analysis, PaRMA'04*
- **δ -libres et leur pseudo fermeture** : *Free-sets: a condensed representation of boolean data for the approximation of frequency queries, DMKD*
- **Fusion de concepts** : *Mining Formal Concepts with a Bounded Number of Exceptions from Transactional Data, KDID 2004*
- **Bi-ensembles denses et pertinents** : *accepté à CAp*

Introduction

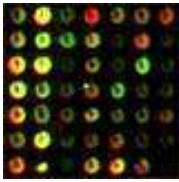
Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Classification hiérarchique de concepts

- Post-traitement de concepts par clustering hiérarchique
- Prendre en compte simultanément les objets et les items

$$d(C_i, C_j) = \frac{1}{2} \frac{O_i \Delta O_j}{O_i \cup O_j} + \frac{1}{2} \frac{I_i \Delta I_j}{I_i \cup I_j}$$

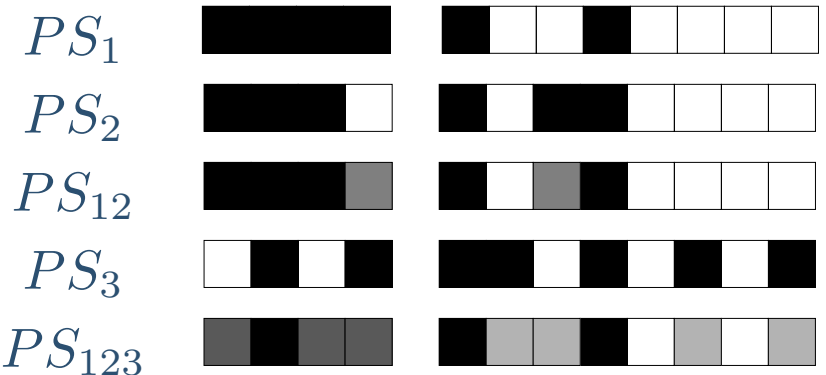
avec $C_i = (O_i, I_i)$, $C_j = (O_j, I_j)$ et $S_i \Delta S_j = S_i \cup S_j \setminus S_i \cap S_j$

- Pseudo concept : une représentation unique des concepts d'une même classe
- Basé sur la fréquence d'apparition des objets et des items dans la classe

$(\{E_1, E_2, E_3, E_4\}, \{G_1, G_4\})$

$(\{E_1, E_2, E_3\}, \{G_1, G_3, G_4\})$

$(\{E_2, E_4\}, \{G_1, G_2, G_4, G_6, G_8\})$



$$PS_{123} = (\{(E_1; 2/3), (E_2; 1), (E_3; 2/3), (E_4; 2/3)\}, \{(G_1; 1), (G_2; 1/3), (G_3; 1/3), (G_4; 1), (G_6; 1/3), (G_8; 1/3)\}, 3)$$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

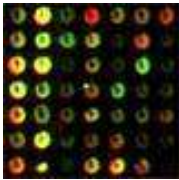
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

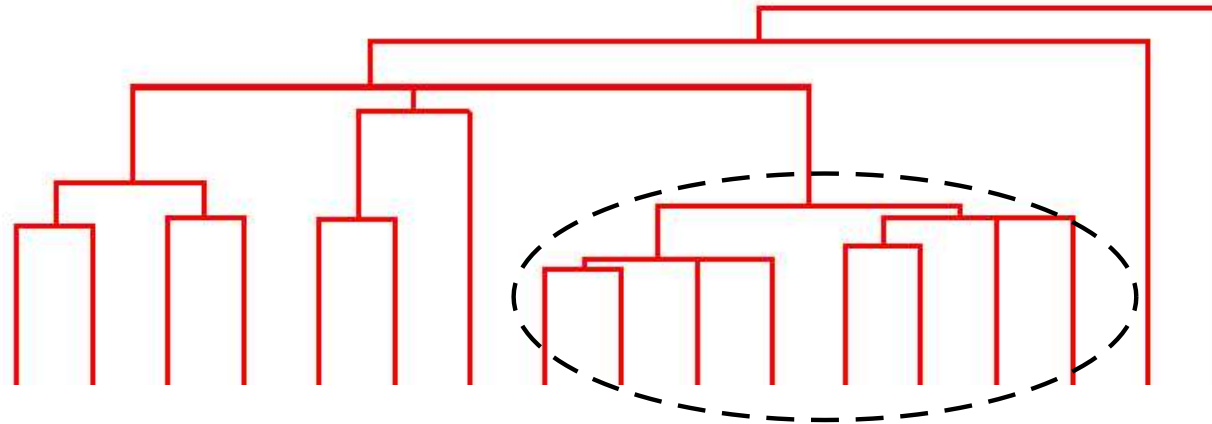
❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Classification hiérarchique de concepts



	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

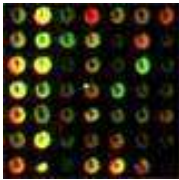
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

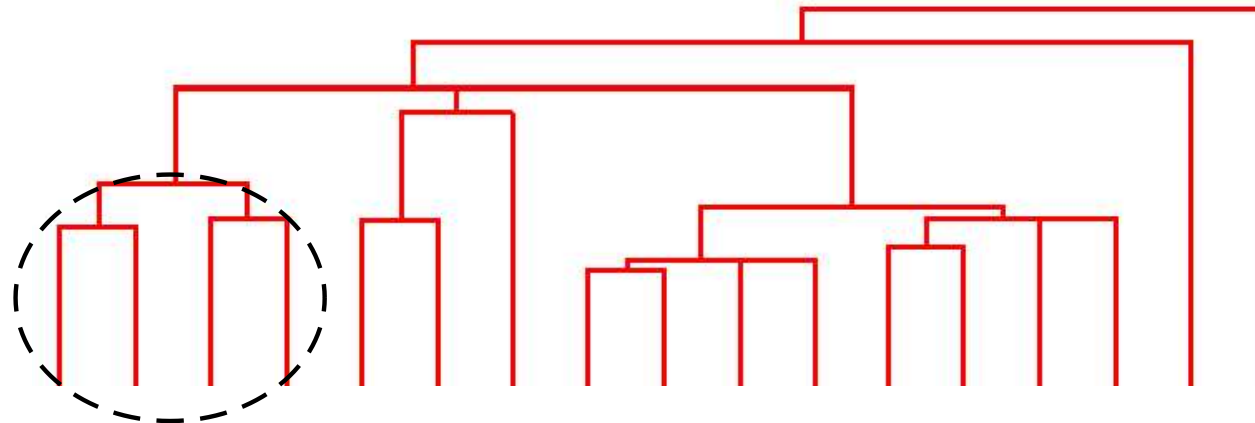
❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Classification hiérarchique de concepts



	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	0	0

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

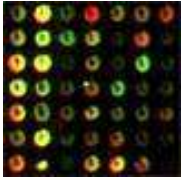
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Classification hiérarchique de concepts

Bons points

- Visualisation
- Extraction faisable

Inconvénients

- Pas de propriétés formelles sur les motifs extraits

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

❖ Ensembles

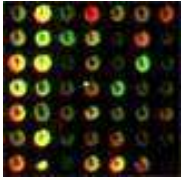
fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Ensembles fermés, libres et δ -libres

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations
❖ Classification hiérarchique de concepts

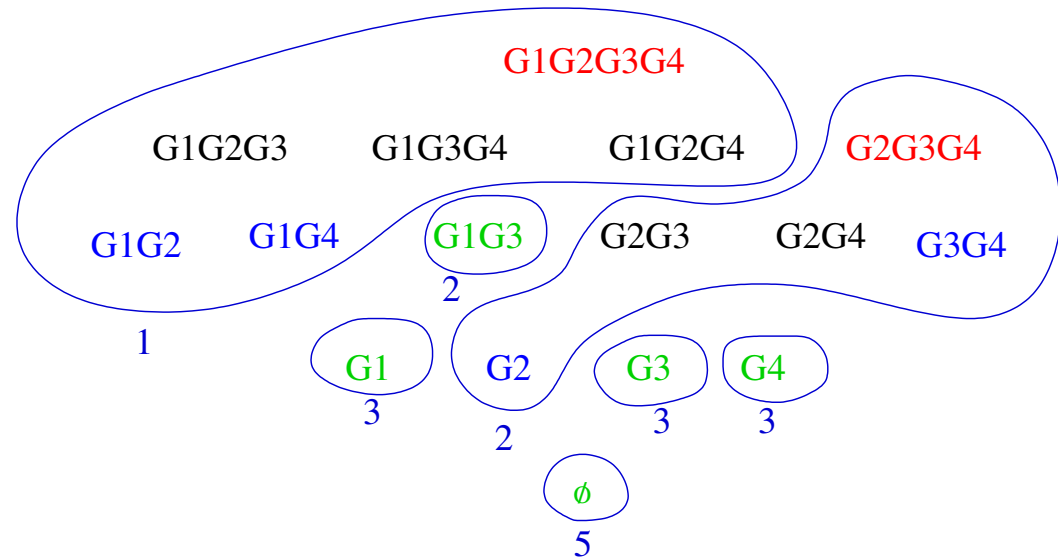
❖ Ensembles fermés, libres et δ -libres

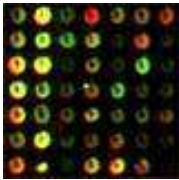
❖ Bi-ensembles à partir des δ -libres
❖ Fusion de concepts
❖ Concepts denses et pertinents

Conclusion

- $\text{ferm}_{item}(Y) = Y'$ où Y' est le plus grand sur-ensemble de Y tel que $\psi(Y') = \psi(Y)$
- $\text{ferm}_{item}(Y) = \phi(\psi(Y))$
- Relation d'équivalence : $(X, Y) \sim (X', Y') \Leftrightarrow \text{ferm}_{item}(Y) = \text{ferm}_{item}(Y')$
- Classes d'équivalence
 - ♦ L'itemset **fermé** est l'élément maximum de cette classe
 - ♦ Les itemsets **libres** sont les éléments minimaux de cette classe (également appelés motifs clés)

	G_1	G_2	G_3	G_4
	1	0	1	0
	1	1	1	1
	0	1	1	1
	0	0	0	1
	1	0	0	0

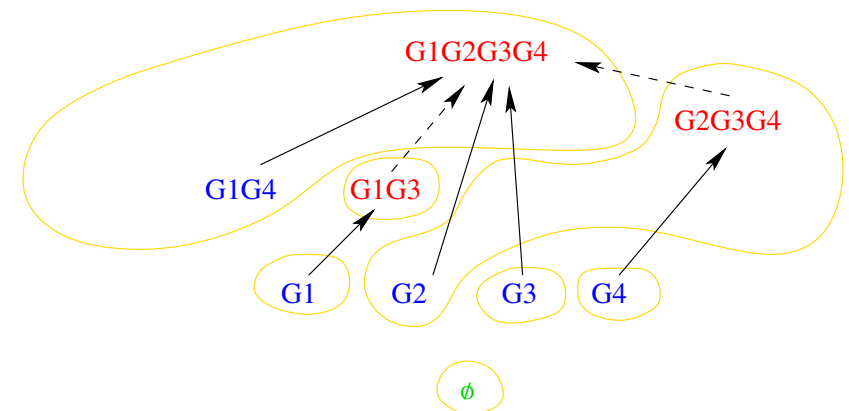
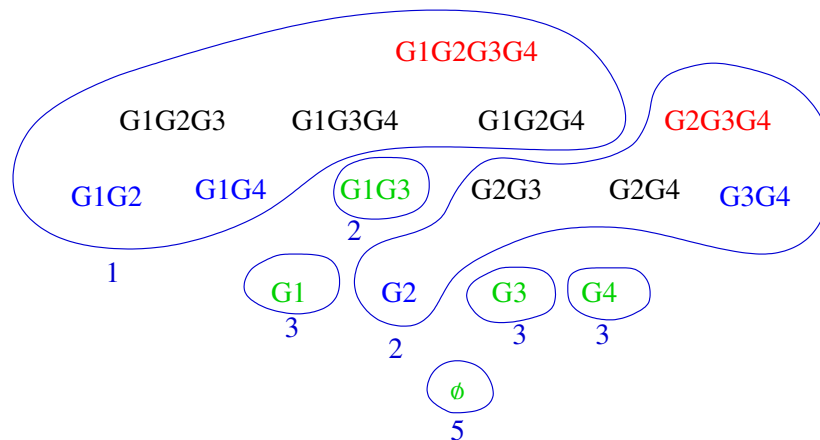




Ensembles fermés, libres et δ -libres

- Y est libre ssi $\forall Z \subset Y, \psi(Z) \neq \psi(Y)$
- Généralisation avec δ un entier positif
 Y est δ -libre ssi $\forall Z \subset Y, \#(\psi(Z)) - \#(\psi(Y)) > \delta$

G_1	G_2	G_3	G_4
1	0	1	0
1	1	1	1
0	1	1	1
0	0	0	1
1	0	0	0



Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

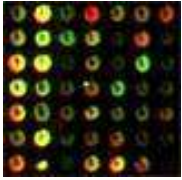
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Bi-ensembles à partir des δ -libres

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Un 1-libre et son support : $\{\{E_2E_3E_4TF_1\}, \{G_2\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

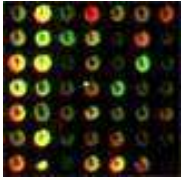
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Bi-ensembles à partir des δ -libres

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Un 1-libre et son support : $\{\{E_2E_3E_4TF_1\}, \{G_2\}\}$
- Sa pseudo fermeture : $\{\{E_2E_3E_4TF_1\}, \{G_1G_2G_3G_4G_5\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

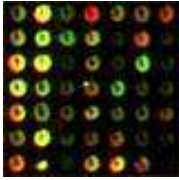
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Bi-ensembles à partir des δ -libres

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Un 1-libre et son support : $\{\{E_2E_3E_4TF_1\}, \{G_2\}\}$
- Sa pseudo fermeture : $\{\{E_2E_3E_4TF_1\}, \{G_1G_2G_3G_4G_5\}\}$
- Un autre 1-libre et son support : $\{\{TF_1 TF_2\}, \{G_5G_6\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

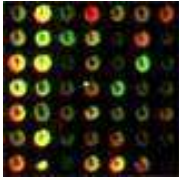
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Bi-ensembles à partir des δ -libres

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Un 1-libre et son support : $\{\{E_2E_3E_4TF_1\}, \{G_2\}\}$
- Sa pseudo fermeture : $\{\{E_2E_3E_4TF_1\}, \{G_1G_2G_3G_4G_5\}\}$
- Un autre 1-libre et son support : $\{\{TF_1 TF_2\}, \{G_5G_6\}\}$
- Sa pseudo fermeture : $\{\{TF_1 TF_2\}, \{G_2G_3G_4G_5G_6G_7\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

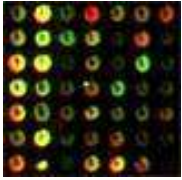
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Bi-ensembles à partir des δ -libres

Bons points :

- Complétude
- Nombre de valeurs 0 borné sur les colonnes
- Maximalité en colonne
- Pertinence en colonne
- Il existe une fonction ϕ
- Faisabilité du calcul

Inconvénients :

- Nombre de valeurs 0 non borné sur les lignes
- Pas de maximalité sur les lignes
- Pas de pertinence sur les lignes
- Pas de fonction ψ

⇒ Approche dissymétrique

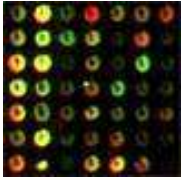
Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

Trois contraintes :

- $\mathcal{C}_{\alpha\beta}$: le nombre de valeurs 0 par ligne et par colonne est borné (respectivement par α et β).
- Les bi-ensembles sont *consistants* :
 - ◆ Soit on ajoute tous les items identiques (par rapport à l'ensemble d'objets) ou tous les objets identiques (par rapport l'ensemble d'items) dans le bi-ensemble si $\mathcal{C}_{\alpha\beta}$ est satisfaite,
 - ◆ Soit on les exclut tous si ce n'est pas le cas.
- Les bi-ensembles sont maximaux.

Idée :

- Fusion de concepts

Introduction

Extraction de concepts formels

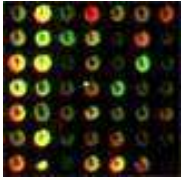
Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

- ❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

Adaptation d'algorithmes d'extraction de motifs maximaux (e.g., itemsets maximaux) :

- Items : les concepts extraits
- Ensemble d'items : l'union des concepts
- Contrainte anti-monotone : $\mathcal{C}_{\alpha\beta}$
- Pas d'objet, contrainte seulement syntaxique.

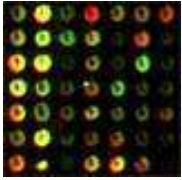
Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Quatre concepts

- ◆ $\{\{E_2E_3E_4TF_1\}, \{G_3G_4\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

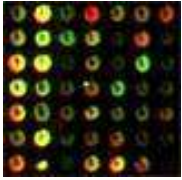
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Quatre concepts

- ◆ $\{\{E_2E_3E_4TF_1\}, \{G_3G_4\}\}$
- ◆ $\{\{E_3E_4TF_1TF_2\}, \{G_4G_5\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

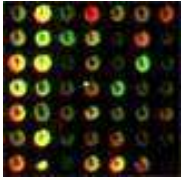
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Quatre concepts

- ◆ $\{\{E_2E_3E_4TF_1\}, \{G_3G_4\}\}$
- ◆ $\{\{E_3E_4TF_1TF_2\}, \{G_4G_5\}\}$
- ◆ $\{\{E_1E_2E_3E_4TF_1\}, \{G_3\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

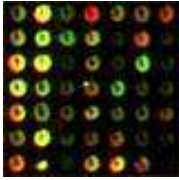
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Quatre concepts

- ◆ $\{\{E_2E_3E_4TF_1\}, \{G_3G_4\}\}$
- ◆ $\{\{E_3E_4TF_1TF_2\}, \{G_4G_5\}\}$
- ◆ $\{\{E_1E_2E_3E_4TF_1\}, \{G_3\}\}$
- ◆ $\{\{E_1E_3E_4TF_1TF_2\}, \{G_5\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

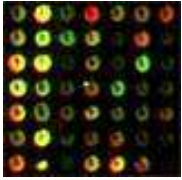
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- Quatre concepts

- ◆ $\{\{E_2E_3E_4TF_1\}, \{G_3G_4\}\}$
- ◆ $\{\{E_3E_4TF_1TF_2\}, \{G_4G_5\}\}$
- ◆ $\{\{E_1E_2E_3E_4TF_1\}, \{G_3\}\}$
- ◆ $\{\{E_1E_3E_4TF_1TF_2\}, \{G_5\}\}$

- Leur fusion : $\{\{E_1E_2E_3E_4TF_1TF_2\}, \{G_3G_4G_5\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

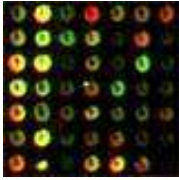
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

Deux motifs issus du processus de fusion :

- $\{\{E_1 E_2 E_3 E_4 TF_1 TF_2\}, \{G_3 G_4 G_5\}\}$ ———
- $\{\{E_1 E_2 E_3 E_4 TF_1\}, \{G_1 G_3 G_5\}\}$ - - - - -

⇒ Pas de correspondance de Galois

On a (O, I) et (O', I') tels que $O \subset O'$ et $I' \not\subset I$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

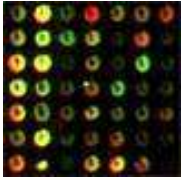
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Fusion de concepts

Bons points :

- Nombre de valeurs 0 borné en ligne et en colonne
- Maximalité sur les deux dimensions
- Pertinence en ligne et en colonne

Inconvénients :

- Pas de complétude
- On n'a ni une fonction ϕ ni une fonction ψ
- L'extraction est un peu difficile

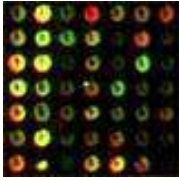
Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Concepts denses et pertinents

Introduction

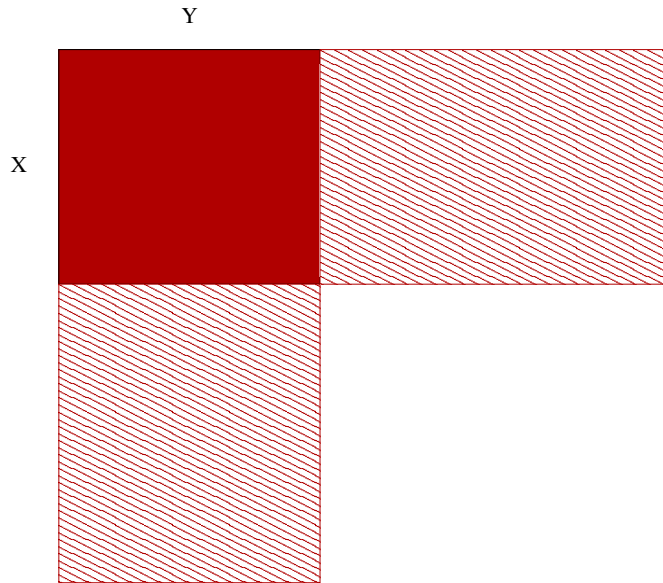
Extraction de concepts formels

Motifs locaux robustes au bruit

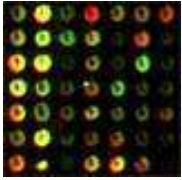
- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



$$\max_{s \in S} (\#\{g \in G \mid (s, g) \notin \mathbf{r}\}) \leq \min\{\alpha, \min_{s \in S \setminus S} (\#\{g \in G \mid (s, g) \notin \mathbf{r}\}) - \delta\}$$
$$\max_{g \in G} (\#\{s \in S \mid (s, g) \notin \mathbf{r}\}) \leq \min\{\beta, \min_{g \in G \setminus G} (\#\{s \in S \mid (s, g) \notin \mathbf{r}\}) - \delta'\}$$



Concepts denses et pertinents

En fonction de α , β , δ et δ' , on retrouve des collections bien connues :

- 1-RECTANGLE
- ITEMSET
- OBJECTSET
- CONCEPT FORMEL

et des nouvelles

- Des 1-RECTANGLES à trous
- Des ITEMSETS à trous
- Des OBJECTSETS à trous
- Des CONCEPTS FORMELS à trous

Ces bi-ensembles sont munis d'une correspondance de Galois.

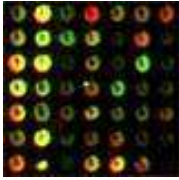
Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Concepts denses et pertinents

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
E_1	1	0	1	0	1	0	0
E_2	1	1	1	1	0	1	0
E_3	1	1	1	1	1	0	0
E_4	1	1	1	1	1	0	0
TF_1	0	1	1	1	1	1	1
TF_2	0	0	0	1	1	1	0
TF_3	1	0	0	0	0	1	0

- $\{\{E_1 E_2 E_3 E_4 TF_1 TF_2\}, \{G_3 G_4 G_5\}\}$

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

❖ Motivations

❖ Classification hiérarchique de concepts

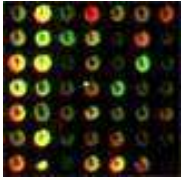
❖ Ensembles fermés, libres et δ -libres

❖ Bi-ensembles à partir des δ -libres

❖ Fusion de concepts

❖ Concepts denses et pertinents

Conclusion



Concepts denses et pertinents

Bons points :

- Complétude
- Nombre de valeurs 0 borné en lignes et en colonnes
- Pertinence des motifs
- Fonctions
- Correspondance de Galois

Inconvénients :

- Calculs très difficiles en pratique.

Mais approche incomplète possible pour l'extension de certains concepts

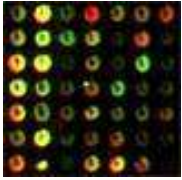
Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

- ❖ Motivations
- ❖ Classification hiérarchique de concepts
- ❖ Ensembles fermés, libres et δ -libres
- ❖ Bi-ensembles à partir des δ -libres
- ❖ Fusion de concepts
- ❖ Concepts denses et pertinents

Conclusion



Introduction

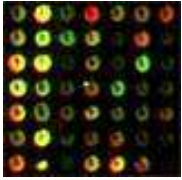
Extraction de
concepts formels

Motifs locaux
robustes au bruit

Conclusion

❖ Conclusion et
perspectives

Conclusion



Conclusion et perspectives

- Les Motifs denses et pertinents possèdent toutes les propriétés requises... sauf la faisabilité
- Les δ -libres et leur fermeture sont extractable mais leur définition n'est pas complètement satisfaisante
- Quel compromis pouvons-nous trouver ?

Introduction

Extraction de concepts formels

Motifs locaux robustes au bruit

Conclusion

❖ Conclusion et perspectives

Ce travail a partiellement été supporté par l'ACI DM 46 BINGO.