Conception et implémentation de nouvelles fonctionnalités dans un prototype de fouille de données

Kévin Emamirad

Université du Québec en Outaouais
Département d'informatique et d'ingénierie

emak01@uqo.ca

mercredi 31 mai 2017

Plan de la présentation

1. Introduction

2. Rappels

Introduction

- L'analyse formelle de concepts (AFC) (Ganter et Wille 1999) est un formalisme de représentation des connaissances et de fouille de données qui est
  - utilisé dans divers domaines (informatique, linguistique, sociologie, biologie, etc.), et
  - produit des visualisations graphiques des structures inhérentes aux données sous forme de treillis de concepts (Galois).
- Au cours des deux dernières décennies, on a vu apparaître plusieurs d'outils d'analyse formelle de concepts tels *ToscanaJ*, *ConExp*, *Coron*, *Java Lattices*, et *Lattice Miner*.

- But : Enrichir *Lattice Miner* (Roberge 2007), un prototype de fouille de données qui exploite l'AFC avec les objectifs suivants :
  1. production d'implications avec négation (Missaoui, Nourine et Renaud 2012),
  2. enrichissement du module de génération des règles triadiques pour obtenir exhaustivement et précisément les trois formes d'implications triadiques définies par (Ganter et Obiedkov 2004),
  3. validation intensive de la procédure de production de la base d'implications de (Guigues et Duquenne 1986) et de la procédure de calcul des relations de flèches. Cette dernière est utile dans le processus de décomposition de contextes formels (Viaud et al. 2015), et
  4. amélioration de la convivialité de l'interface usager.

Rappels I

Définition 1 : Contexte formel

Soit $\mathbb{K} = (G, M, I)$ *un contexte formel* où G, M et I sont respectivement un ensemble d'objets, une collection d'attributs et une relation binaire entre G et M. L'expression $(g, m) \in I$ ou encore gIm signifie que l'objet g possède l'attribut m.

	a	b	c	d	e	f
1	$\times$		$\times$		$\times$	
2		$\times$		$\times$		$\times$
3			$\times$	$\times$		
4		$\times$	$\times$			$\times$
5	$\times$	$\times$				$\times$
6	$\times$	$\times$		$\times$	$\times$	

Tableau 1: Exemple de contexte $\mathbb{K}$

Rappels II

Définition 2 : Concept formel

Un *concept formel* c est une paire d'ensembles $c := (A, B)$ avec $A \subseteq G$, $B \subseteq M$, $A = B'$ et $B = A'$, où A' est l'ensemble des attributs partagés par les objets dans A et B' est l'ensemble des objets ayant tous leurs attributs dans B. Les sous-ensembles A et B sont appelés respectivement l'extension et l'intention du concept c. Les valeurs de A' et B' sont obtenues comme suit : $A' := \{m \in M \mid gIm \; \forall g \in A\}$ et $B' := \{g \in G \mid gIm \; \forall \; m \in B\}$.

Défintion 3 : Sous-ensemble fermé

Un sous-ensemble X est fermé si $X'' = X$. Un concept objet pour l'objet g est une paire de la forme $\gamma(g) = (g'', g')$ alors que le concept attribut pour l'attribut m est $\mu(m) = (m', m'')$. Les sous-ensembles fermés de G sont les extensions alors que les sous-ensembles fermés de M sont les intentions de $\mathbb{K}$.

Rappels III

Défintion 4 : Treillis de concepts

Un *treillis de concepts* $\underline{\mathfrak{B}}(G, M, I)$ est un treillis résultant de l'ordre partiel existant entre les concepts du contexte $\mathbb{K} = (G, M, I)$.

$$(A, B) \leq (C, D) \Leftrightarrow A \subseteq C \text{ et } D \subseteq B$$

(A, B) est alors un sous-concept ou prédecesseur immédiat de (C, D) alors que ce dernier est un successeur de (A, B).

Définition 5 : Les bornes du treillis

La borne inférieure $\wedge$ (*meet*) d'un ensemble de concepts (X_i, Y_i) avec $i = 1, \ldots k$ est le plus grand des prédécesseurs communs. De même, la borne supérieure $\vee$ (*join*) d'un ensemble de concepts est le petit des successeurs communs.

Rappels IV

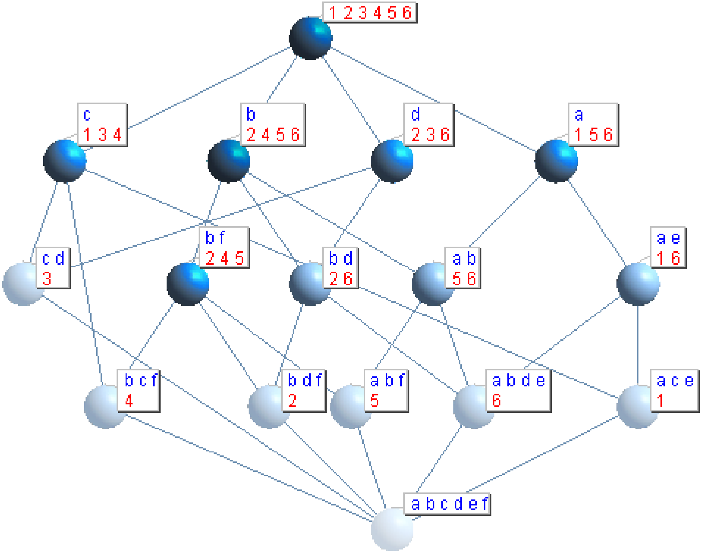

Figure 1: Exemple de treillis de concepts pour le contexte du tableau 1

Blocks of Highlighted Text

Block 3

Suspendisse tincidunt sagittis gravida. Curabitur condimentum, enim sed venenatis rutrum, ipsum neque consectetur orci, sed blandit justo nisi ac lacus.

Multiple Columns

Heading

1. Statement
2. Explanation
3. Example

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Tableau 2: Table caption

Theorem

Théorème (Mass–energy equivalence)

$E = mc^2$

Verbatim

Exemple (Theorem Slide Code)

```
\begin{frame}
\frametitle{Theorem}
\begin{theorem}[Mass--energy equivalence]
$E = mc^2$
\end{theorem}
\end{frame}
```

Figure

Uncomment the code on this slide to include your own image from the same directory as the template .TeX file.

Citation

An example of the `\parencite` command to cite within the presentation :

This statement requires citation (Ganter et Wille 1999).

References I

Jean-Louis Guigues et Vincent Duquenne. « Familles minimales d'implications informatives résultant d'un tableau de données binaires ». In : *Mathématiques et Sciences Humaines* 95 (1986), p. 5–18.

Bernhard Ganter et Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag New York, Inc., 1999.

Bernhard Ganter et Sergei A. Obiedkov. « Implications in Triadic Formal Contexts ». In : *ICCS*. 2004, p. 186–195.

Geneviève Roberge. « Visualisation des résultats d'une fouille de données dans les treillis de concepts ». Mém.de mast. Université du Québec en Outaouais, 2007.

References II

Rokia Missaoui, Lhouari Nourine et Yoan Renaud. « Computing Implications with Negation from a Formal Context ». In : *Fundam. Inf.* 115.4 (déc. 2012), p. 357–375.

Jean-François Viaud et al. « Décomposition sous-directe d'un treillis en facteurs irréductibles ». In : *Journées francophones d'Ingénierie des Connaissances IC 2015*. collection AFIA. Rennes, France, juin 2015.

Questions ?