

Treillis de Galois, Analyse Formelle de Concepts

Applications au génie logiciel

27 janvier 2010

Sommaire

Fondements

Données complexes

Analyse formelle de concepts (AFC)

- Méthode d'analyse de données
- Application de la théorie des treillis
- Extraction de concepts *Unité de base de la pensée humaine*
- Contextes restreints

Les planètes du système solaire

Objets étudiés : planètes

Attributs : caractéristiques des planètes

	Taille			Dist soleil		Satellite	
	petite	moyenne	grande	proche	loin	oui	non
Mercure	x			x			x
Vénus	x			x			x
Terre	x			x		x	
Mars	x			x		x	
Jupiter			x		x	x	
Saturne			x		x	x	
Uranus		x			x	x	
Neptune		x			x	x	
Pluton	x				x	x	

Contexte binaire et ses applications caractéristiques

Contexte (O, A, R)

- deux ensembles finis O et A
- une relation binaire $R \subseteq O \times A$.

Définition (applications caractéristiques d'une relation binaire)

Attributs communs à un ensemble d'objets

$$f : \mathcal{P}(O) \rightarrow \mathcal{P}(A)$$

$$X \mapsto f(X) = \{y \in A \mid \forall x \in X, (x, y) \in R\} = X'$$

Objets partageant un ensemble d'attributs

$$g : \mathcal{P}(A) \rightarrow \mathcal{P}(O)$$

$$Y \mapsto g(Y) = \{x \in O \mid \forall y \in Y, (x, y) \in R\} = Y'$$

Contexte binaire et ses applications caractéristiques

$$f(\{Mercury, Terre\}) = \\ \{Taille : petite, Soleil : proche\}$$

$$g(\{Taille : petite, Soleil : proche\}) = \\ \{Mercury, Venus, Terre, Mars\}$$

Concept

Un concept formel C est un couple (E, I) such that $f(E) = I$ (ou de manière équivalente) $E = g(I)$

$E = \{ e \in O \mid \forall i \in I, (e, i) \in R \}$
est *l'extension* (objets couverts),

$I = \{ i \in A \mid \forall e \in E, (e, i) \in R \}$
est *l'intension* (caractéristiques partagées).

Concept

	Taille			Dist soleil		Satellite	
	petite	moyenne	grande	proche	loin	oui	non
Mercure	x			x			x
Vénus	x			x			x
Terre	x			x		x	
Mars	x			x		x	
Jupiter			x		x	x	
Saturne			x		x	x	
Uranus		x			x	x	
Neptune		x			x	x	
Pluton	x				x	x	

Exemple

Le concept (E, I) des *petites planètes proches du soleil*

$$E = \{ \text{Mercure, Venus, Terre, Mars} \},$$

$$I = \{ \text{Taille : petite, Soleil : proche} \}$$

Un concept est un ensemble maximal d'objets possédant un ensemble maximal d'attributs

Concept

	Taille			Dist soleil		Satellite	
	petite	moyenne	grande	proche	loin	oui	non
Mercure	x			x			x
Vénus	x			x			x
Terre	x			x		x	
Mars	x			x		x	
Jupiter			x		x	x	
Saturne			x		x	x	
Uranus		x			x	x	
Neptune		x			x	x	
Pluton	x				x	x	

Contre-exemple

$$(E = \{ \text{Mercure}, \text{Terre} \}, I = \{ \text{Taille : petite}, \text{Soleil : proche} \})$$

$$E \neq g(I) = \{ \text{Mercure}, \text{Venus}, \text{Terre}, \text{Mars} \}$$

Un concept est un ensemble maximal d'objets possédant un ensemble maximal d'attributs

Spécialisation entre concepts

L'ensemble de tous les concepts \mathcal{C} forme un treillis \mathcal{L} lorsqu'il est muni de l'ordre suivant :

$$(E_1, I_1) \leq_{\mathcal{L}} (E_2, I_2) \Leftrightarrow E_1 \subseteq E_2 \\ \text{(or de manière équivalente } I_2 \subseteq I_1 \text{).}$$

Le concept des *petites planètes proches* du soleil

$$E_1 = \{ \textit{Mercure}, \textit{Venus}, \textit{Terre}, \textit{Mars} \},$$

$$I_1 = \{ \textit{Taille : petite}, \textit{Soleil : proche} \}$$

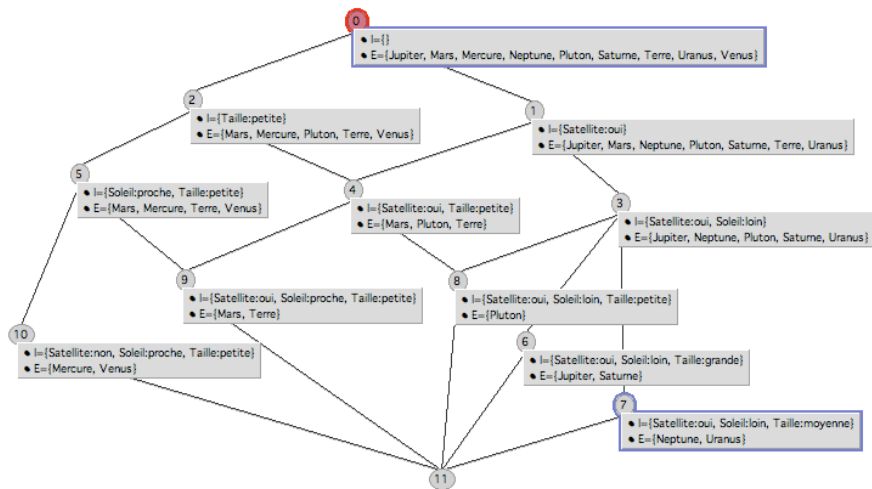
spécialise

Le concept des planètes *proches* du soleil

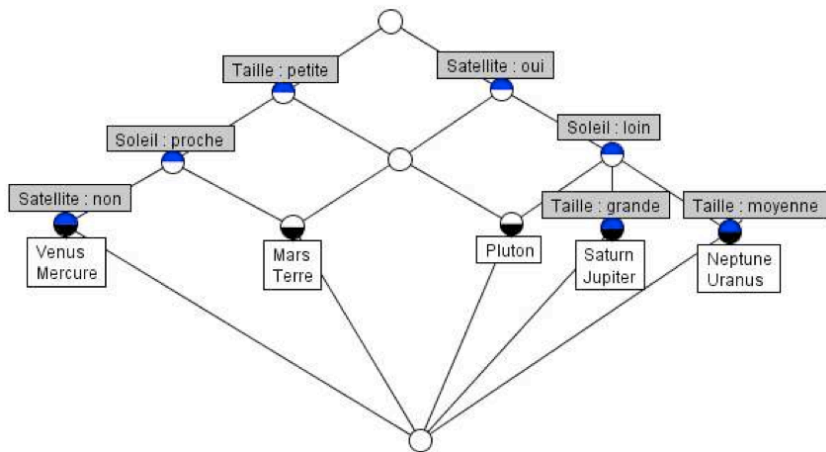
$$E_2 = \{ \textit{Mercure}, \textit{Venus}, \textit{Terre}, \textit{Mars}, \textit{Pluton} \},$$

$$I_2 = \{ \textit{Taille : petite} \}$$

Classification des planètes - Treillis



Classification des planètes - Treillis, étiquettes simplifiées



Conexp : <http://conexp.sourceforge.net/index.html>

Règles d'implication d'un contexte formel

A_1 et A_2 deux ensembles d'attributs

$$A_1 \Rightarrow A_2 \text{ ssi } A'_1 \subseteq A'_2$$

$$\text{(ou encore } f(A_1) \subseteq f(A_2)\text{)}$$

une planète sans satellite est de petite taille et proche du soleil

$$\{\textit{Satellite : non}\} \Rightarrow \{\textit{Soleil : proche, Taille : petite}\}$$

$$f(\{\textit{Satellite : non}\}) = \{\textit{Mercure, Venus}\}$$

$$\subseteq$$

$$f(\{\textit{Soleil : proche, Taille : petite}\}) =$$

$$\{\textit{Mars, Mercure, Terre, Venus}\}$$

Règles d'implication d'un contexte formel

$A_1 \Rightarrow A_2$ se ramène à :

$$A_1 \Rightarrow a, \forall a \in A_2$$

une planète sans satellite est de petite taille et proche du soleil

$$\{ \textit{Satellite} : \textit{non} \} \Rightarrow \{ \textit{Soleil} : \textit{proche}, \textit{Taille} : \textit{petite} \}$$

se ramène à

$$\{ \textit{Satellite} : \textit{non} \} \Rightarrow \{ \textit{Taille} : \textit{petite} \}$$

et

$$\{ \textit{Satellite} : \textit{non} \} \Rightarrow \{ \textit{Soleil} : \textit{proche} \}$$

Règles d'implication d'un contexte formel

Certains ensembles de règles sont redondants

$$\{\textit{Satellite} : \textit{non}\} \Rightarrow \{\textit{Soleil} : \textit{proche}, \textit{Taille} : \textit{petite}\}$$

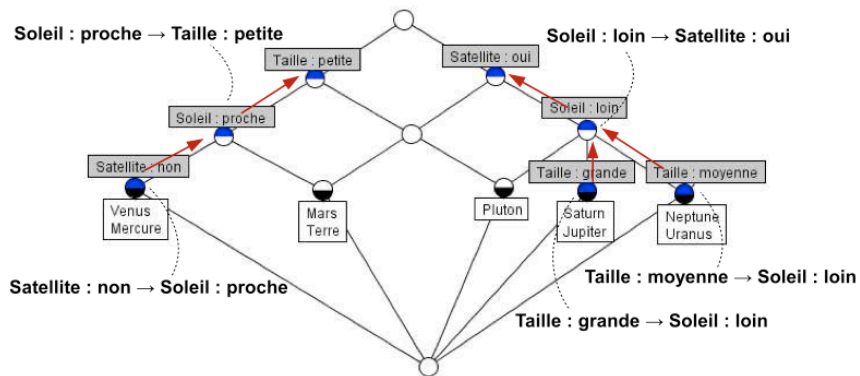
$$\{\textit{Satellite} : \textit{non}\} \Rightarrow \{\textit{Taille} : \textit{petite}\}$$

$$\{\textit{Satellite} : \textit{non}\} \Rightarrow \{\textit{Soleil} : \textit{proche}\}$$

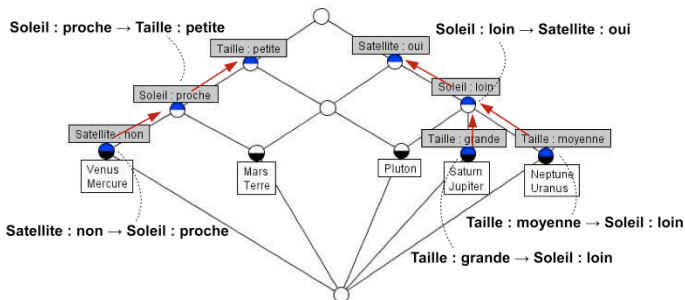
$$\{\textit{Soleil} : \textit{proche}\} \Rightarrow \{\textit{Taille} : \textit{petite}\}$$

On peut déduire de la réduction transitive du treillis l'ensemble minimal non redondant des implications du contexte qui ont un support non nul (il existe des objets vérifiant l'implication).

Regles d'implications sur les planètes



Regles d'implications sur les planètes



$$\begin{aligned} \{ \text{Satellite : non} \} &\Rightarrow \{ \text{Soleil : proche} \} \\ \{ \text{Soleil : proche} \} &\Rightarrow \{ \text{Taille : petite} \} \\ \{ \text{Taille : grande} \} &\Rightarrow \{ \text{Soleil : loin} \} \\ \{ \text{Taille : moyenne} \} &\Rightarrow \{ \text{Soleil : loin} \} \\ \{ \text{Soleil : loin} \} &\Rightarrow \{ \text{Satellite : oui} \} \end{aligned}$$

Sommaire

Fondements

Données complexes

Modélisation de données complexes

Sans prétendre à l'exhaustivité :

- Attributs valués (entier, réels, termes, structures, objets symboliques, etc.) (Ganter et Wille, Polaillon, ...)
- Description floue (Yahia et al., Belohlavek, ...)
- Hiérarchisation des valeurs (Godin et al., Carpineto et Romano, ...)
- Relations entre objets (Priss, Huchard-Valtchev et al., ...)
- Description logique (Chaudron et al., Ferré et al., ...)
- Graphes (Liquière, Prediger et Wille, ...)
-

Contextes multi-valués

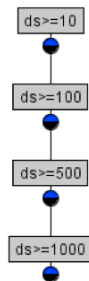
Definition

Un contexte multi-valué est un quadruplet $K = (O, A, V, J)$ où O et A sont les ensembles d'objets et d'attributs respectivement et $J \subseteq O \times A \times V$ représente la valuation des attributs. $(o, a, v) \in J$ signifie que l'objet o a la valeur v pour l'attribut a .

	Taille	Dist soleil	Satellite
	(km)	(10^6 km)	(nombre)
Mercure	4 878	58	0
Vénus	12 400	108	0
Terre	12 756	150	1
Mars	6 800	228	2
Jupiter	142 800	778	16
Saturne	120 800	1 427	19
Uranus	47 600	2 870	5
Neptune	44 600	4 500	8
Pluton	2 320	9 950	1

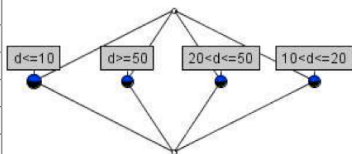
Echelle conceptuelle de l'attribut distance au soleil

	$ds \geq 10$	$ds \geq 100$	$ds \geq 500$	$ds \geq 1000$
58	x			
108	x	x		
150	x	x		
228	x	x		
778	x	x	x	
1 427	x	x	x	x
2 870	x	x	x	x
4 500	x	x	x	x
9 950	x	x	x	x



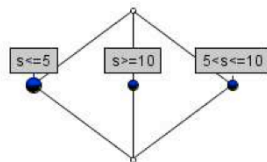
Echelle conceptuelle de l'attribut diamètre

	$d \leq 10$	$10 < d \leq 20$	$20 < d \leq 50$	$d \geq 50$
4 878	x			
12 400		x		
12 756		x		
6 800	x			
142 800				x
120 800				x
47 600			x	
44 600			x	
2.320	x			



Echelle conceptuelle de l'attribut Satellite

	$s \leq 5$	$5 < s \leq 10$	$s \geq 10$
0	x		
0	x		
1	x		
2	x		
16			x
19			x
5	x		
8		x	
1	x		



Contexte après échantillonnage

	Diamètre (km)				Distance au Soleil (10^6 km)				Satellite		
	$d \leq 10$	$10 < d \leq 20$	$20 < d \leq 50$	$d \geq 50$	$ds \geq 10$	$ds \geq 100$	$ds \geq 500$	$ds \geq 1000$	$s \leq 5$	$5 < s \leq 10$	$s \geq 10$
Mercure	x				x				x		
Venus		x			x	x			x		
Terre		x			x	x			x		
Mars	x				x	x			x		
Jupiter				x	x	x	x				x
Saturne				x	x	x	x	x			x
Uranus			x		x	x	x	x	x		
Neptune			x		x	x	x	x		x	
Pluton	x				x	x	x	x	x		

Treillis après échantillonnage

