

Représentation condensée de règles d'association multidimensionnelles

Alexandre Bazin, Aurélie Bertaux, Christophe Nicolle

LORIA, CIAD

03/09/19

Données symboliques

Données = relations entre choses de même nature

Deux types de choses : “Clients” et “Produits”

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Poisson</i>	<i>Pâtes</i>
<i>Bob</i>	×	×			
<i>Frank</i>		×	×	×	
<i>Alice</i>		×		×	×
<i>Claire</i>			×		×
<i>Michel</i>				×	×

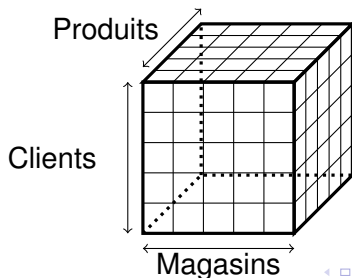
= “*Bob* a acheté des *Patates*”, “*Bob* a acheté de la *Viande*”, “*Frank* a acheté du *Riz*”, ...

Données symboliques multidimensionnelles

Trois types de choses : “Clients”, “Produits” and “Magasins”

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

= “*Bob* a acheté du *Riz* chez *Monoprix*”, “*Alice* a acheté de la *Viande* chez *Carrefour*”, “*Alice* a acheté de la *Viande* chez *Leclerc*”, ...



Règles d'association

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Poisson</i>	<i>Pâtes</i>
<i>Bob</i>	×	×			
<i>Frank</i>		×	×	×	
<i>Alice</i>		×		×	×
<i>Claire</i>			×		×
<i>Michel</i>				×	×

- Tous les clients qui ont acheté de la *Viande* et du *Riz* ont aussi acheté du *Poisson* !
- Deux tiers des clients qui ont acheté des *Pâtes* ont aussi acheté du *Poisson* !

Ou :

- $\{Viande, Riz\} \xrightarrow{1} \{Poisson\}$
- $\{Pâtes\} \xrightarrow{2/3} \{Poisson\}$

Règles dans le cas multidimensionnel

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

- La moitié des clients qui ont acheté du *Riz* chez *Monoprix* y ont aussi acheté des *Patates*
- Aucun de ceux qui ont acheté du *Riz* chez *Monoprix* n'en ont aussi acheté chez *Leclerc*
- Tous les clients qui ont acheté de la *Viande* chez *Leclerc* l'ont aussi fait chez *Carrefour*

Ou :

- $\{Riz\} \times \{Monoprix\} \xrightarrow{0.5} \{Patates\}$
- $\{Riz\} \times \{Monoprix\} \xrightarrow{0} \{Leclerc\}$
- $\{Viande\} \times \{Leclerc\} \xrightarrow{1} \{Carrefour\}$

Règles dans le cas multidimensionnel

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

- La moitié des clients qui ont acheté de la *Viande* quelque part l'ont fait chez *Monoprix*
- Un quart des achats de *Viande* étaient accompagnés de *Patates*

Ou :

- $\{Viande\} \xrightarrow{0.5} \{Monoprix\}$
- $\{Viande\} \xrightarrow{0.25} \{Patates\}$

Support et confiance

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Poisson</i>	<i>Pâtes</i>
<i>Bob</i>	×	×			
<i>Frank</i>		×	×	×	
<i>Alice</i>		×		×	×
<i>Claire</i>			×		×
<i>Michel</i>				×	×

$Support(\{Pâtes\}) =$ Nombre de clients qui ont acheté des *Pâtes*

$Support(\{Pâtes, Poisson\}) =$ Nombre de clients qui ont acheté à la fois des *Pâtes* et du *Poisson*

$$\begin{aligned} &Confiance(\{Pâtes\} \rightarrow \{Poisson\}) \\ = &\frac{Support(\{Pâtes, Poisson\})}{Support(\{Pâtes\})} = \frac{|\{Alice, Michel\}|}{|\{Alice, Claire, Michel\}|} = \frac{2}{3} \end{aligned}$$

Support dans le cas multidimensionnel

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

$$\{Riz\} \times \{Monoprix\} \rightarrow \{Patates\}$$

Bob and Alice ont tous deux acheté du *Riz* chez *Monoprix* donc
 $Support(\{Riz\} \times \{Monoprix\}) = 2$

De la même façon,

$$Support(\{Viande\} \times \{Leclerc\}) = 1$$

$$Support(\{Patates, Riz\} \times \{Monoprix\}) = 1$$

$$Support(\{Riz\} \times \{Leclerc, Carrefour\}) = 1$$

Support dans le cas multidimensionnel

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

$\{Viande\} \rightarrow \{Patates\} ?$

Frank et Alice ont acheté de la *Viande* quelque part donc

$Support_{Clients}(\{Viande\}) = 2$

De la *Viande* a été achetée par quelqu'un chez *Monoprix*, *Carrefour* et *Leclerc* donc

$Support_{Magasins}(\{Viande\}) = 3$

Frank a acheté de la *Viande* chez *Monoprix* et *Carrefour* et Alice a acheté de la *Viande* chez *Carrefour* et *Leclerc* donc

$Support_{Clients, Magasins}(\{Viande\}) = 4$

Confiance dans le cas multidimensionnel

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

$$\text{Confiance}(\{Riz\} \times \{Monoprix\} \rightarrow \{Patates\}) =$$

$$\frac{\text{Support}_{Clients}(\{Riz, Patates\} \times \{Monoprix\})}{\text{Support}_{Clients}(\{Riz\} \times \{Monoprix\})} = 0.5$$

$$\text{Confiance}(\{Viande\} \rightarrow \{Patates\}) =$$

$$\frac{\text{Support}_{Clients, Magasins}(\{Viande, Patates\})}{\text{Support}_{Clients, Magasins}(\{Patates\})} = 0.25$$

KNT Nguyen, L Cerf, M Plantevit, JF Boulicaut. Multidimensional association rules in boolean tensors. *Proceedings of the 2011 SIAM International Conference on Data Mining*.

Réduction du nombre de règles

Avec perte d'information

- Ne permet pas de retrouver la confiance de toutes les règles
- Mesures d'intérêt (support, confiance...)



Ne calculer/garder que les règles "intéressantes"

Sans perte d'information

- Permet de retrouver la confiance de toutes les règles
- En 2D, règles entre concepts



Calcul du diagramme de couverture du treillis des concepts

Concepts

	Patates	Viande	Riz	Poisson	Pâtes
Bob	×	×			
Frank		×	×	×	
Alice		×		×	×
Claire			×		×
Michel				×	×

Concepts :

$(\{Bob\}, \{Patates, Viande\}),$

$(\{Frank, Alice\}, \{Viande, Poisson\}),$

$(\{Alice, Claire, Michel\}, \{Pâtes\}), \dots$

Concepts

	Patates	Viande	Riz	Poisson	Pâtes
Bob	×	×			
Frank		×	×	×	
Alice		×		×	×
Claire			×		×
Michel				×	×

Concepts :

$(\{Bob\}, \{Patates, Viande\}),$

$(\{Frank, Alice\}, \{Viande, Poisson\}),$

$(\{Alice, Claire, Michel\}, \{Pâtes\}), \dots$

Concepts

	Patates	Viande	Riz	Poisson	Pâtes
Bob	×	×			
Frank		×	×	×	
Alice		×		×	×
Claire			×		×
Michel				×	×

Concepts :

$(\{Bob\}, \{Patates, Viande\})$,

$(\{Frank, Alice\}, \{Viande, Poisson\})$,

$(\{Alice, Claire, Michel\}, \{Pâtes\})$, ...

Concepts

	Patates	Viande	Riz	Poisson	Pâtes
Bob	×	×			
Frank		×	×	×	
Alice		×		×	×
Claire			×		×
Michel				×	×

Concepts :

$(\{Bob\}, \{Patates, Viande\})$,

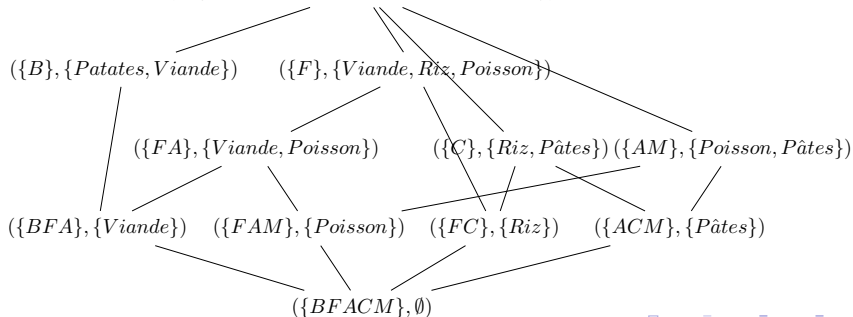
$(\{Frank, Alice\}, \{Viande, Poisson\})$,

$(\{Alice, Claire, Michel\}, \{Pâtes\})$, ...

Règles entre concepts

	Patates	Viande	Riz	Poisson	Pâtes
Bob	×	×			
Frank		×	×	×	
Alice		×		×	×
Claire			×		×
Michel				×	×

$(\emptyset, \{Patates, Viande, Riz, Poisson, Pâtes\})$



Règles entre concepts

$(\{F\}, \{Viande, Riz, Poisson\})$



$(\{FA\}, \{Viande, Poisson\})$

$Viande, Poisson \rightarrow Riz$

confiance = $\frac{1}{2}$

Concepts

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

Concepts:

$(\{Frank\}, \{Viande\}, \{Monoprix, Carrefour\}),$

$(\{Frank, Alice\}, \{Viande\}, \{Carrefour\}), \dots$

Concepts

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

Concepts:

$(\{Frank\}, \{Viande\}, \{Monoprix, Carrefour\}),$

$(\{Frank, Alice\}, \{Viande\}, \{Carrefour\}), \dots$

Concepts

	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		

Concepts:

$(\{Frank\}, \{Viande\}, \{Monoprix, Carrefour\}),$

$(\{Frank, Alice\}, \{Viande\}, \{Carrefour\}), \dots$

Transformation des données

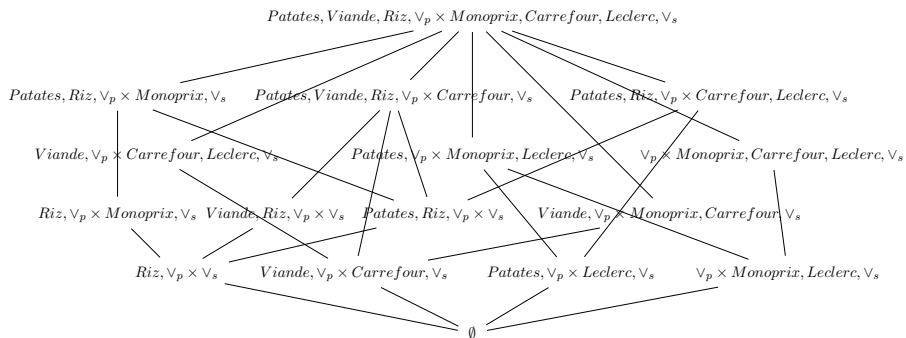
	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>
<i>Bob</i>	×		×				×		
<i>Frank</i>		×		×	×	×	×		×
<i>Alice</i>			×		×			×	
	<i>Monoprix</i>			<i>Carrefour</i>			<i>Leclerc</i>		



	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	\vee_p	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	\vee_p	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	\vee_p	<i>Patates</i>	<i>Viande</i>	<i>Riz</i>	\vee_p
<i>Bob</i>	×		×	×					×			×	×		×	×
<i>Frank</i>		×		×	×	×	×	×	×		×	×	×	×	×	×
<i>Alice</i>			×	×		×		×		×		×		×	×	×
	<i>Monoprix</i>				<i>Carrefour</i>				<i>Leclerc</i>				\vee_m			

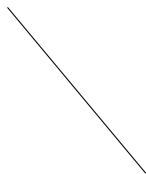
Règles entre concepts

	Patates	Viande	Riz	\forall_p	Patates	Viande	Riz	\forall_p	Patates	Viande	Riz	\forall_p	Patates	Viande	Riz	\forall_p	
Bob	x		x	x					x				x				x
Frank		x		x	x	x	x	x	x		x	x	x	x	x	x	x
Alice			x	x		x		x		x				x	x		x
	<i>Monoprix</i>				<i>Carrefour</i>				<i>Leclerc</i>				\forall_m				



Règles entre concepts

$(\{B\}, \{Patates, \forall_p\}, \{Monoprix, Leclerc, \forall_s\})$



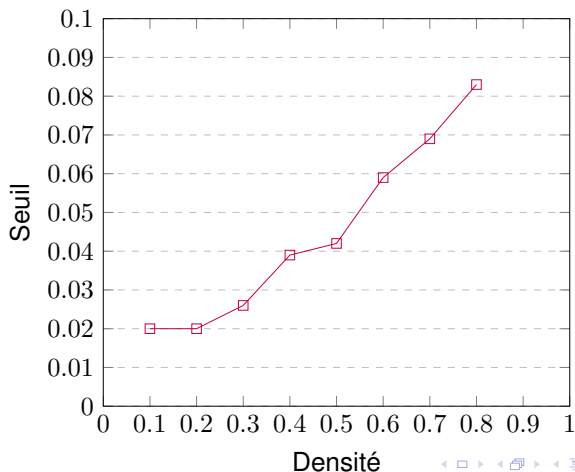
$(\{F, B\}, \{Patates, \forall_p\}, \{Leclerc, \forall_s\})$

$\{Patates, \forall_p\} \times \{Leclerc, \forall_s\} \rightarrow \{Monoprix\}$

confiance = $\frac{1}{2}$

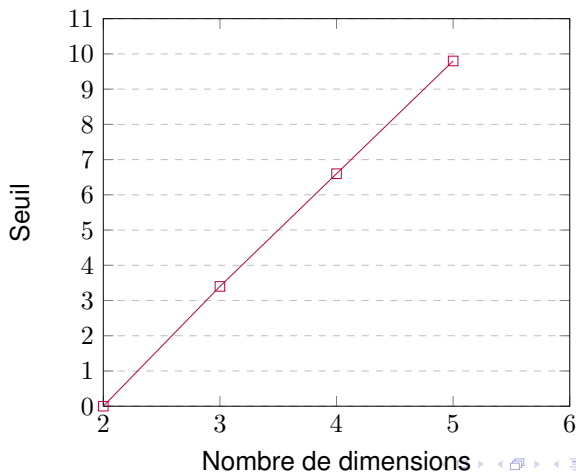
Concepts vs. Fréquents (Densité)

Seuil de fréquence pour lequel #concepts = #fréquents : 500 datasets aléatoires $50 \times 10 \times 4$ / point



Concepts vs. Fréquents (Dimensionnalité)

Seuil de fréquence pour lequel $\# \text{concepts} = \# \text{fréquents}$: 500 datasets aléatoires 30×5^n / point



Merci !