

lina LABORATOIRE D'INFORMATIQUE DE NANTES ATLANTIQUE
Équipe COD / KOD Team



Vers la fouille de règles d'association guidée par des ontologies et des schémas de règles

Claudia Marinica
Fabrice Guillet et Henri Briand

Atelier "Qualité des Données et des Connaissances"
Extraction et Gestion des Connaissances 2008

29/01/2008

lina LABORATOIRE D'INFORMATIQUE DE NANTES ATLANTIQUE
Équipe COD / KOD Team

Sommaire




- 1 Contexte et problématique
- 2 Etat de l'art
- 3 Approche proposée
- 4 Conclusions et perspectives

29/01/2008 2 Claudia Marinica

lina LABORATOIRE D'INFORMATIQUE DE NANTES ATLANTIQUE
Équipe COD / KOD Team

Contexte et problématique




❖ A la confluence de deux domaines :

- Extraction de Connaissances dans les Données (Datamining) - Règles d'association
- Gestion et Ingénierie de Connaissances, et leur application dans le Web Sémantique

29/01/2008 3 Claudia Marinica

lina LABORATOIRE D'INFORMATIQUE DE NANTES ATLANTIQUE
Équipe COD / KOD Team

Contexte et problématique



❖ Extraction de Règles d'Association

- extraction non supervisée de tendances implicatives dans les données : **X -> Y**
- But : extraction des connaissances intéressantes et utiles pour l'utilisateur (analyste ou décideur) ;

❖ La Gestion et l'Ingénierie de Connaissances (Knowledge Management)

- permet la manipulation et la diffusion de connaissances dans une communauté ;

29/01/2008 4 Claudia Marinica

❖ Problématique :

- Grand nombre de règles d'associations :
 - Triviales, inutilisables
 - inintéressants pour le décideur

=> besoin d'aide en phase de post-traitement

❖ Objectif :

- Intégration des connaissances du décideur

❖ Approche :

- Ontologies + Schémas de règles
=> **superviser la découverte de règles d'association**

- 1 Contexte et problématique
- 2 **État de l'art**
- 3 Approche proposée
- 4 Conclusion et perspectives

❖ Règles d'association

- découverte non supervisée de tendances implicatives dans les données.
- règle d'association** [Agrawal et al., 1993] - une implication :

$X \rightarrow Y$, où $X, Y \subset I$ et $X \cap Y = \emptyset$

+ support, confiance

Exemple : bread -> milk

ensemble d' items : $I = \{I_1, I_2, \dots, I_p\}$
 transaction : $t = \{I_1, I_2, \dots, I_m\}$, donc $t \subseteq I$
 base de transactions : $T = \{t_1, t_2, \dots, t_n\}$

❖ Mesures d'intérêt :

- Indicateurs numériques sur la qualité des règles d'association

❖ Classification

- Mesures objectives** (Piatetsky-Shapiro, 1991, Hilderman and Hamilton, 2001, Guillet and Hamilton, 2007)
 - indices statistiques qui évaluent la contingence d'une règle dans les données
 - limitation : beaucoup trop de règles, la plus part qui n'intéressent pas le décideur
- Mesures subjectives**
 - dépendent des buts, connaissances, croyances du décideur
 - combinées à des algorithmes supervisés spécifiques afin de comparer les règles extraites avec ce que le décideur connaît

Etat de l'art

Mesures subjectives

- ❖ Système KEFIR (Piatetsky-Shapiro et al, 1994)
 - Déviations – estimation du profit et prise des actions correctives
- ❖ Rule Templates (Klemettinen et al., 1994)
 - Modèle basé sur les *templates (patrons)*
- ❖ Système de croyances (Silberschatz et Tuzhilin, 1995)
 - *Hard Beliefs* et *Soft Beliefs* ;
- ❖ Technique floue de « matching » (Liu et Hsu, 1996)
 - Connaissances sous forme de règles floues
- ❖ Contradiction logique (Padmanabhan et Tuzhilin., 1997)
- ❖ Classification des connaissances de l'utilisateur (Liu et al., 1999)
 - *General Impression (GI)*,
 - *Reasonably Precise Concept (RPC)*, *Precise Knowledge (PK)*
- ❖ **Règles d'association généralisées** (Srikant & Agrawal, 1995)
 - Les items hiérarchisés à travers une taxonomies

29/01/2008

9

Claudia Marinica

Etat de l'art

Ontologies et règles d'association

- ❖ **Ontologie** - *une spécification explicite d'une conceptualisation. Une conceptualisation est un résumé, une vision simplifiée du monde que nous voulions représenter avec certaines objectifs* (Gruber T.R., 1993).
 - ensemble de concepts $C = \{ C_1, C_2, \dots, C_o \}$
 - ensemble de relations/propriétés $R = \{ R_1, R_2, \dots, R_r \}$
 - relation de subsomption, \leq :
 - $C_1 \leq C_2$ - si le concept C_1 subsume le concept C_2 .
 - Réaliser une structuration de type hiérarchique des items en tenant compte des connaissances du décideur

29/01/2008

10

Claudia Marinica

Sommaire

- 1 Contexte et problématique
- 2 État de l'art
- 3 **Approche proposée**
- 4 Conclusions et perspectives

29/01/2008

11

Claudia Marinica

Approche proposée

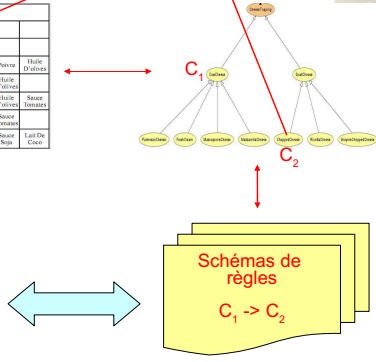
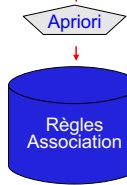
- ❖ Problématique - règles d'association extraites:
 - nombre trop élevé
 - inintéressantes pour le décideur
- ❖ Points de départ
 - Règles d'association généralisées -> ontologie
 - Impressions générales -> schémas de règles
- ❖ Proposition - amélioration de la phase de post-traitement :
 - Représentation des connaissances sous deux formes : ontologique et implicite

29/01/2008

12

Claudia Marinica

N°	Transaction										
1	Echalotes	Maquereau Fume	Moutarde A L'ancienne	Cumin	Gruyère Râpé	Huile D'olives					
2	Oignons	Salami	Parmesan	Huile D'olives	Sauce Tomates						
3	Tomates	Oignons	Coeur D'artichauts	Noix Muscades	Sucres-En Poudre	Sel	Poivre	Huile D'olives			
4	Poires	Lardons	Sel	Poivre	Romarin	Parmesan	Huile D'olives				
5	Olives Noires	Poivron Vert	Chorizo	Œufs	Origan	Mozzarella	Huile D'olives	Sauce Tomates			
6	Olives Noires	Ananas	Jambon Blanc	Origan	Fromage Râpé	Huile D'olives	Sauce Tomates				
7	Poivron Rouge	Coriandre Fraiche	Pousses D'épinard	Crevettes Roses	Sucres-En Poudre	Gruyère Râpé	Sauce Soja	Lait De Coco			



29/01/2008

13

Claudia Marinica

Illustration - Données -

❖ L'ensemble de données porte sur le domaine des types de pizzas

❖ Caractéristiques :

- 99 attributs
- 35 transactions
- Chaque item correspond à un ingrédient de pizza
- Chaque transaction – sous-ensemble d'items => correspond à un type de pizza

29/01/2008

14

Claudia Marinica

Illustration - Données -

N°	Transaction									
1	Echalotes	Maquereau Fume	Moutarde A L'ancienne	Cumin	Gruyère Râpé	Huile D'olives				
2	Oignons	Salami	Parmesan	Huile D'olives	Sauce Tomates					
3	Tomates	Oignons	Coeur D'artichauts	Noix Muscades	Sucres-En Poudre	Sel	Poivre	Huile D'olives		
4	Poires	Lardons	Sel	Poivre	Romarin	Parmesan	Huile D'olives			
5	Olives Noires	Poivron Vert	Chorizo	Œufs	Origan	Mozzarella	Huile D'olives	Sauce Tomates		
6	Olives Noires	Ananas	Jambon Blanc	Origan	Fromage Râpé	Huile D'olives	Sauce Tomates			
7	Poivron Rouge	Coriandre Fraiche	Pousses D'épinard	Crevettes Roses	Sucres-En Poudre	Gruyère Râpé	Sauce Soja	Lait De Coco		

22/11/2007

15

Claudia Marinica

Illustration - Règles d'association -

- Extraction de règles d'association avec l'outil ARMiner :

-8% support
-70% confiance
=> 253 règles

Règles d'association	Support	Confiance
Tomates, Sel -> Poivre	0.085	1.0
JambonBlanc -> HuileDolive	0.085	1.0
Ouefs, HuileDolive -> Sel, Poivre	0.0859	0.75
Ouefs, Poivre -> Sel, HuileDolive	0.085	0.75
Sel, Poivre, HuileDolive, SauceTomate -> Ail	0.085	1.0
Sel, Mozzarella -> Poivre, Ail	0.085	0.75
OlivesNoires, FromageRapé, HuileDolive, SauceTomate -> Origan	0.085	1.0
OlivesNoires, Origan, HuileDolive, SauceTomate -> FromageRapé	0.085	0.75
Mozarella -> HuileDolive	0.171	0.85
Mozarella, SauceTomate -> HuileDolive	0.114	1.0
Origan, HuileDolive, SauceTomate -> OlivesNoires, FromageRapé	0.085	0.75
Origan, FromageRapé, SauceTomate -> OlivesNoires, HuileDolive	0.085	1.0
Origan, FromageRapé, HuileDolive -> Origan, SauceTomate	0.085	0.75
OlivesNoires, FromageRapé, SauceTomate -> Origan, HuileDolive	0.085	1.0
OlivesNoires, FromageRapé, HuileDolive -> Origan, SauceTomate	0.085	1.0
OlivesNoires, Origan, SauceTomate -> FromageRapé, HuileDolive	0.085	0.75
OlivesNoires, FromageRapé, HuileDolive -> Origan	0.085	1.0
OlivesNoires, Origan, FromageRapé -> HuileDolive	0.085	1.0
Origan, FromageRapé -> OlivesNoires, HuileDolive	0.085	0.75
OlivesNoires, FromageRapé -> Origan, HuileDolive	0.085	1.0

29/01/2008

Claudia Marinica

Illustration
- Ontologie -

Approche proposée

- ❖ Utilisation de l'ontologie pizza.owl (Rector et al. 2004) proposée dans les exemples de l'outil Protégé
- ❖ Trois concepts – racine :
 - « Pizza » ;
 - « PizzaTopping » ;
 - « Spiciness »
- ❖ Deux propriétés :
 - « hasTopping »
 - « hasSpiciness »



29/01/2008

17

Claudia Marinica

Illustration
- Ontologie -

Approche proposée

- ❖ Connexion aux données :
 - Chaque concept est instancié dans les données
 - Deux possibilités de connexion :
 - Un concept est associé à un item
« Black olives » -> « Olives noires »
« FishTopping » -> « Somon frais » OU « Anchois salés » OU « Anchois » OU « Poulpes » OU « Moules » OU « Coques » OU « Crevettes roses » OU « Maquereau fumé »
 - Un concept est construit à l'aide d'une condition logique (nécessaire et suffisante) portant sur un sous-ensemble des concepts de l'ontologie

$$\text{VegetarianPizza} = \overline{\text{FishTopping}} \text{ET} \overline{\text{MeatTopping}}$$

29/01/2008

18

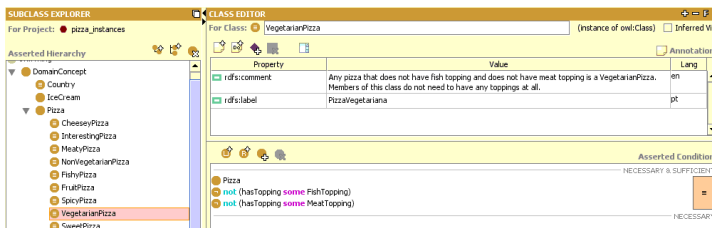
Claudia Marinica

Illustration
- Ontologie -

Approche proposée

- ❖ Connexion concept – condition logique (nécessaire et suffisante) portant sur un sous-ensemble des concepts de l'ontologie

$$C_d = f(C_p, \dots, C_k)$$
 Exemple :
 « VegetarianPizza » = f (FishTopping, MeatTopping) =
 = not (hasTopping some FishTopping) AND
 not (hasTopping some MeatTopping)



29/01/2008

Claudia Marinica

Illustration
- Schémas de règles -

Approche proposée

- Représentation des connaissances du décideur
- $X_1, X_2, \dots, X_{p1} \rightarrow Y_1, Y_2, \dots, Y_{p2}$
 où chaque X_i, Y_i sont des concepts C_{di} , contraintes sur les concepts (C)
- Exemple :
 Soit le schéma de règles :
 $S : C_{d1} \rightarrow C_{d2}$ avec $C_{d1} = f_1(I_p, \dots, I_k)$ et $C_{d2} = f_2(I'_p, \dots, I'_k)$
 Soit $R : A \rightarrow B$ une règle d'association
 R est **conforme** avec S :
 – f_1 est vraie sur A
 – f_2 est vraie sur B

29/01/2008

20

Claudia Marinica

- Exemple :
 - S : « CheesePizza » -> « VegetarianPizza »
 - « CheesePizza = hasTopping some CheeseTopping »
 - Le schéma de règles sélectionne les règles :
 - ayant au moins un item de type fromage dans l'antécédent
 - n'ayant pas d'items de type viande ou poisson dans le conséquent
- R : « OlivesNoires, FromageRapé, HuileDolive -> Origan, SauceTomate »
- => R est conforme avec S

29/01/2008

21

Claudia Marinica

- Post-traitement de règles d'association avec le schéma de règles « CheesePizza -> VegetarianPizza »

=> 88 règles sont sélectionnées

Règles d'association avec le schéma de règles ChessePizza -> VegetarianPizza	Support	Confience
Sel, Mozarella -> Poivre, Ail	0.085	0.75
OlivesNoires, FromageRapé, HuileDolive, SauceTomate -> Origan	0.085	1.0
Mozarella -> HuileDolive	0.171	0.85
Mozarella, SauceTomate -> HuileDolive	0.114	1.0
Origan, FromageRapé, SauceTomate -> OlivesNoires, HuileDolive	0.085	1.0
Origan, FromageRapé, HuileDolive -> OlivesNoires, SauceTomate	0.085	0.75
OlivesNoires, FromageRapé, SauceTomate -> Origan, HuileDolive	0.085	1.0
OlivesNoires, FromageRapé, HuileDolive -> Origan, SauceTomate	0.085	1.0
OlivesNoires, FromageRapé, HuileDolive -> Origan	0.085	1.0
OlivesNoires, Origan, FromageRapé -> HuileDolive	0.085	1.0
Origan, FromageRapé -> OlivesNoires, HuileDolive	0.085	0.75
OlivesNoires, FromageRapé -> Origan, HuileDolive	0.085	1.0

29/01/2008

22

Claudia Marinica

- Contexte et problématique
- État de l'art
- Approche proposée
- Conclusions et perspectives**

29/01/2008

23

Claudia Marinica

- ❖ Proposer une nouvelle méthode d'extraction de règles d'association à l'aide de connaissances du décideur
 - Ontologies + Schémas de règles
 - => **superviser la découverte de règles d'association**
- ❖ Exemplifier notre proposition sur un exemple pédagogique
- ❖ Intégration dans l'algorithme de la fouille

29/01/2008

24

Claudia Marinica

