LIRMM
State of the Art
E. Hüllermeier
F. Berzal & al.
C. Fiot & al.
Proposition
Definitions
Frequency Computing
Join Operation
Experiments
Generating Rules
Conclusion

Fast Extraction of Gradual Association Rules: A Heuristic Based Method

Lisa Di Jorio, Anne Laurent and Maguelonne Teisseire

October 29, 2008

Outline

1. State of the Art
   - E. Hüllermeier
   - F. Berzal & al.
   - C. Fiot & al.
2. Proposition
   - Definitions
   - Frequency Computing
   - Join Operation
3. Experiments
4. Generating Rules
5. Conclusion

What is a Gradual Rule?

Gradualness

- Represents variation between elements
- "The more X is A, the more Y is B"
- Fuzzy domain
- No consensus definition

Example

- The more a truc weight is **medium**, the more the speed is **slow**
- The later we wake up, the higher **the certainty** to miss the train
- The higher the age, the higher the salary

Many applications, in various domains

Example

Example database

Objet	Age (A)	Salary (S)	Car (C)
o_1	22	1200	1
o_2	28	1850	1
o_3	24	1200	0
o_4	35	2200	1
o_5	38	2000	1
o_6	44	3400	1
o_7	52	3400	2
o_8	41	5000	2

Example: A^+S^+

Example database

Objet	Age (A)	Salary (S)
o_1	22	1200
o_3	24	1200
o_2	28	1850
o_4	35	2200
o_5	38	2000
o_6	44	3400
o_7	52	3400
o_8	41	5000

- We are working on **values** of items
- Comparisons are done **between objects**

State of the Art

- **Eyke Hüllermeier** : Association Rules for Expressing Gradual Dependencies. PKDD 2002
- **Berzal & al.** : An alternative approach to discover gradual dependencies. IJUFKS 2007
- **Fiot & al.**: Gradual trends in fuzzy sequential patterns. IPMU 2008

Eyke Hüllermeier's proposal

Two kind of rules

- **deviation rule:** $(A \rightarrow^d B)$ important gap from the conditional means
- **tendency rule:** $(A \rightarrow^t B)$ gradual dependency between two items A and B

Principle

1. Build the "*contingency diagram*"
2. Generate information about the diagram
3. Generate association rules

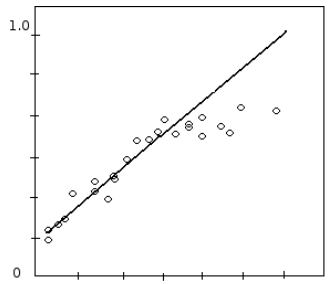
Contingency Diagram

Binary context...

	$B(y)=0$	$B(Y)=1$	
$A(x)=0$	n_{00}	n_{01}	$n_{0\bullet}$
$A(x)=1$	n_{10}	n_{11}	$n_{1\bullet}$
	$n_{\bullet 0}$	$n_{\bullet 1}$	n

- $freq(A \rightarrow B) = n_{11}/n$
- $conf(A \rightarrow B) = n_{11}/n_{1\bullet}$

...Fuzzy context



- Least squares method
- Quality measure R^2
- Deriving rule as $A \rightarrow^t B[\alpha, \beta]$

Limits

- Computation complexity: contingency diagram + R^2
- Looking for rule which are only increasing ($\delta > 0$), or only decreasing ($\delta < 0$)
- Cannot combine the two variations

Berzal & al. proposal

Gradual rule

- The more X is A, the more Y is B
- The less X is A, the more Y is B
- The more X is A, the less Y is B
- The less X is A, the less Y is B

Principle

- Adapting classical algorithms to the gradual extraction process
- Defining gradual items and consequently gradual itemsets
- Defining frequency (support) and confidence

Limits

- Memory complexity: maximum of 3-itemsets
- Let $X = \{X_1, X_2, ..., X_m\}$ be a set of attributes from $\mathcal{D}$, and for each item $X_i \in X$ the set of fuzzy restrictions n_i defined by fuzzy sets $A_{i1}, ..., A_{in_i}$. Then $|GI^{\mathcal{D}}| = 2 \times \sum_{i=0}^{m} n_i$

Fiot & al. proposal

Evolution Patterns

- A sequential pattern is a list of ordered itemsets.
- Measures variations from one time to another
- 2-step approach:
  - Transform the database into a "variation" database
  - Mine it by the means of generalised sequential pattens

Different problematic from ours.

Motivation

Gradualness

- "A variation of A **comes along** with a variation of B"
- Statistical approach: extract itemsets, but **no increasing and decreasing variation** combination
- Data mining approach: does not go beyond 3-itemsets

Our aim

- **Adapt** classical approaches
- No fuzzy theory: **directly working on values**
- Extract n-itemsets
- Be **scalable**

Proposition

Overview

Generate and Prune Principle

1. Generate gradual 1-itemsets
   - Frequency counting
   - Prune
2. Generate gradual n-itemsets
   - Joining on gradual $(n-1)$-itemsets
   - Frequency counting
   - Prune

We need to define:

- What gradual items and gradual itemsets are
- How to count the frequency of a gradual itemset
- How to build the join operation

Item

Definition: gradual item

A gradual item is denoted i^* with $* \in \{+, -\}$, and $dom(i)$ having a total order relation

- $* = +$ means "*value of i is increasing*"
- $* = -$ means "*value of i is decreasing*"

What is variation?

- $+$ corresponds to $\geq$
- $-$ corresponds to $\leq$
- As we are comparing objects, order is expressed as :
  - If $f(o_1, i) \geq f(o_2, i)$ then we write $i+$
  - If $f(o_1, i) \leq f(o_2, i)$ then we write $i-$

Gradual Itemset

Definition

A gradual itemset is a non-empty list of gradual items

Example database

Objet	Age (A)	Salary (S)
o_1	22	1200
o_3	24	1200
o_2	28	1850
o_5	38	2000
o_6	44	3400
o_7	52	3400

TAB: $(A^+ S^+)$

Frequency of a Gradual Itemset

Definition : frequency

Let $\mathcal{O}' \subseteq \mathcal{O}$ be the **maximal** set of objects ordered on i_1 and then on i_2 ... and then on i_n such as $\forall o_j, o_k \in \mathcal{O} such\ as\ j \neq k$

$$\begin{cases} f(o_j, i_1) *_1 f(o_k, i_1) \wedge ... \wedge f(o_j, i_n) *_n f(o_k, i_n) if\ r(o_j) > r(o_k) \\ f(o_j, i_1)\ c(*_1)\ f(o_k, i_1) \wedge ... \wedge f(o_j, i_n)\ c(*_n)\ f(o_k, i_n) if\ r(o_j) < r(o_k) \end{cases}$$

Then $Freq(i_1^{*1} i_2^{*2} ... i_n^{*n}) = \frac{|\mathcal{O}'|}{|\mathcal{O}|}$

Proposition

It could exist more than one $\mathcal{O}_n \subseteq \mathcal{O}$ respecting this definition

How to find the maximal set $\mathcal{O}$ for a gradual itemset ?

Example

Example database

Objet	Age (A)	Salary (S)
o_1	22	1200
o_3	24	1200
o_2	28	1850
o_4	35	2200
o_5	38	2000
o_6	44	3400
o_7	52	3400
o_8	41	5000

Frequency Computing

Conflicting set : gradual 2-itemset case

Let $(i_1^{*1} i_2^{*2})$ be a gradual 2-itemset and $\mathcal{O}$ the set of objects ordered on i_1^{*1} and then on i_1^{*2}.
The **conflicting set** $\mathcal{O}_{|i}$ associated to an object $o_i \in \mathcal{O}$ contains objects discarded if keeping o_i into the final set $\mathcal{O}'$

Why?

- If $\mathcal{O}_{|i} = \emptyset$, it respects the gradual itemset without any conflict
- Otherwise, $\forall o \in \mathcal{O}_{|i}$ will be deleted
- Delete objects o_i such that $|\mathcal{O}_{|i}|$ is maximal

Conflicting Set Building

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	
o_3	24	1200	
o_2	28	1850	
o_4	35	2200	
o_5	38	2000	
o_8	41	5000	
o_6	44	3400	
o_7	52	3400	

Conflicting Set Building

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$
o_3	24	1200	$\emptyset$
o_2	28	1850	$\emptyset$
o_4	35	2200	
o_5	38	2000	
o_8	41	5000	
o_6	44	3400	
o_7	52	3400	

Conflicting Set Building

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$
o_3	24	1200	$\emptyset$
o_2	28	1850	$\emptyset$
o_4	35	2200	$\{o_5\}$
o_5	38	2000	$\{o_4\}$
o_8	41	5000	
o_6	44	3400	
o_7	52	3400	

Conflicting Set Building

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$
o_3	24	1200	$\emptyset$
o_2	28	1850	$\emptyset$
o_4	35	2200	$\{o_5\}$
o_5	38	2000	$\{o_4\}$
o_8	41	5000	$\{o_6, o_7\}$
o_6	44	3400	$\{o_8\}$
o_7	52	3400	$\{o_8\}$

Discarding Conflicts

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$
o_3	24	1200	$\emptyset$
o_2	28	1850	$\emptyset$
o_4	35	2200	$\{o_5\}$
o_5	38	2000	$\{o_4\}$
~~o_8~~	~~41~~	~~5000~~	~~$\{o_6, o_7\}$~~
o_6	44	3400	~~$\{o_8\}$~~ $= \emptyset$
o_7	52	3400	~~$\{o_8\}$~~ $= \emptyset$

Discarding Conflicts

Example

Objet	A	S	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$
o_3	24	1200	$\emptyset$
o_2	28	1850	$\emptyset$
~~o_4~~	~~35~~	~~2200~~	~~$\{o_5\}$~~
o_5	38	2000	~~$\{o_4\}$~~
o_6	44	3400	$\emptyset$
o_7	52	3400	$\emptyset$

Example : Frequency

Example

Objet	A	S
o_1	22	1200
o_3	24	1200
o_2	28	1850
o_5	38	2000
o_6	44	3400
o_7	52	3400

Frequency computing

$$Freq(A^+S^+) = \frac{6}{8} = 0.75\ (75\%)$$

Frequency : Levels 1 and 2

Level 1

- For a 1-item, total order does not have any conflict
- $Freq(i^*) = 100\%$

Definition

Let $s_1 = (i_1^{*1} \ \ldots \ i_n^{*n})$ and $s_2 = (i_j^{*j} \ \ldots \ i_k^{*k}) = s_2$ be two gradual itemsets such as $s_1 \cap s_2 = \emptyset$. Then the gradual rule $s_1 \Rightarrow s_2$ has a frequency $Freq(s_1 \cup s_2)$ and a confidence :

$$Conf(s_1 \Rightarrow s_2) = \frac{Freq(s_1 \cup s_2)}{Freq(s_1)}$$

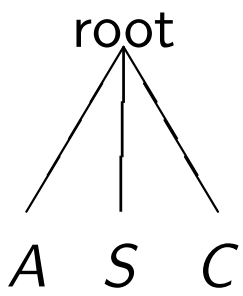
Level 2 confidence

We have $Conf(i_1^{*1} \Rightarrow i_2^{*2}) = Conf(i_2^{*2} \Rightarrow i_1^{*1}) = F(i_1^{*1} i_2^{*2})$

Generation Process

Problem

- Classically: levels k are generated from $(k-1)$ ones
- Objets respecting an itemset: intersection
- Levelwise algorithm by the mean of a prefix tree



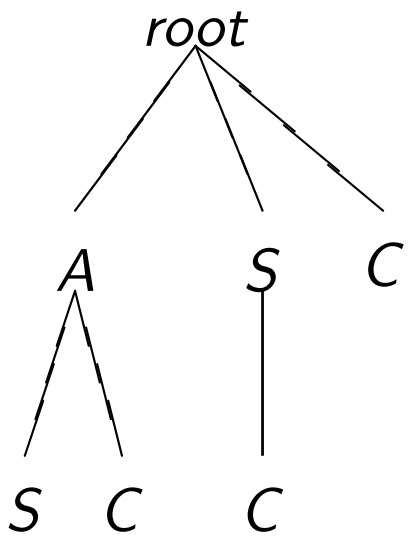
Example

- $A = \{o_1, o_2, o_3\}$
- $S = \{o_1, o_2\}$
- $C = \{o_2, o_3\}$

Generation Process

Problem

- Classically: levels k are generated from $(k-1)$ ones
- Objets respecting an itemset: intersection
- Levelwise algorithm by the mean of a prefix tree



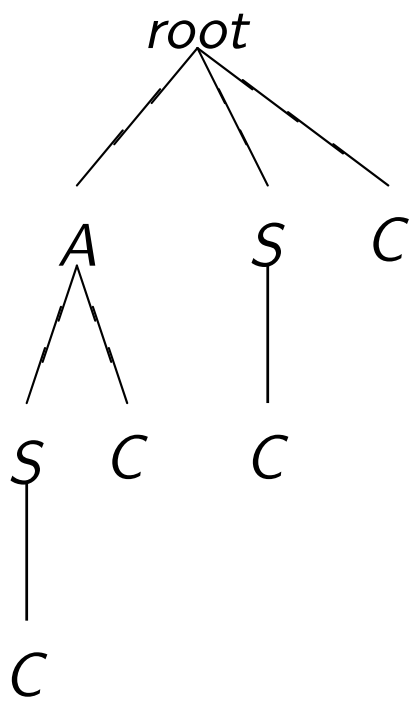
Example

- $A = \{o_1, o_2, o_3\}$
- $S = \{o_1, o_2\}$
- $C = \{o_2, o_3\}$
- $A \cap S = \{o_1, o_2\}$
- $A \cap C = \{o_2, o_3\}$
- $S \cap C = \{o_2\}$

Generation Process

Problem

- Classically: levels k are generated from $(k-1)$ ones
- Objets respecting an itemset: intersection
- Levelwise algorithm by the mean of a prefix tree



Example

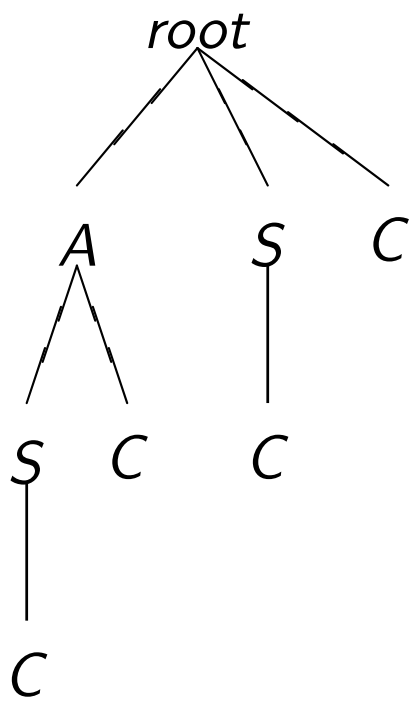
- $AS = \{o_1, o_2\}$
- $AC = \{o_2, o_3\}$
- $SC = \{o_2\}$
- $AS \cap AC = \{o_2\}$

Intersection impossible in our context: how to generate level k candidates?

Generation Process

Problem

- Classically: levels k are generated from $(k-1)$ ones
- Objets respecting an itemset: intersection
- Levelwise algorithm by the mean of a prefix tree



Example

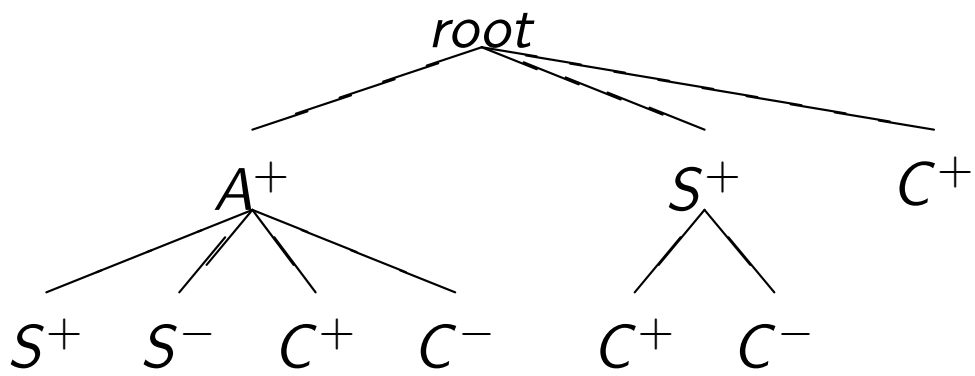
- $AS = \{o_1, o_2\}$
- $AC = \{o_2, o_3\}$
- $SC = \{o_2\}$
- $AS \cap AC = \{o_2\}$

Intersection impossible in our context: how to generate level k candidates?

Generation Process

Solution

- Generate i^+ and i^-
- Union of sets associated to gradual $k-1$-itemsets
- Adapt conflicting sets to two items



Property

- $c(\geq) = \leq$ and $c(\leq) = \geq$
- Let S be a sequence, then $Freq(c(S)) = Freq(S)$

Conflicting Sets Having Two Dimensions

Joining with conflicting sets

- $Join((i_1^{*1} \ldots i_{n-1}^{*n-1} i_n^{*n}), (i_1^{*1} \ldots i_{n-1}^{*n-1} i_k^{*k})) = (i_1^{*1} \ldots i_{n-1}^{*n-1} i_n^{*n} i_k^{*k})$
- Objects ordered during the generation

Conflicting set

- Keep two conflicting sets: one for i_n^{*n} and one for i_k^{*k}
- Discarding object having the more conflict number
- Conflicting sets are updated together

Example

Example

Objet	A	S	$\mathcal{O}_{\mid i}$	C	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$	1	$\{o_3 o_7\}$
o_3	24	1200	$\emptyset$	0	$\{o_1 o_7\}$
o_2	28	1850	$\emptyset$	1	$\{o_7\}$
o_4	35	2200	$\{o_5\}$	1	$\{o_7\}$
o_5	38	2000	$\{o_4\}$	1	$\{o_7\}$
o_6	44	3400	$\emptyset$	1	$\{o_7\}$
o_7	52	3400	$\emptyset$	2	$\{o_1 o_2 o_4 o_5 o_6\}$

Example

Example

Objet	A	S	$\mathcal{O}_{\mid i}$	C	$\mathcal{O}_{\mid i}$
o_1	22	1200	$\emptyset$	1	$\{o_3\}$
o_3	24	1200	$\emptyset$	0	$\{o_1\}$
o_2	28	1850	$\emptyset$	1	$\emptyset$
o_4	35	2200	$\{o_5\}$	1	$\emptyset$
o_5	38	2000	$\{o_4\}$	1	$\emptyset$
o_6	44	3400	$\emptyset$	1	$\emptyset$

Example

Example

Objet	A	S	$\mathcal{O}_{\mid i}$	C	$\mathcal{O}_{\mid i}$
~~o_1~~	~~22~~	~~1200~~	~~$\emptyset$~~	~~1~~	~~$\{o_3\}$~~
o_3	24	1200	$\emptyset$	0	~~$\{o_1\}$~~
o_2	28	1850	$\emptyset$	1	$\emptyset$
o_4	35	2200	$\{o_5\}$	1	$\emptyset$
o_5	38	2000	$\{o_4\}$	1	$\emptyset$
o_6	44	3400	$\emptyset$	1	$\emptyset$

Example

Example

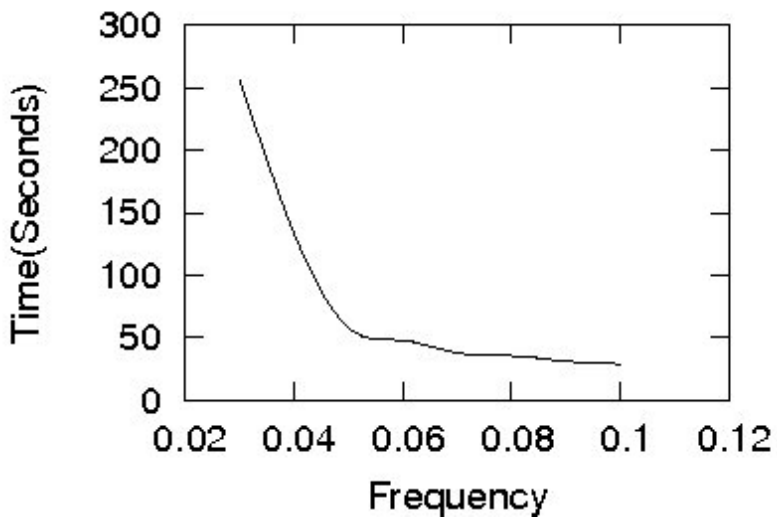
Objet	A	S	$\mathcal{O}_{\mid i}$	C	$\mathcal{O}_{\mid i}$
o_3	24	1200	$\emptyset$	0	$\emptyset$
o_2	28	1850	$\emptyset$	1	$\emptyset$
~~o_4~~	~~35~~	~~2200~~	~~$\{o_5\}$~~	~~1~~	~~$\emptyset$~~
o_5	38	2000	~~$\{o_4\}$~~	1	$\emptyset$
o_6	44	3400	$\emptyset$	1	$\emptyset$

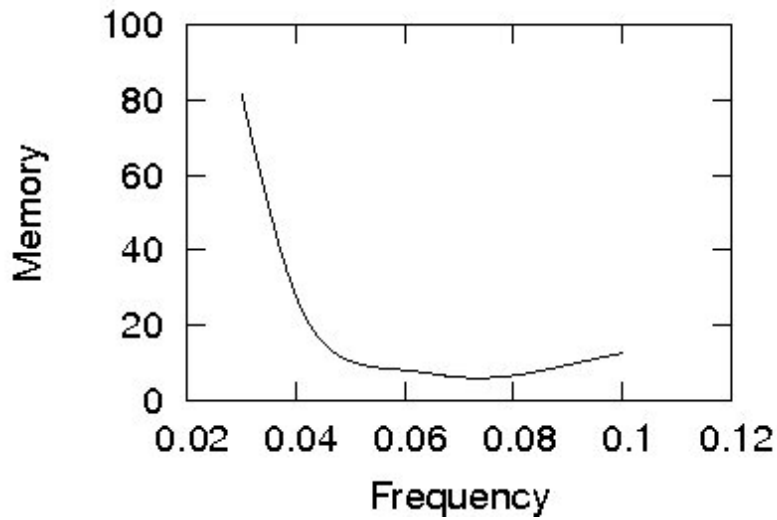
Conflicting Set

Limits...

- More than one maximal possible conflicting set
- Each choice leads to the potential loss of a candidate at the following level
- Frequency is inexact
- Rebuilding conflicting set: loss of time

Experiments





Experiments

- Implemented in C++
- Adaptation of IBM Generator
- Variation of number of items / number of objects

From Gradual Itemsets to Gradual Rules

- Post generation
- Need to mesure the variation strength

Example

Obj	A^+	C^-
o_1	22	1
o_2	28	1
o_4	35	1
o_5	38	1
o_6	44	1

Measuring strength

- Number of common objects
- Variance or co-variance
- Entropy

Conclusion

A new approach...

- Generating gradual n-itemset
- No statistical methods
- Many kinds of gradualness (increasing and decreasing)
- Scalable

Future work

- Use of order relation restrictions
- Extention to sequential patterns

LIRMM

State of the Art

E. Hüllermeier

F. Berzal & al.

C. Fiot & al.

Proposition

Definitions

Frequency Computing

Join Operation

Experiments

Generating Rules

Conclusion

Thank you for your attention
–Questions–