

# Contraintes de temps étendues pour les motifs séquentiels généralisés

Céline Fiot

Laboratoire d'Informatique de Robotique de Microélectronique



## Motifs séquentiels généralisés

Problématique

**Motivations**

Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- imposer des contraintes temporelles entre les itemsets des séquences fréquentes [3]
- plusieurs algorithmes prenant en compte ces contraintes
  - pendant le processus d'extraction (GSP, [3])
  - dans une phase assimilable à un prétraitement (GTC, [1])



## Limites

Problématique

**Motivations**

Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- un découpage strict  $\Rightarrow$  une bonne connaissance a priori des contraintes à spécifier
- une proposition de détermination automatique d'une des contraintes, en fonction d'une autre [2]



## Objectifs

- aboutir à plus de flexibilité dans la spécification des contraintes
- pouvoir déterminer la précision des contraintes des fréquents extraits

Problématique

**Motivations**

Proposition

Contraintes de temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion



## Proposition : extension des contraintes de temps

Problématique

Motivations

**Proposition**

Contraintes de  
temps étendues

Contraintes de temps

Extension des définitions

Précision temporelle  
d'une séquence

GETC

Expérimentations

Conclusion

- une définition de contraintes de temps plus souples
- un algorithme efficace pour prendre en compte toutes les contraintes de temps (étendues ou non)
- une mesure du respect des différentes contraintes



## Contraintes de temps, définition [3]

- L'écart maximum entre les items d'un itemsets  $i$  :  
*windowSize*

$$fin_i - deb_i \leq windowSize$$

- La durée minimum qui sépare deux itemsets consécutifs  $i - 1$  et  $i$  : *minGap*

$$deb_i - fin_{i-1} > minGap$$

- La durée maximum qui sépare le début de l'itemset  $i - 1$  et la fin de l'itemset suivant  $i$  : *maxGap*

$$fin_i - deb_{i-1} \leq maxGap$$

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion



## Contraintes de temps, définition [3]

- L'écart maximum entre les items d'un itemsets  $i$  :  
 $windowSize$

$$fin_i - deb_i \leq windowSize$$

- La durée minimum qui sépare deux itemsets consécutifs  
 $i - 1$  et  $i$  :  $minGap$

$$deb_i - fin_{i-1} > minGap$$

- La durée maximum qui sépare le début de l'itemset  $i - 1$   
et la fin de l'itemset suivant  $i$  :  $maxGap$

$$fin_i - deb_{i-1} \leq maxGap$$

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion



## Contraintes de temps, définition [3]

- L'écart maximum entre les items d'un itemsets  $i$  :  
 $windowSize$

$$fin_i - deb_i \leq windowSize$$

- La durée minimum qui sépare deux itemsets consécutifs  
 $i - 1$  et  $i$  :  $minGap$

$$deb_i - fin_{i-1} > minGap$$

- La durée maximum qui sépare le début de l'itemset  $i - 1$   
et la fin de l'itemset suivant  $i$  :  $maxGap$

$$fin_i - deb_{i-1} \leq maxGap$$

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

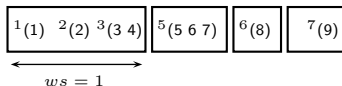




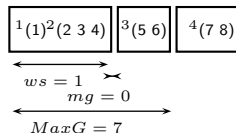
## Contraintes de temps, exemple

$windowSize=1$ ,  $mingap=0$  et  $maxgap=7$ ,  $s=\langle(1\ 2\ 3\ 4)\rangle$

Date	1	2	3	4	5	6	7
Client 1	1	2	3 4	-	5 6 7	8	9
Client 2	1	2 3 4	5 6	7 8	-	-	-



(a) séquence du client 1



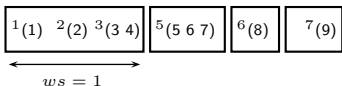
(b) séquence du client 2

$$s \not\subseteq d1, s \subseteq d2$$

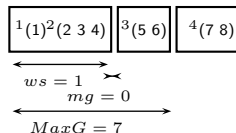
## Contraintes de temps, exemple

$windowSize=1$ ,  $mingap=0$  et  $maxgap=7$ ,  $s=\langle(1\ 2\ 3\ 4)\rangle$

Date	1	2	3	4	5	6	7
Client 1	1	2	3 4	-	5 6 7	8	9
Client 2	1	2 3 4	5 6	7 8	-	-	-



(a) séquence du client 1



(b) séquence du client 2

$$s \not\subseteq d1, s \subseteq d2$$

## Mise en œuvre

- Définition de valeurs limites utiles pour les contraintes *au delà desquelles on ne trouvera aucune séquence les respectant*
- Par analogie avec la théorie des sous-ensembles flous [4] *relaxer les contraintes  $windowSize$ ,  $minGap$  et  $maxGap$  depuis une valeur initiale  $ws$ ,  $mg$  et  $MG$  jusqu'à un certain seuil*
- Définition de degré de respect de chacune des contraintes  *$\rho_{ws}$ ,  $\rho_{mg}$  et  $\rho_{MG}$*

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Mise en œuvre

- Définition de valeurs limites utiles pour les contraintes *au delà desquelles on ne trouvera aucune séquence les respectant*
- Par analogie avec la théorie des sous-ensembles flous [4] *relaxer les contraintes  $windowSize$ ,  $minGap$  et  $maxGap$  depuis une valeur initiale  $ws$ ,  $mg$  et  $MG$  jusqu'à un certain seuil*
- Définition de degré de respect de chacune des contraintes  *$\rho_{ws}$ ,  $\rho_{mg}$  et  $\rho_{MG}$*

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Mise en œuvre

- Définition de valeurs limites utiles pour les contraintes *au delà desquelles on ne trouvera aucune séquence les respectant*
- Par analogie avec la théorie des sous-ensembles flous [4] *relaxer les contraintes  $windowSize$ ,  $minGap$  et  $maxGap$  depuis une valeur initiale  $ws$ ,  $mg$  et  $MG$  jusqu'à un certain seuil*
- Définition de degré de respect de chacune des contraintes  $\rho_{ws}$ ,  $\rho_{mg}$  et  $\rho_{MG}$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendus

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion

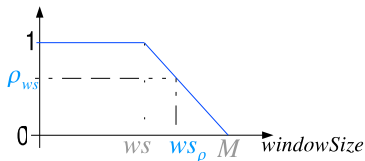


## Extension de *windowSize* (1/2)

- Valeur maximale utile : l'écart maximum pour tous les clients entre les dates min et max des transactions

$$M = \max_{c \in \mathcal{C}} (D_{c_{max}} - D_{c_{min}}) \geq \textit>windowSize$$

- Variation de *windowSize* entre *ws* et *M*... deux cas possibles
  - $ws \geq M$  : on ne fera pas varier *windowSize*
  - $ws < M$  : afin de générer d'autres séquences candidates, on va faire varier la valeur de *windowSize* entre *ws* et *M*

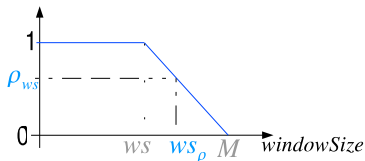


## Extension de *windowSize* (1/2)

- Valeur maximale utile : l'écart maximum pour tous les clients entre les dates min et max des transactions

$$M = \max_{c \in \mathcal{C}} (D_{c_{max}} - D_{c_{min}}) \geq \textit>windowSize$$

- Variation de *windowSize* entre *ws* et *M*... deux cas possibles
  - $ws \geq M$  : on ne fera pas varier *windowSize*
  - $ws < M$  : afin de générer d'autres séquences candidates, on va faire varier la valeur de *windowSize* entre *ws* et *M*



## Extension de *windowSize* (2/2)

- Définition du degré de respect de la contrainte *windowSize*

- Pour  $windowSize = x$ ,  $\rho_{ws} = \mu_{ws}(x)$ , c'est-à-dire :

$$\rho_{ws} = \begin{cases} \frac{1}{ws-M}x - \frac{M}{ws-M}, & \text{pour } x \geq M \\ 1, & \text{sinon} \end{cases}$$

- La valeur maximale de *windowSize*, pour un seuil de précision  $\rho_{ws}$  est donc :

$$ws_{\rho} = \lfloor (ws - M)\rho_{ws} + M \rfloor$$

Problématique

Motivations  
PropositionContraintes de  
temps étenduesContraintes de temps  
**Extension des définitions**  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion





## Extension de *windowSize*, un exemple

$s = \langle (1\ 2\ 3\ 4) \rangle$ ,  $ws = 1$ ,  $mg = 0$  et  $MG = 7$

Date	1	2	3	4	5	6	7
Client 1	1	2	3 4	-	5 6 7	8	9
Client 2	1	2 3 4	5 6	7 8	-	-	-

- Avec des contraintes de temps strictes,  $s \not\subseteq d1$ ,  $s \subseteq d2$
- Avec *windowSize* étendue,  $\rho_{ws} = 0.5$ 
  - $ws = 1$ ,  $M = \max((7 - 1), (4 - 1)) = 6 \Rightarrow$   
 $ws_{\rho} = \lfloor (1 - 6) * 0.5 + 6 \rfloor = 2$
  - dans ce cas,  $s \subseteq d1$ ,  $s \subseteq d2$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Extension de *windowSize*, un exemple

$s = \langle (1\ 2\ 3\ 4) \rangle$ ,  $ws = 1$ ,  $mg = 0$  et  $MG = 7$

Date	1	2	3	4	5	6	7
Client 1	1	2	3 4	-	5 6 7	8	9
Client 2	1	2 3 4	5 6	7 8	-	-	-

- Avec des contraintes de temps strictes,  $s \not\subseteq d1$ ,  $s \subseteq d2$
- Avec *windowSize* étendue,  $\rho_{ws} = 0.5$ 
  - $ws = 1$ ,  $M = \max((7 - 1), (4 - 1)) = 6 \Rightarrow$   
 $ws_{\rho} = \lfloor (1 - 6) * 0.5 + 6 \rfloor = 2$
  - dans ce cas,  $s \subseteq d1$ ,  $s \subseteq d2$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Extension de *windowSize*, un exemple

$s = \langle (1\ 2\ 3\ 4) \rangle$ ,  $ws = 1$ ,  $mg = 0$  et  $MG = 7$

Date	1	2	3	4	5	6	7
Client 1	1	2	3 4	-	5 6 7	8	9
Client 2	1	2 3 4	5 6	7 8	-	-	-

- Avec des contraintes de temps strictes,  $s \not\subseteq d1$ ,  $s \subseteq d2$
- Avec *windowSize* étendue,  $\rho_{ws} = 0.5$ 
  - $ws = 1$ ,  $M = \max((7 - 1), (4 - 1)) = 6 \Rightarrow$   
 $ws_{\rho} = \lfloor (1 - 6) * 0.5 + 6 \rfloor = 2$
  - dans ce cas,  $s \subseteq d1$ ,  $s \subseteq d2$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
**Extension des définitions**  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Extension de *maxGap* et *minGap*

- Degré de respect de la contrainte *maxGap*

$$\rho_{MG}(x) = \begin{cases} \frac{1}{MG-M}x - \frac{M}{MG-M}, & \text{pour } x \geq M \\ 1, & \text{sinon} \end{cases}$$

- Valeur maximale de *maxGap*, pour un seuil de précision

$\rho_{MG}$  :

$$MG_{\rho} = \lfloor (MG - M)\rho_{MG} + M \rfloor$$

- Degré de respect de la contrainte *minGap*

$$\rho_{mg}(x) = \begin{cases} \frac{1}{mg-m}x - \frac{m}{mg-m}, & \text{pour } x \leq mg \\ 1, & \text{sinon} \end{cases}$$

- Valeur maximale de *minGap*, pour un seuil de précision

$\rho_{mg}$  :

$$mg_{\rho} = \lceil (mg - m)\rho + m \rceil$$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions

 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



## Extension de *maxGap* et *minGap*

- Degré de respect de la contrainte *maxGap*

$$\rho_{MG}(x) = \begin{cases} \frac{1}{MG-M}x - \frac{M}{MG-M}, & \text{pour } x \geq M \\ 1, & \text{sinon} \end{cases}$$

- Valeur maximale de *maxGap*, pour un seuil de précision

$\rho_{MG}$  :

$$MG_{\rho} = \lfloor (MG - M)\rho_{MG} + M \rfloor$$

- Degré de respect de la contrainte *minGap*

$$\rho_{mg}(x) = \begin{cases} \frac{1}{mg-m}x - \frac{m}{mg-m}, & \text{pour } x \leq mg \\ 1, & \text{sinon} \end{cases}$$

- Valeur maximale de *minGap*, pour un seuil de précision

$\rho_{mg}$  :

$$mg_{\rho} = \lceil (mg - m)\rho + m \rceil$$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions

**Précision temporelle  
 d'une séquence**

GETC

Expérimentations

Conclusion

## Précision temporelle d'une séquence $s$

- la précision avec laquelle la séquence  $s$  respecte les contraintes de temps initiale

la précision d'une séquence  $s = \langle s_1 \cdots s_n \rangle$  pour un client  $c$  :

$$\rho(s, c) = \perp_{s \in \mathcal{C}} \left( \overline{\perp}_{i \in [1, n]} (\mu_{ws}(\text{date}(s_{u_i}) - \text{date}(s_{l_i}))), \right. \\ \left. \overline{\perp}_{i \in [2, n]} (\mu_{mg}(\text{date}(s_{l_i}) - \text{date}(s_{u_{i-1}}))), \right. \\ \left. \mu_{MG}(\text{date}(s_{u_i}) - \text{date}(s_{l_{i-1}}))) \right) \quad (1)$$

Pour l'ensemble de la base,

$$\Upsilon(s) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \rho(s, c) \quad (2)$$

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions

**Précision temporelle  
 d'une séquence**

GETC

Expérimentations

Conclusion

## Précision temporelle d'une séquence $s$

- la précision avec laquelle la séquence  $s$  respecte les contraintes de temps initiale

la précision d'une séquence  $s = \langle s_1 \cdots s_n \rangle$  pour un client  $c$  :

$$\rho(s, c) = \perp_{s \in \mathcal{C}_c} \left( \overline{\top}_{i \in [1, n]} (\mu_{ws}(\text{date}(s_{u_i}) - \text{date}(s_{l_i}))), \right. \\ \left. \overline{\top}_{i \in [2, n]} (\mu_{mg}(\text{date}(s_{l_i}) - \text{date}(s_{u_{i-1}}))), \right. \\ \left. \mu_{MG}(\text{date}(s_{u_i}) - \text{date}(s_{l_{i-1}}))) \right) \quad (1)$$

Pour l'ensemble de la base,

$$\Upsilon(s) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \rho(s, c) \quad (2)$$

## GETC - Principe

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

Basé sur le principe de GTC :

- transformer une séquence d'un client en un graphe de séquences respectant les contraintes de temps
- utiliser les graphes de séquences des différents clients pour déterminer les fréquents avec PSP
- calculer la précision des séquences extraites





## GETC - Principe

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

Basé sur le principe de GTC :

- transformer une séquence d'un client en un graphe de séquences respectant les contraintes de temps
- utiliser les graphes de séquences des différents clients pour déterminer les fréquents avec PSP
- calculer la précision des séquences extraites



## GETC - Principe

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

Basé sur le principe de GTC :

- transformer une séquence d'un client en un graphe de séquences respectant les contraintes de temps
- utiliser les graphes de séquences des différents clients pour déterminer les fréquents avec PSP
- calculer la précision des séquences extraites



# GETC - Toutes les séquences fréquentes ?

## Théorème

L'inclusion de de séquence est impossible après l'exécution de l'algorithme GETC.

## Théorème

L'algorithme GETC construit exactement toutes les solutions de la plus grande taille possible pour les séquences respectant *minGap* et *maxGap*.

## Théorème

L'algorithme *addWindowSize* construit exactement tous les sommets susceptibles de contribuer à la construction de toutes les solutions de la plus grande taille possible pour les séquences respectant *minGap* et *maxGap*.

## Théorème

L'algorithme GETC construit exactement toutes les séquences de la plus grande taille possible pour les séquences respectant *windowSize*, *minGap* et *maxGap*.

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion



# GETC - Construction des graphes de séquences

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Création de l'ensemble des sommets
  - ajout des itemsets de la séquence du client  $c$
  - ajout de l'ensemble des combinaisons d'itemsets permises pour les différentes valeurs de  $windowSize$
- Pour chaque sommet, on ajoute les arcs respectant les contraintes  $minGap$  et  $maxGap$
- Suppression des séquences incluses



## GETC - Construction des graphes de séquences

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Création de l'ensemble des sommets
  - ajout des itemsets de la séquence du client  $c$
  - ajout de l'ensemble des combinaisons d'itemsets permises pour les différentes valeurs de  $windowSize$
- Pour chaque sommet, on ajoute les arcs respectant les contraintes  $minGap$  et  $maxGap$
- Suppression des séquences incluses



## GETC - Construction des graphes de séquences

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Création de l'ensemble des sommets
  - ajout des itemsets de la séquence du client  $c$
  - ajout de l'ensemble des combinaisons d'itemsets permises pour les différentes valeurs de  $windowSize$
- Pour chaque sommet, on ajoute les arcs respectant les contraintes  $minGap$  et  $maxGap$
- Suppression des séquences incluses



## Construction des graphes de séquences : un exemple

Problématique

 Motivations  
 Proposition

Contraintes de temps étendues

 Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence

GETC

Expérimentations

Conclusion

Client 1		Client 2		Client 3	
Date	Items	Date	Items	Date	Items
1	1	1	2 3	1	1 2
4	2 3	3	4	4	3
5	3 4	6	5	5	3 4
6	4	8	6	6	4
8	4			9	5 6
10	5				
12	6				
17	7				
18	8				

$minSupp=70\%$ ,  $ws=2$  -  $\rho_{ws} = 0.87$ ,  $mg=2$  -  $\rho_{mg} = 0.5$ ,  
 $MG=4$  -  $\rho_{MG} = 0.85$ .

Quelles sont les séquences maximales fréquentes respectant les contraintes de temps étendues ?

# Construction des graphes de séquences : un exemple (1/4)

Problématique  
 Motivations  
 Proposition

Contraintes de temps étendues

Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion

Construction du graphe de séquences du client 3.  
 Création de l'ensemble des sommets initiaux

(1 2)    (3)    (3 4)    (4)    (5 6)

●           ●           ●           ●           ●

⏟           ⏟           ⏟           ⏟           ⏟

$I_1$             $II_4$             $III_5$             $IV_6$             $V_9$





# Construction des graphes de séquences : un exemple (2/4)

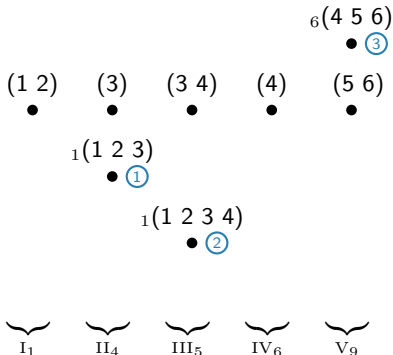
Problématique  
 Motivations  
 Proposition

Contraintes de temps étendues  
 Contraintes de temps  
 Extension des définitions  
 Précision temporelle d'une séquence  
 GETC

Expérimentations

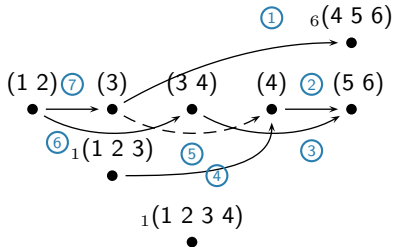
Conclusion

Construction du graphe de séquences du client 3.  
 Création des sommets combinaisons ( $ws_\rho = 4$ )



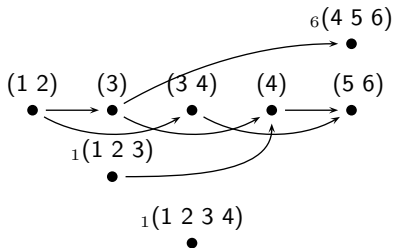
## Construction des graphes de séquences : un exemple (3/4)

Construction du graphe de séquences du client 3.  
 Construction des arcs selon  $minGap$  et  $maxGap$  ( $mg_\rho = 1$  et  $MG_\rho = 6$ )



# Construction des graphes de séquences : un exemple (4/4)

Construction du graphe de séquences du client 3.  
Suppression des inclusions - Graphe final pour le client 3



Problématique  
Motivations  
Proposition

Contraintes de temps étendus  
Contraintes de temps  
Extension des définitions  
Précision temporelle d'une séquence  
GETC

Expérimentations

Conclusion



## GETC - Calcul de la précision temporelle

Problématique

Motivations  
Proposition

Contraintes de  
temps étendus

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Pour chaque sommet du graphe du client  $c$ 
  - valuation par la précision selon la valeur de  $windowSize$
- Pour chaque arc  $(u, v)$  du graphe
  - valuation par la précision  

$$\top(\mu_{mg}(deb_v - fin_u), \mu_{MG}(fin_v - deb_u))$$
 selon les valeurs de  $minGap$  et  $maxGap$
- Pour chaque séquence de l'arbre des fréquents,
  - on incrémente sa précision pour le client  $c$  grâce aux valuations du graphe



## GETC - Calcul de la précision temporelle

Problématique

Motivations  
Proposition

Contraintes de temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Pour chaque sommet du graphe du client  $c$ 
  - valuation par la précision selon la valeur de  $windowSize$
- Pour chaque arc  $(u, v)$  du graphe
  - valuation par la précision  
 $\top(\mu_{mg}(deb_v - fin_u), \mu_{MG}(fin_v - deb_u))$  selon les valeurs de  $minGap$  et  $maxGap$
- Pour chaque séquence de l'arbre des fréquents,
  - on incrémente sa précision pour le client  $c$  grâce aux valuations du graphe



## GETC - Calcul de la précision temporelle

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

- Pour chaque sommet du graphe du client  $c$ 
  - valuation par la précision selon la valeur de  $windowSize$
- Pour chaque arc  $(u, v)$  du graphe
  - valuation par la précision  
 $\top(\mu_{mg}(deb_v - fin_u), \mu_{MG}(fin_v - deb_u))$  selon les  
valeurs de  $minGap$  et  $maxGap$
- Pour chaque séquence de l'arbre des fréquents,
  - on incrémente sa précision pour le client  $c$  grâce aux  
valuations du graphe



# Calcul de la précision temporelle : un exemple (1/3)

Problématique  
Motivations  
Proposition

Contraintes de temps étendues  
Contraintes de temps  
Extension des définitions  
Précision temporelle d'une séquence  
GETC

Expérimentations

Conclusion

- valuation des graphes :
  - pour chaque sommet,  $windowSize \leq 2 \Rightarrow v_{ws} = 1$ ,  
 $windowSize=3 \Rightarrow v_{ws} = 0.93$  et  
 $windowSize=4, \Rightarrow v_{ws} = 0.87$ ,
  - pour chaque arc,
    - $minGap=2 \Rightarrow v_{mg} = 1$  et  $minGap=1 \Rightarrow v_{mg} = 0.5$
    - $maxGap \leq 4 \Rightarrow v_{MG} = 1$ ,  $maxGap=5 \Rightarrow v_{MG} = 0.92$   
et  $maxGap=6 \Rightarrow v_{MG} = 0.85$

# Calcul de la précision temporelle : un exemple (2/3)

Problématique  
 Motivations  
 Proposition

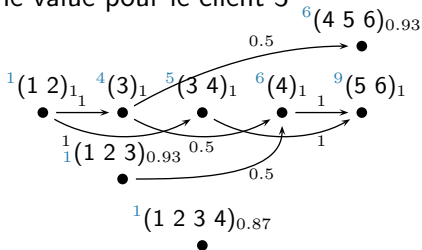
Contraintes de temps étendues

Contraintes de temps  
 Extension des définitions  
 Précision temporelle d'une séquence  
 GETC

Expérimentations

Conclusion

Graphe valué pour le client 3



client 3



# Calcul de la précision temporelle : un exemple (3/3)

Précision temporelle des motifs séquentiels extraits :

Motifs séquentiels	$q_{CI1}$	$q_{CI2}$	$q_{CI3}$	$\Upsilon$
$\langle (2\ 3\ 4) \rangle$	1	1	0.87	0.96
$\langle (2\ 3)(4)(5\ 6) \rangle$	0.5	0.5	0.5	0.5
$\langle (2)(4\ 5) \rangle$	0.85	0.5	1	0.78
$\langle (3\ 4)(5) \rangle$	0.85	1	1	0.95
$\langle (3\ 4)(6) \rangle$	0.5	0.5	1	0.67
$\langle (3)(4\ 5) \rangle$	0.85	0.5	0.5	0.62

Problématique  
 Motivations  
 Proposition

Contraintes de temps étendues  
 Contraintes de temps  
 Extension des définitions  
 Précision temporelle d'une séquence

GETC

Expérimentations

Conclusion



## Expérimentations (1/5)

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion

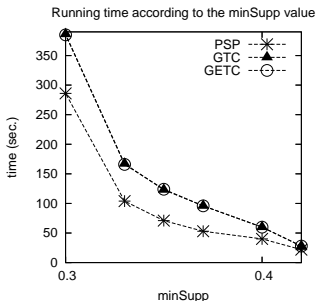
- plusieurs jeux de données synthétiques (1000 clients, env. 20 transactions/clients, env. 15 items/trans., choisis parmi 1000)
- GETC vs. GTC, GETC vs. PSP



## Expérimentations (2/5)

### Comparer les temps d'exécution sans contrainte de temps

( $wS = 0$ ,  $m.g = 0$  et  $MG = \infty$  pour GTC et GETC avec  $\rho_{wS} = \rho_{m.g} = \rho_{MG} = 1$ )



### Résultats

GETC et GTC similaires, temps d'exécution quasiment identiques, les motifs extraits sont les mêmes.

Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

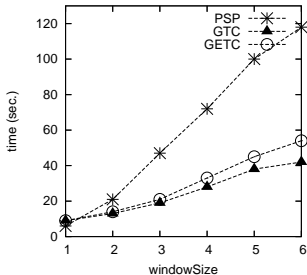
Conclusion

## Expérimentations (3/5)

### Comparer les temps d'exécution avec des contraintes de temps

$$\rho_{ws} = \rho_{mg} = \rho_{MG} = 1$$

Running time according to the windowSize value



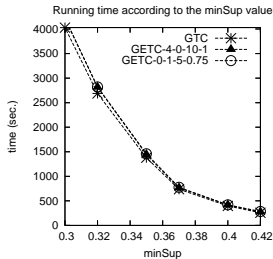
### Résultats

comportement linéaire de GETC, proche de GTC + traitement de la précision.

## Expérimentations (4/5)

### Comparaison de GTC et GETC en fonction de $minSupp$

- *GETC* avec  $windowSize=0$ ,  $minGap=1$  et  $maxGap=5$  avec une précision de 0.75, qui nous donne  $ws_{\rho} = 4$ ,  $mg_{\rho} = 0$  et  $MG_{\rho} = 10$ ,
- *GETC* avec  $windowSize=4$ ,  $minGap=0$  et  $maxGap=10$  avec une précision de 1,
- *GTC* avec  $windowSize=4$ ,  $minGap=0$  et  $maxGap=10$ ,

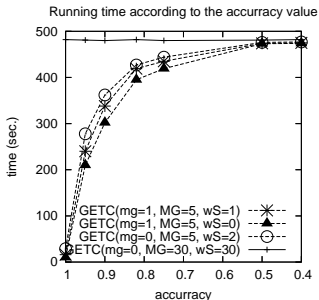


### Résultats

GETC et GTC  $\Rightarrow$  temps équivalents, les mêmes motifs séquentiels + leur précision temporelle avec GETC.

## Expérimentations (5/5)

Evolution du temps d'extraction en fonction de la précision pour un support minimum de 0.37



### Résultats

Le temps d'extraction atteint une valeur limite.

## Conclusion

- une extension des contraintes de temps pour les motifs séquentiels
- une méthode efficace, plus flexible
- un indicateur des contraintes de temps utilisées pour générer une séquence
- des motifs séquentiels flous généralisés ?

Problématique

Motivations  
Proposition

Contraintes de  
temps étendues

Contraintes de temps  
Extension des définitions  
Précision temporelle  
d'une séquence  
GETC

Expérimentations

Conclusion



Problématique

 Motivations  
 Proposition

 Contraintes de  
 temps étendues

 Contraintes de temps  
 Extension des définitions  
 Précision temporelle  
 d'une séquence  
 GETC

Expérimentations

Conclusion

## Bibliographie



F. Masegla, P. Poncelet et M. Teisseire

*Extraction efficace de motifs séquentiels généralisés : le prétraitement des données.*

(BDA '99)



N. Meger et C. Rigotti

*Constraint-based mining of episode rules and optimal window sizes.*

(PKDD'04)



R. Srikant et R. Agrawal

*Mining sequential patterns : Generalizations and performance improvements.*

(EDBT '96)



L.A. Zadeh

*Fuzzy Sets.*

Information and Control (3)-8, 1965.