A Symbolic Approach for Extracting Disjunctive Patterns from Data and Discovering Conceptual Clustering Models

Said Jabbour

CRIL - CNRS, Université d'Artois, France

CAVIAR@2025

Context

- Pattern mining is a key problem in many applications (Data Mining, Machine Learning, etc.)
- Mining generalized patterns (generators, disjunctive association rules, etc.)
- **Disjunctive association rules** extend the expressive power of classical association rules (e.g., when customers buy bread they also buy butter or milk).
- Despite their generality, classical patterns are not sufficiently expressive to capture interesting disjunctive relationships among items.

Itemsets & Association Rules

- Ω: a set of items
- An **itemset** X over Ω: a subset of Ω.
- **A transaction database**: collection of transactions $D = \{(1, T_1), \ldots, (m, T_m)\}$.
- **Cover** of X: $\mathcal{C}(X, \mathcal{D}) = \{i \in [1..m] \mid (i, T) \in D \text{ and } X \subseteq T_i\}$.
- **Support** of X: $Supp(X, \mathcal{D}) = |\mathcal{C}(X, \mathcal{D})|$.
- **Generalized disjunctive itemset (GDI)**: a disjunctive collection of itemsets $[X_1, \ldots, X_p]$ or $X_1 \vee \ldots \vee X_p$ [Nanavati et al., 2001]
- **Support of a GDI**:

$$Supp([X], \mathcal{D}) = \frac{|\bigcup_{X_i \in X} \mathcal{C}(X_i, \mathcal{D})|}{|\mathcal{D}|}$$

tid	Itemset								
t_1		b	c	d					
t_2	a		c	d					
t_3	a	b		d	e				
t_4	a	b	c		e	f			
t_5	a		c			f		h	
t_6				d	e		g	h	i
t_7					e	f	g	h	i
t_8					e	f	g		i
t_9						f	g	h	

Itemsets & Association Rules

- **Association rule** (**AR**): an implication $X \to Y$ where $X \subseteq \Omega$ and $Y \subseteq \Omega$ and $X \cap Y = \emptyset$.

$$supp(X \to Y, \mathcal{D}) = \frac{Supp(X \cup Y, \mathcal{D})}{|\mathcal{D}|}$$

$$conf(X \to Y, \mathcal{D}) = \frac{Supp(X \cup Y, \mathcal{D})}{Supp(X, \mathcal{D})}$$

tid	Itemset								
t_1		b	c	d					
t_2	a		c	d					
t_3	a	b		d	e				
t_4	a	b	c		e	f			
t_5	a		c			f		h	
t_6				d	e		g	h	i
t_7					e	f	g	h	i
t_8					e	f	g		i
t_9						f	g	h	

- **Generalized disjunctive association rule** (GDAR): an AR of the form $[X_1, \ldots, X_p] \to [Y_1, \ldots, Y_q]$

$$supp([X] \to [Y], \mathcal{D}) = \frac{|\bigcup_{X_i \in X} \mathcal{C}(X_i, \mathcal{D}) \cap \bigcup_{Y_i \in Y} \mathcal{C}(Y_i, \mathcal{D})|}{|\mathcal{D}|}$$

$$conf([X] \to [Y], \mathcal{D}) = \frac{|\bigcup_{X_i \in X} \mathcal{C}(X_i, \mathcal{D}) \cap \bigcup_{Y_i \in Y} \mathcal{C}(Y_i, \mathcal{D})|}{|\bigcup_{X_i \in X} \mathcal{C}(X_i, \mathcal{D})|}$$

k-Disjunctive Cover

k-disjunctive cover and support

k-disjunctive cover: $\mathcal{C}^k(X, D) = \{(i, T) \in \mathcal{D} \mid T \cap X \neq \emptyset \text{ and } |X \setminus T| \leq k\}$

k-disjunctive support: $Supp^k(X, \mathcal{D}) = |\mathcal{C}^k(X, \mathcal{D})|$

Associated GDI: $\text{GDI}(X) = [X \cap T \mid T \in \mathcal{C}^k(X, \mathcal{D})]$

- $cover(\{a, b, c\}, \mathcal{D}) = \mathcal{C}^0(\{a, b, c\}, \mathcal{D}) = \{t_4\}$
- $\mathcal{C}^1(\{a, b, c\}, \mathcal{D}) = \{t_1, t_2, t_3, t_4, t_5\}$
- $\text{GDI}(\{a, b, c\}) = \{b, c\} \vee \{a, c\} \vee \{a, b\} \vee \{a, b, c\}$

tid	Itemset								
t_1		b	c	d					
t_2	a		c	d					
t_3	a	b		d	e				
t_4	a	b	c		e	f			
t_5	a		c			f		h	
t_6				d	e		g	h	i
t_7					e	f	g	h	i
t_8					e	f	g		i
t_9						f	g	h	

k-Disjunctive Support based Frequent Itemset

Definition

An itemset X is a **k-disjunctive support based frequent itemset** (k-*DSFI*) iff $Supp^k(X, \mathcal{D}) \geq \alpha$.

Proposition

- *Deciding whether there exists an itemset X s.t. $|X| \geq l$ and $Supp^{k>0}(X, \mathcal{D}) = |\mathcal{D}|$ is NP-complete.*
- *Deciding whether there exists a $k-$DSFI X in $\mathcal{D}$ s.t. $|X| \geq l$ and $\alpha > 0$ is NP-complete.*

(k, k')-Disjunctive Support based AR

Definition

Let $\mathcal{D}$ be a transaction database, and k, k' two positive integers. We define the (k, k')-***disjunctive support*** and (k, k')-***disjunctive confidence*** of an AR $X \to Y$ as:

$$supp^{k,k'}(X \to Y, \mathcal{D}) = \frac{|\mathcal{C}^k(X, \mathcal{D}) \cap \mathcal{C}^{k'}(Y, \mathcal{D})|}{|\mathcal{D}|}$$

$$conf^{k,k'}(X \to Y, \mathcal{D}) = \frac{Supp^{k,k'}(X \to Y, \mathcal{D})}{Supp^k(X, \mathcal{D})}$$

Similarly to itemsets, for a given AR $X \to Y$ one can deduce a GDAR w.r.t. the parameters k and k':

$$\text{GDAR}(X \to Y) = \quad [X \cap T \mid T \in \mathcal{C}^k(X, \mathcal{D})] \to [Y \cap T \mid T \in \mathcal{C}^{k'}(Y, \mathcal{D})]$$

(k, k')-Disjunctive Support based AR

Definition

Let $\alpha > 0$ a minimum support threshold, and $\beta > 0$ a minimum confidence threshold. Then, an AR $X \rightarrow Y$ is a (k, k')-***disjunctive support based valid AR*** ((k, k')-*DSVAR*, for short) if and only if $Supp^{k,k'}(X \rightarrow Y, \mathcal{D}) \geq \alpha$ and $Conf^{k,k'}(X \rightarrow Y, \mathcal{D}) \geq \beta$.

- $r : \{d, e\} \rightarrow \{a, b, c\}$ *is a* $(1, 1) - DSVAR$
- $supp^{1,1}(r, \mathcal{D}) = \frac{4}{9}$ $conf^{1,1}(r, \mathcal{D}) = \frac{4}{7}$
- GDAR$(r) : \{d\} \vee \{e\} \vee \{d, e\} \rightarrow \{a, b, c\} \vee \{a, b\} \vee \{a, c\} \vee \{b, c\}$

tid	Itemset							
t_1		b	c	d				
t_2	a		c	d				
t_3	a	b		d	e			
t_4	a	b	c		e	f		
t_5	a		c			f		h
t_6				d	e		g	h
t_7					e	f	g	h
t_8					e	f	g	
t_9						f	g	h

Symbolic Encoding of k-DFSIs Problem

Logical Constraints

Cover constraint	$\bigwedge_{(i,T_i)\in\mathcal{D}} (q_i \leftrightarrow (\sum_{a\in\Omega\setminus T_i} x_a \leq k))$
Frequency constraint	$\sum_{i=1}^{m} q_i \geq m \times \alpha$
Closure constraint	$\bigwedge_{a\in\Omega} (\neg x_a \rightarrow \bigvee_{(i,T_i)\in\mathcal{D}, a\notin T_i} (q_i \wedge \sum_{b\in\Omega\setminus T_i} x_b = k))$
Item frequency	$\bigwedge_{a\in\Omega} (x_a \rightarrow \sum_{(i,T_i)\in\mathcal{D} \mid a\in T_i} q_i \geq \gamma)$

tid	Itemset								
$q_1 \rightarrow t_1$		b	c	d					
$q_2 \rightarrow t_2$	a		c	d					
$q_3 \rightarrow t_3$	a	b		d	e				
$q_4 \rightarrow t_4$	a	b	c		e	f			
$q_5 \rightarrow t_5$	a		c			f		h	
$q_6 \rightarrow t_6$				d	e		g	h	i
$q_7 \rightarrow t_7$					e	f	g	h	i
$q_8 \rightarrow t_8$					e	f	g		i
$q_9 \rightarrow t_9$						f	g	h	
	↑	↑	↑	↑	↑	↑	↑	↑	↑
	x_a	x_b	x_c	x_d	x_e	x_f	x_g	x_h	x_i

Symbolic Encoding of the (k, k')-DSVARs Problem

Logical constraints

$$\bigwedge_{a \in \Omega} (\neg x_a \vee \neg y_a) \quad (1)$$

$$\bigwedge_{i=1}^{m} (p_i \leftrightarrow \sum_{a \in \Omega \setminus T_i} x_a \leq k) \quad (2)$$

$$\bigwedge_{i=1}^{m} (q_i \leftrightarrow p_i \wedge (\sum_{a \in \Omega \setminus T_i} y_a \leq k')) \quad (3)$$

$$\sum_{i=1}^{m} q_i \geq m \times \alpha \quad (4)$$

$$100 \times \sum_{i=1}^{m} q_i - \beta \times \sum_{i=1}^{m} p_i \geq 0 \quad (5)$$

$$\bigwedge_{a \in \Omega} ((x_a \vee y_a) \rightarrow \sum_{(i,T_i) \in \mathcal{D},\ a \in T_i} q_i \geq \gamma) \quad (6)$$

$$\bigwedge_{a \in \Omega} (\neg x_a \wedge \neg y_a \rightarrow ((\sum_{(i,T_i) \in \mathcal{D},\ a \in T_i} q_i < \gamma) \vee (\bigvee_{(i,T_i) \in \mathcal{D},\ a \notin T_i} (q_i \wedge (\sum_{b \in \Omega \setminus T_i} x_b = k) \vee (\sum_{b \in \Omega \setminus T_i} y_b = k')))) \quad (7)$$

Experiments

Setup

- Linux machine, Intel Xeon quad-core processors (32GB of RAM running at 2.66 GHz), 2 hours time out and 10 GB.
- Different datasets from the FIMI[1] repository.
- The minimum confidence threshold β to 95%[2] while the value of γ is identical to α.
- Sequential and Parallel evaluation to find all closed k'-DSVARs ($k = 0$).
- `Minisat` Solver as a allSAT model generation.

[1] `http://fimi.ua.ac.be/data/`

[2] Similar results were observed when β is set to 85% and 90%.

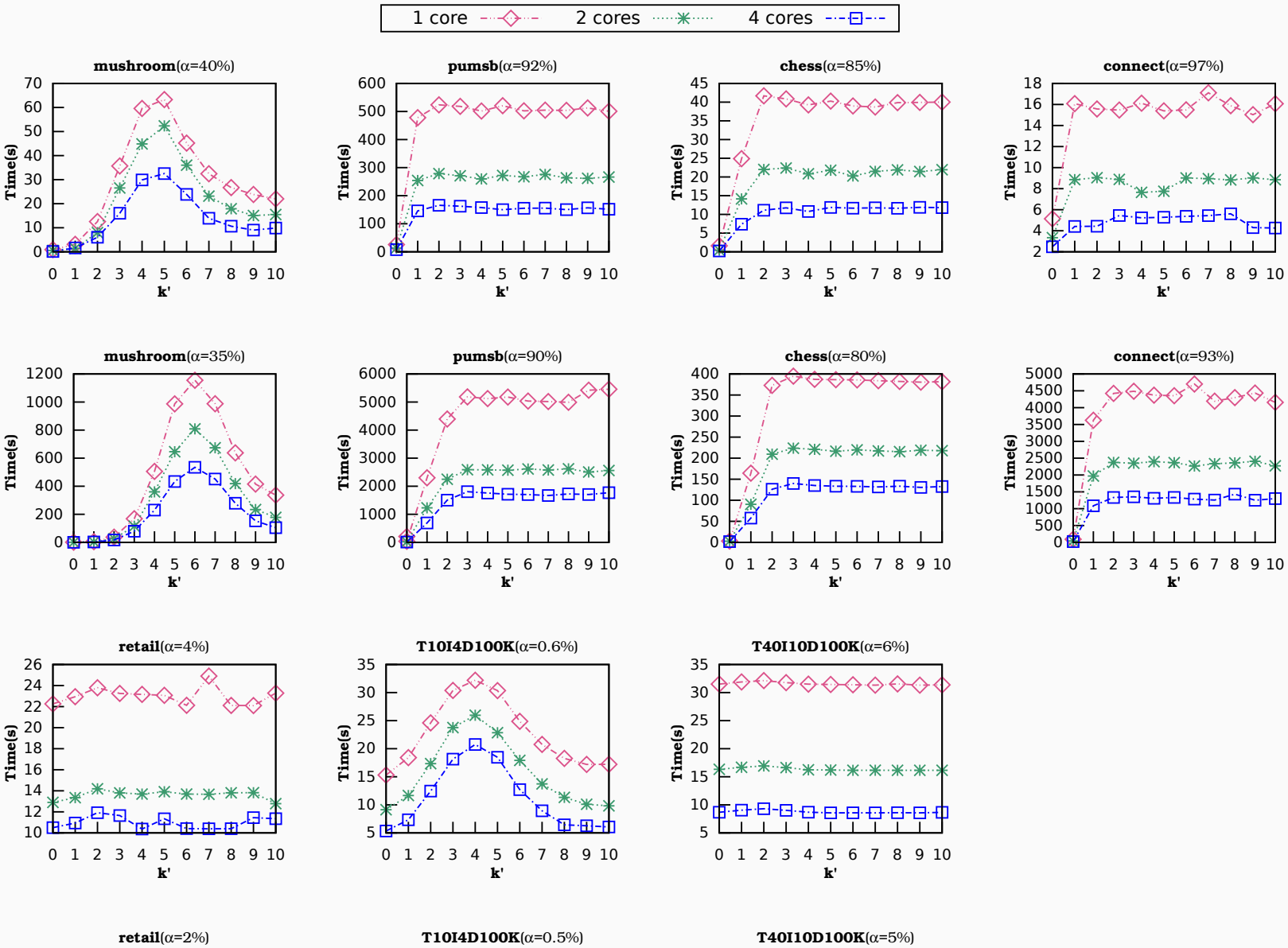
Experiments

Sequential evaluation

Instance	α	k'=0		k'=1		k'=2		k'=3		k'=4	
		#0-CDSVARs	time(s)	#1-CDSVARs	time(s)	#2-CDSVARs	time(s)	#3-CDSVARs	time(s)	#4-CDSVARs	time(s)
chess 3196, 75	75%	287183	8.43	11191898	649.83	48078061	3164.66	80848542	4777.37	85301308	4926.95
	80%	109356	2.76	3488808	149.73	8110917	342.628	8817802	361.38	8818493	356.76
	85%	32201	0.70	778548	23.14	1275841	37.84	1276771	37.13	1276771	36.82
	90%	6471	0.16	85483	2.13	94152	2.37	94152	2.34	94152	2.33
	95%	539	0.05	1358	0.06	1358	0.06	1358	0.06	1358	0.06
mushroom 8124, 112	20%	20062	3.08	107894	91.29	535763	1271.64	–	–	–	–
	25%	4504	1.81	30186	26.41	161307	244.89	971009	1895.89	–	–
	30%	2688	0.60	15800	8.42	67367	64.15	322975	323.41	1363490	1180.99
	35%	1424	0.43	7753	4.98	35956	35.31	175211	166.04	660736	497.63
	40%	576	0.32	3143	2.03	15220	10.81	59547	32.74	159198	60.39
pumsb 49046, 7117	89%	50321	49.26	4132701	7010.22	–	–	–	–	–	–
	90%	30202	38.48	2116159	2300.65	4718682	4393.83	5983488	5183.34	6012634	5122.55
	91%	17662	31.76	1041036	977.62	2071177	1631.89	2557882	1858.79	2571349	1864.72
	92%	9630	25.76	430130	477.98	438667	523.63	438667	518.18	438667	501.55
	93%	4931	20.33	210517	208.91	212549	210.37	212549	206.27	212549	202.29
	94%	2448	19.15	83100	79.57	84093	78.95	84093	77.14	84093	80.23
connect 67557, 129	91%	407225	217.18	–	–	–	–	–	–	–	–
	93%	159895	91.60	4652371	3625.45	5653940	4422.61	5653940	4486.13	5653940	4370.51
	95%	40579	19.14	1018059	709.92	1169408	798.10	1169408	789.23	1169408	785.37
	97%	5974	5.14	27885	16.08	27885	15.57	27885	15.47	27885	16.12
retail 88162, 16470	1%	172	29.30	386	143.03	1204	1260.06	–	–	–	–
	2%	60	24.85	141	32.76	279	59.37	530	110.65	922	165.38
	3%	35	23.43	76	26.17	140	29.28	214	28.93	271	26.19
	4%	18	22.14	43	22.87	65	22.76	76	22.24	78	23.78
	5%	16	21.75	39	22.43	57	22.34	66	21.81	68	21.58

Experiments

Parallel evaluation



On the Discovery of Conceptual Clustering Models disjunctive Patterns

Introduction: Conceptual Clustering ?

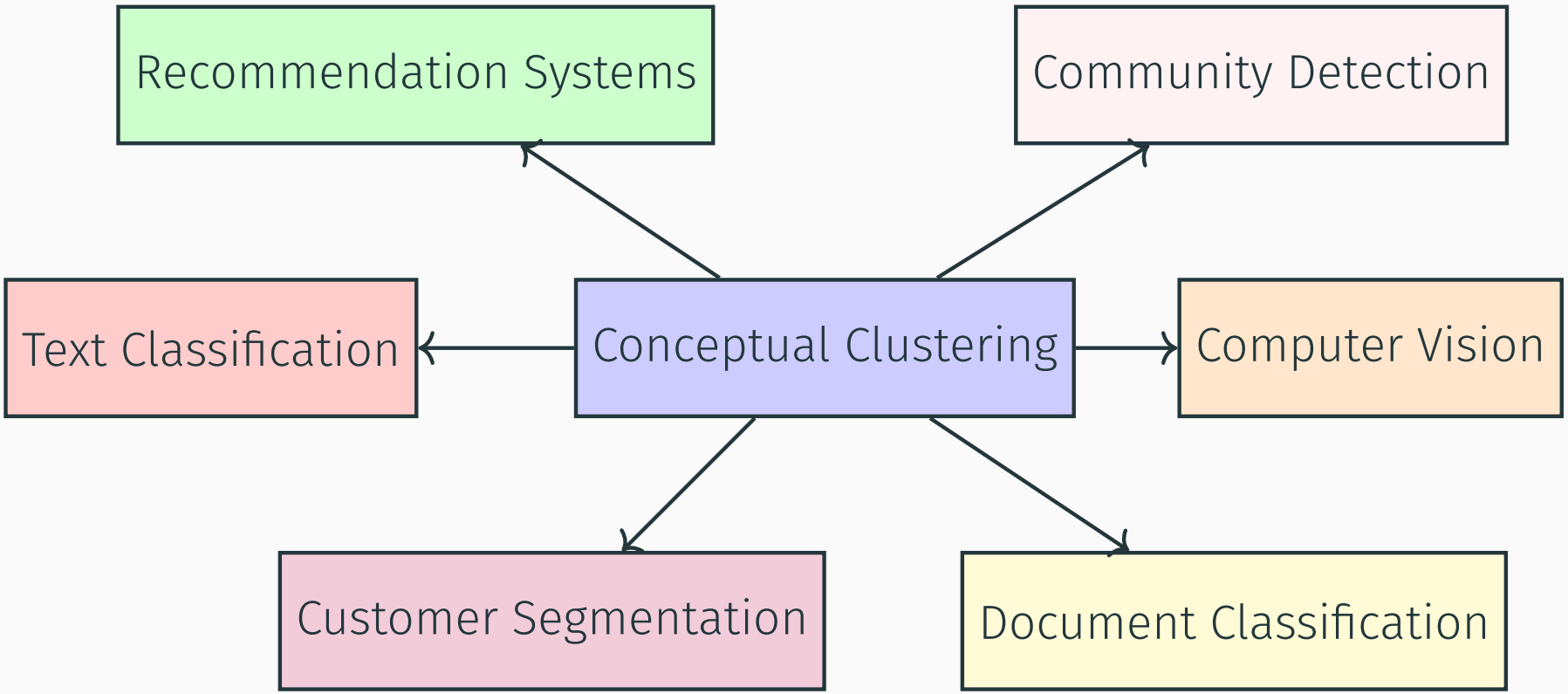
- Conceptual Clustering is an $\mathcal{NP}$-Complete problem.
- Let $\mathcal{D}$ be a dataset with m transactions and n features (items).
- Given $\beta > 0$: the number of clusters.
- The goal is to identify β disjoint clusters covering $\mathcal{D}$, where each cluster is **covered** (or described) by a **closed** pattern.
- **Cover**: $\mathsf{C}(P, \mathcal{D}) = \{i \in [1..m] \mid T_i \in \mathcal{D} \text{ and } P \subseteq T_i\}$.
- **Support**: $\mathrm{Supp}(P, \mathcal{D}) = |\mathsf{C}(P, \mathcal{D})|$.

Introduction: Conceptual Clustering ? Example

- Given $\beta = 2$.
- Given two patterns: $P = \{a, b\}$ and $Q = \{c, d\}$.
- $\mathrm{C}(P, \mathcal{D}) = \{T_1, T_2\}$ and $\mathrm{C}(Q, \mathcal{D}) = \{T_3, T_4, T_5\}$.
- $\mathrm{Supp}(P, \mathcal{D}) = 2$ and $\mathrm{Supp}(Q, \mathcal{D}) = 3$.
- P and Q are **closed** patterns.

Transactions	Items			
T_1	a	b		d
T_2	a	b	c	
T_3	a		c	d
T_4		b	c	d
T_5	a		c	d

Introduction: Conceptual Clustering applications



Introduction: Existing Conceptual Clustering approaches

- Using Satisfiability Problem (SAT)
- Using Constraint Programming (CP)
- Using Integer Linear Programming (ILP)
  - Enumerate the set of Closed Patterns (minimum support = 1)
  - Encode the Conceptual Clustering Problem into an ILP
  - Extract clusters using an ILP solver (Simplex, BnB, etc)

Introduction: CCP approach: limitation

- Plausible clusters: $\{T_1, T_2, T_3, T_4\}$ and $\{T_5, T_6, T_7, T_8\}$.
- With classical closed Patterns (CCP), **it is not possible to have 2 clusters.**

Transactions	Items
T_1	a b c
T_2	a b d
T_3	a c d
T_4	b c d
T_5	f g h
T_6	e g h
T_7	e f h
T_8	e f g

$\Longrightarrow$ We need a new pattern representation to represent clusters.

Contribution

Contribution: SAT + ILP approach

- **Main idea**: allow transactions to not contain the whole items of the representative pattern.
- k-Relaxed frequent Patterns (k-RFP) to represent clusters: relax the support definition.

Definition

Let $\mathcal{D}$ be a dataset and k a positive integer. Then, the ***k*-cover** of a pattern P is $\mathrm{C}^k(P, \mathcal{D}) = \{i \in [1..m] \mid T_i \cap P \neq \emptyset \text{ and } |P \setminus T_i| \leq k\}$. The ***k*-support** of P is defined as usual as: $\mathrm{Supp}^k(P, \mathcal{D}) = |\mathrm{C}^k(P, \mathcal{D})|$.

Contribution: Encoding of k-RFP in Propositional Logic

Cover Constraint

$$\bigwedge_{T_i \in \mathcal{D}} (q_i \leftrightarrow (\sum_{a \in \Omega \setminus T_i} x_a \leq k))$$

Frequency Constraint

$$\sum_{i=1}^{m} q_i \geq m \times \alpha$$

Closure Constraint

$$\bigwedge_{a \in \Omega} (\neg x_a \rightarrow \bigvee_{T_i \in \mathcal{D}, a \in T_i} (q_i \wedge \sum_{b \in \Omega \setminus T_i} x_b = k))$$

Item Frequency Constraint

$$\bigwedge_{T_i \in \mathcal{D}} (x_a \rightarrow (\sum_{T_i \in \mathcal{D} \mid a \in T_i} q_i \geq \gamma))$$

Contribution: k-RFP: Example

- $k = 1$, $\alpha = 3$ and $\gamma = 3$
- $P = \{a, b, c, d\}$ and $Q = \{e, f, g, h\}$ **are the 1-RFPs**.
- Two clusters can be identified using P and Q $\{T_1, T_2, T_3, T_4\}$ and $\{T_5, T_6, T_7, T_8\}$.

Transactions	Items
T_1	a b c
T_2	a b d
T_3	a c d
T_4	b c d
T_5	f g h
T_6	e g h
T_7	e f h
T_8	e f g

Contribution: CCP-Based k-RFP using ILP Model M1

$$\texttt{Maximize} \quad \sum_{c \in C} v_c \cdot y_c$$

$$\texttt{Subject to} \quad (1) \quad \sum_{c \in C} a_{T_i,c} \cdot y_c = 1, \quad \forall\, T_i \in \mathcal{D}$$

$$(2) \quad \sum_{c \in C} y_c = \beta_0$$

$$y_c \in \{0,1\}, c \in C$$

$$i = 1, \ldots, m$$

Contribution: CCP-Based k-RFP using ILP Model M2

$$\texttt{Maximize} \quad \sum_{c \in C} v_c \cdot y_c$$

$$\texttt{Subject to} \quad (1) \quad \sum_{c \in C} a_{T_i,c} \cdot y_c \leq \theta, \quad \forall\, T_i \in \mathcal{D}$$

$$(2) \quad \sum_{c \in C} y_c = \beta_0$$

$$y_c \in \{0,1\}, c \in C$$

$$i = 1, \ldots, m$$

Experiments: k-RFP vs CCP using ILP Model M1

$\mathcal{D}$	α	k-RFP			Classical Closed pattern		
		#k-RFP	#Found sol	Average time (s)	#CCP	#Found sol	Average time (s)
Lymph	10%	3605378	**27**	3760.40	51862	8	**62.85**
	20%	759630	**27**	415.81	13934	2	**2.37**
	30%	202602	**27**	74.12	4910	1	**0.92**
	40%	60470	0	-	2058	0	-
Mushroom	10%	128962	**27**	930.68	3287	6	**34.21**
	20%	19712	**27**	74.59	817	1	**6.03**
	30%	4055	0	-	293	**1**	**3.06**
	40%	1135	0	-	107	0	-
Primary-Tumor	10%	256991	**27**	42.71	32183	7	**19.70**
	20%	76081	**27**	13.89	9891	2	**6**
	30%	30372	**27**	8.88	3614	1	**1.42**
	40%	14778	0	-	1382	0	-
Soybean	10%	69191	**27**	16.19	2907	6	**2.57**
	20%	11900	0	-	844	**2**	**0.53**
	30%	3383	0	-	380	0	-
	40%	1484	0	-	205	0	-
Tic-tac-toe	10%	4479	**28**	19.94	191	2	**0.17**
	20%	811	**28**	1.45	26	1	**0.10**
	30%	171	0	-	18	0	-
	40%	15	0	-	5	0	-
Vote	10%	280386	**28**	67.38	37399	3	**14.93**
	20%	34098	0	-	7227	0	-
	30%	6606	0	-	658	0	-
	40%	693	0	-	79	0	-
Zoo-1	10%	92711	**28**	4.71	3291	7	**0.32**
	20%	35081	**28**	2.02	1743	2	**0.23**
	30%	12614	**28**	0.86	818	1	**0.14**
	40%	3555	**28**	**0.28**	316	0	-

Experiments: CCA-k-RFP-M2 vs CCP-M2

$\mathcal{D}$	θ	CCA-k-RFP-M2		CCP-M2	
		#Found sol	Average time (s)	#Found sol	Average time (s)
Lymph	2	28	4156.28	28	**2536.88**
	3	**28**	3519.44	26	**351.28**
	4	**28**	6115.981	26	**623.40**
	5	**28**	3998.53	24	**218.67**
Mushroom	2	**28**	**355.77**	2	15244.37
	3	**28**	362.99	1	**195.29**
	4	**28**	355.79	1	**187.84**
	5	**28**	332.93	1	**196.36**
Primary-Tumor	2	28	**47.96**	28	198.12
	3	28	**48.54**	28	183.98
	4	28	**59.25**	28	159.136
	5	28	**52.73**	28	123.17
Soybean	2	28	**11.97**	28	137.062
	3	28	**11.72**	28	100.11
	4	28	**12.72**	28	80.66
	5	28	**10.68**	28	73.489
Tic-tac-toe	2	**28**	**27.01**	27	1693.54
	3	**28**	**22.46**	27	447.67
	4	**28**	**19.60**	25	227.98
	5	**28**	**16.24**	24	221.14
Vote	2	28	**77.77**	28	2716.59
	3	28	**103.17**	28	2660.54
	4	28	**531.33**	28	1436.24
	5	28	**240.68**	28	963.59
Zoo-1	2	28	5.12	28	**0.61**
	3	28	5.27	28	**0.53**
	4	28	5.25	28	**0.54**
	5	28	5.24	28	**0.53**

Experiments: CCA-k-RFP-M1 vs CCP-M1

$\mathcal{D}$	CCA-k-RFP-M1			CCP-M1		
	$k = 1$, $\beta = 30$, $\alpha = 10\%$			$\beta = 30$		
	#k-RFP	ICS	Time (s)	#CCP	ICS	Time (s)
Lymph	3605378	**3723.17**	3305.40	154220	277.93	**270.67**
Mushroom	128962	**5652568.44**	**932.69**	221524	-	-
Primary-	256991	**15984.25**	**44.11**	87230	1895.89	241.95
Soybean	69191	**40133.02**	**14.23**	31759	10795.60	87.50
Tic-tac-toe	4479	**47165.55**	**19.52**	42711	29278.23	1466.06
Vote	280386	**16992.90**	**68.75**	227031	8563.77	1268.51
Zoo-1	92711	**768.37**	4.74	4567	267.79	**0.422**

Experiments: CCA-k-RFP-M1 vs other methods

$\mathcal{D}$	CCA-k-RFP-M1		k-Means		BIRCH		SPECTRAL		Aggl-Clust	
	ICS	Time (s)	ICS	Time (s)	ICS	Time (s)	ICS	Time (s)	ICS	Time (s)
Lymph	**3723.17**	3305.40	252.80	0.45	257.89	0.04	452.48	0.429	250.86	**0.008**
Mushroom	**5652568.44**	932.69	883931.50	**3.17**	926729.75	15.90	1442391.09	100.98	1072855.32	11.32
Primary-Tumor	**15984.25**	44.11	1745.32	0.47	2030.70	0.06	5632.47	0.76	1792.95	**0.014**
Soybean	**40133.02**	14.23	5067.89	0.64	5433.03	0.182	17066	1.14	5040.35	**0.04**
Tic-tac-toe	**47165.55**	19.52	7751.78	0.89	7049.87	0.32	7353.16	2.07	7059.29	**0.08**
Vote	**16992.90**	68.75	2794.59	0.51	3189.58	0.12	14807.83	0.86	2533.51	**0.02**
Zoo-1	**768.37**	4.74	218.60	0.39	248.17	0.02	467.03	0.23	164.77	**0.005**

Experiments: CCA-k-RFP-M2 vs Neo-k-Means

$\mathcal{D}$	CCA-k-RFP-M2		Neo-k-Means	
	ICS	Time (s)	ICS	Time (s)
Lymph	**7470.67**	3288.04	461	**0.21**
Mushroom	**12259500.13**	332.13	1967290.99	**7.36**
Primary-Tumor	**36752.61**	50.74	2965.39	**0.16**
Soybean	**71741.53**	13.12	6344.39	**0.35**
Tic-tac-toe	**92792.95**	34.43	10917.74	**0.58**
Vote	**31761.66**	77.37	6083.92	**0.27**
Zoo-1	**1657.55**	5.01	285.84	**0.06**

Conclusion & future work

Conclusion & Perspectives

Conclusion

- New class of disjunctive itemsets and association rules is introduced.
- New symbolic encoding for disjunctive patterns is proposed.
- Sequential and parallel empirical evaluation over different real-world datasets is performed.

Perspectives

- Investigate other types of patterns (e.g., minimal non-redundant rules).
- Improve our model enumeration for (k, k')-based disjunctive association rules for $k > 0$ and $k' > 0$.