





	6.1 Introduction	
	Frequent Patterns	
Inductive bias Frequent patte	f association rules s are more interesting than infrequent ones	
Major challens Efficiently find	ng (all) frequent patterns	
Types of freque • Frequent item s • Generalized fre • Quantitative fre • Frequent sequent \rightarrow 7 Mining b	nt patterns s (boolean attributes) ent item <i>sets</i> (boolean attributes with concept hierarchies) ent <i>sets</i> (numerical attributes) <i>es</i> (sequence data))
→ /. Mining t	logical data	







	6.2 Basic	Associa	ation Rules			
		Example	2			
[TransactionID	Items				
	2000	A,B,C	. 500/			
	1000	A,C	minsup = 50%, minconf = 50%			
	4000	A,D	minconj = 50%			
	5000	B,E,F				
Support (A): 75% (A, C):	, (B), (C): 50%, (50%, (A, B), ((D), (E), (F): A, D), (B, C)	25%, , (B, E), (B, F), (E, F): 25%			
Association	rules					
$A \Rightarrow C$	$A \Rightarrow C$ (support = 50%, confidence = 66.6%)					
$C \Rightarrow A$	(support = 50%, con)	nfidence = 10	0%)			
	SFU, C	MPT 740, 03-3, N	artin Ester 285			



























Methods of Efficency Improvement

Partitioning of the database [Savasere, Omiecinski & Navathe 1995]

- Item set is only frequent if frequent in at least one partition
- Form memory-resident partitions of the database

more efficient on partitions, but expensive combination of intermediate results

Sampling [Toivonen 1996]

- Apply algorithm to sample to find frequent item sets
- Count support of these frequent item sets in the whole database
- Determine further candidates and support counting on the whole database

SFU, CMPT 740, 03-3, Martin Ester

298







	Body	Implies	Head	Supp (%)	Conf (%)	F	G	Н	1
1	cost(x) = '0.00~1000.00'	==>	revenue(x) = '0.00~500.00'	28.45	40.4				
2	cost(x) = '0.00~1000.00'	==>	revenue(x) = '500.00~1000.00'	20.46	29.05				
3	cost(x) = '0.00~1000.00'	==>	order_qty(x) = '0.00~100.00'	59.17	84.04				
1	cost(x) = 10.00~1000.00'	==>	revenue(x) = '1000.00~1500.00'	10.45	14.84				
5	cost(x) = '0.00~1000.00'	==>	region(x) = 'United States'	22.56	32.04				
6	cost(x) = '1000.00~2000.00'	==>	order_qty(x) = '0.00~100.00'	12.91	69.34				
7	order_gty(x) = '0.00~100.00'	==>	revenue(x) = '0.00~500.00'	28.45	34.54				
8	order gty(x) = '0.00~100.00'	==>	cost(x) = '1000.00~2000.00'	12.91	15.67				
9	order_qty(x) = '0.00~100.00'	==>	region(x) = 'United States'	25.9	31.45				
0	order_qty(x) = '0.00~100.00'	==>	cost(x) = '0.00~1000.00'	59.17	71.86				
1	order_qty(x) = '0.00~100.00'	==>	product_line(x) = 'Tents'	13.52	16.42				
2	order_qty(x) = '0.00~100.00'	==>	revenue(x) = '500.00~1000.00'	19.67	23.88				
3	product_line(x) = Tents'	==>	order_qty(x) = '0.00~100.00'	13.52	98.72				
4	region(x) = 'United States'	==>	order_qty(x) = '0.00~100.00'	25.9	81.94				
5	region(x) = 'United States'	==>	cost(x) = '0.00~1000.00'	22.56	71.39				
6	revenue(x) = '0.00~500.00'	==>	cost(x) = '0.00~1000.00'	28.45	100	DE	BMin	er Sv	stem
7	revenue(x) = '0.00~500.00'	==>	order_qty(x) = '0.00~100.00'	28.45	100				
8	revenue(x) = '1000.00~1500.00'	==>	cost(x) = '0.00~1000.00'	10.45	96.75	1 Ha	an et	al. 19	9961
9	revenue(x) = '500.00~1000.00'	==>	cost(x) = '0.00~1000.00'	20.46	100	L			1
0	revenue(x) = '500.00~1000.00'	==>	order_qty(x) = '0.00~100.00'	19.67	96.14				
1									
2									
3	cost(x) = 10.00~1000.00'	==>	revenue(x) = 10.00~500.00' AND order_qty(x) = 10.00~100.00'	28.45	40.4				
4	cost(x) = 10.00~1000.00'	==>	revenue(x) = '0.00~500.00' AND order_qty(x) = '0.00~100.00'	28.45	40.4				
5	cost(x) = 10.00~1000.00'	==>	revenue(x) = '500.00~1000.00' AND order_qty(x) = '0.00~100.00'	19.67	27.93				
6	cost(x) = 10.00~1000.00'	==>	revenue(x) = "500.00~1000.00" AND order_qty(x) = "0.00~100.00"	19.67	27.93				
27	cost(x) = '0.00~1000.00' AND order gtv(x) = '0.00~100.00'	==>	revenue(x) = '500.00~1000.00'	19.67	33.23				













6.2 0	Constraints for A	Associatio	on Rules
Types of Constraints	$\begin{array}{c} S \ \theta \ v, \theta \in \{=, \leq, \geq\} \\ \hline v \in S \\ S \supseteq V \\ S \subseteq V \\ S = V \\ \hline min(S) \leq v \\ min(S) \geq v \\ min(S) \geq v \\ max(S) \geq v \\ max(S) \leq v \\ max(S) \leq v \\ max(S) \leq v \\ max(S) \geq v \\ count(S) \leq v \\ count(S) \leq v \\ count(S) \leq v \\ count(S) \leq v \\ sum(S) \geq v \\ sum(S) \leq v \\ sum(S) \geq v \\ sum(S) = v \\ avg(S) \ \theta \ v, \theta \in \{=, \leq, \geq\} \\ (frequent constraint) \end{array}$	yes no no yes partly no yes partly yes no partly yes no partly yes no partly no partly no partly no yes no partly	anti-monotone?
	SFU, CMPT 740, 03	-3, Martin Ester	309











	E.	xample	
	TransaktionsID	Items	
	1	Anorak	
	2	Windcheater, Hiking boot	
	3	Anorak, Hiking boot	
	4	Walking-shoes	
	5	Walking-shoes	
	6	Windcheater	
Support of {Jacke Support of {Jacke ,Hiking-boots ⇒	ets}: 4 of 6 = 67% ets, Hiking boots}: 2 of Jackets": Support 3	6 = 33% 3%, Confidence 100%	





















6.3 Generalized Association Rules
Discussion
Fix support
• Same <i>minsup</i> value for all levels of the item taxonomy
+ Effiziency: pruning successors of infrequent item sets
- Reduced effectiveness
<i>minsup</i> too high \Rightarrow no low-level associations
minsup too low \Rightarrow too many high-level associations
Variable support
• Different <i>minsup</i> values for different levels of the item taxonomy
+ Good effectiveness
Find association rules at appropriate support level
- Inefficient: no pruning of successors of infrequent item sets
SFU, CMPT 740, 03-3, Martin Ester 326