MACHINE INTELLIGENCE

Dr. N MEHALA
Department of Computer Science and Engineering

MACHINE INTELLIGENCE

Module 4 [Unsupervised Learning]

Dr. N MEHALA
Department of Computer Science and Engineering

- **Two-step approach:**
  1. Frequent Itemset Generation
     - Generate all itemsets whose support $\geq$ minsup
  2. Rule Generation
     - Generate high confidence rules from each frequent itemset, where each rule is a binary partitioning of a frequent itemset

Mining Association Rules : Rule Generation

- Given a frequent itemset L, find all non-empty subsets $f \subset L$ such that $f \rightarrow L - f$ satisfies the minimum confidence requirement
  - If {A,B,C,D} is a frequent itemset, candidate rules:

ABC $\rightarrow$D,	ABD $\rightarrow$C,	ACD $\rightarrow$B,	BCD $\rightarrow$A,
A $\rightarrow$BCD,	B $\rightarrow$ACD,	C $\rightarrow$ABD,	D $\rightarrow$ABC
AB $\rightarrow$CD,	AC $\rightarrow$ BD,	AD $\rightarrow$ BC,	BC $\rightarrow$AD,
BD $\rightarrow$AC,	CD $\rightarrow$AB,		

- If $|L| = k$, then there are $2^k - 2$ candidate association rules (ignoring $L \rightarrow \varnothing$ and $\varnothing \rightarrow L$)

Rule Generation

- How to efficiently generate rules from frequent itemsets?
  - In general, confidence does not have an anti-monotone property

    c(ABC $\rightarrow$D) can be larger or smaller than c(AB $\rightarrow$D)

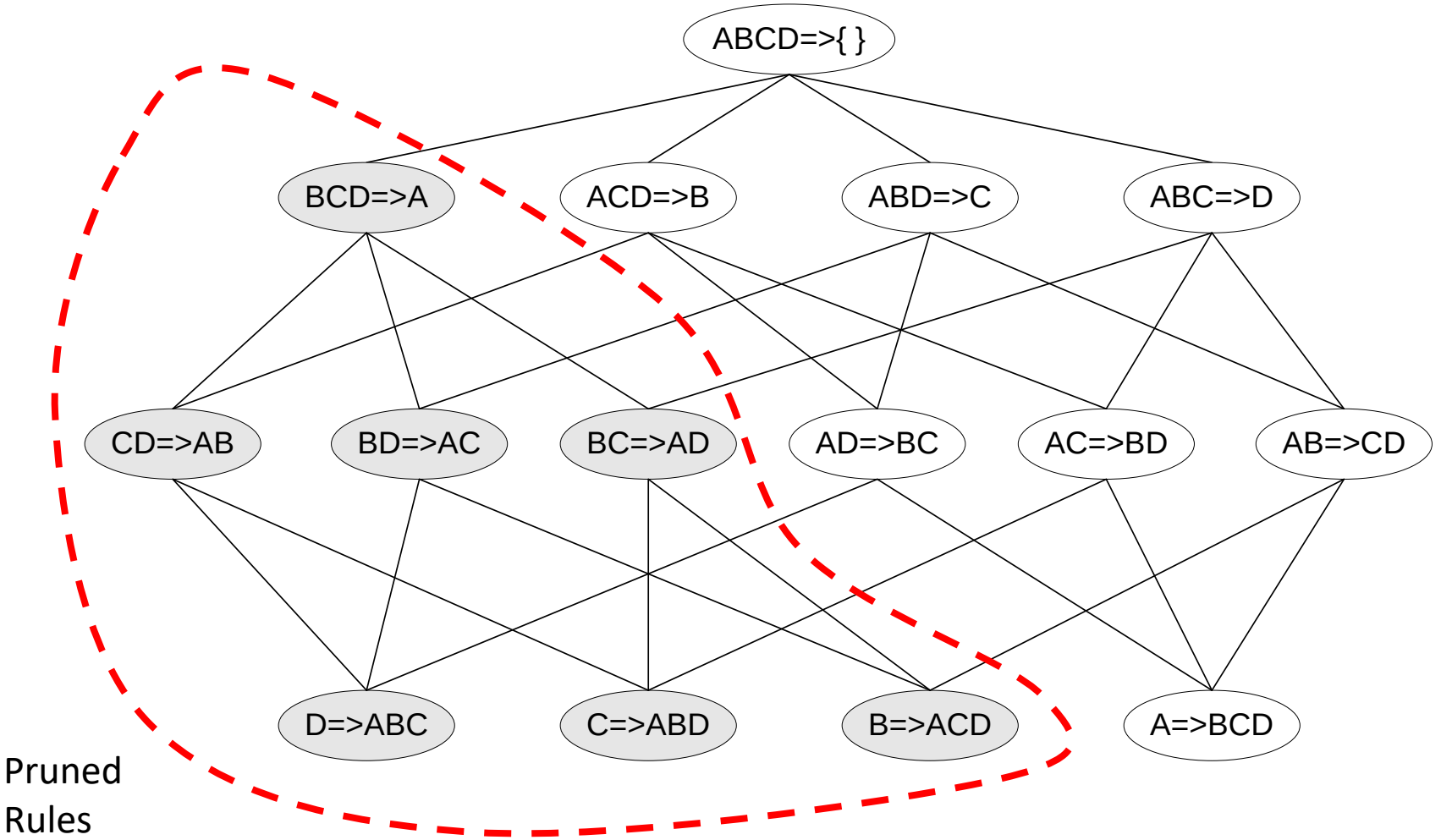
  - But confidence of rules generated from the same itemset has an anti-monotone property
  - e.g., L = {A,B,C,D}:

$$c(ABC \rightarrow D) \geq c(AB \rightarrow CD) \geq c(A \rightarrow BCD)$$

    - Confidence is anti-monotone w.r.t. number of items on the RHS of the rule

MACHINE INTELLIGENCE

Rule Generation for Apriori Algorithm



THANK YOU

Dr. N MEHALA

Department of Computer Science and Engineering

mehala@pes.edu

MACHINE INTELLIGENCE

Dr. N MEHALA
Department of Computer Science and Engineering

MACHINE INTELLIGENCE

Module 4 [Unsupervised Learning]

Dr. N MEHALA
Department of Computer Science and Engineering

MACHINE INTELLIGENCE

Association Rule Mining : Example

CONSIDER THE FOLLOWING DATASET WITH THE FOLLOWING INPUTS.

Minimum support (i.e coverage): 60%
Minimum confidence (i.e. accuracy): 80%

Trans_id	Itemlist
T1	{K, A, D, B}
T2	{D, A C, E, B}
T3	{C, A, B, E}
T4	{B, A, D}

- STEP 1: FREQUENT ITEMSET GENERATION
- STEP2: RULE GENERATION

Association Rule Mining : Example

STEP 1: FREQUENT ITEMSET GENERATION

MACHINE INTELLIGENCE

Association Rule Mining : Example

Minimum support (i.e coverage): 60%
Minimum confidence (i.e. accuracy): 80%

Trans_id	Itemlist
T1	{K, A, D, B}
T2	{D, A C, E, B}
T3	{C, A, B, E}
T4	{B, A, D}

1-ItemSet	Support Count
{A}	4
{B}	4
{C}	2
{D}	3
{E}	2
{K}	1

Association Rule Mining : Example

Now let's form the item sets containing 2 items.

1-ItemSet	Support Count
{A}	4
{B}	4
{D}	3

1-ItemSet	Support Count
{A}	4
{B}	4
{C}	2
{D}	3
{E}	2
{K}	1

2-ItemSet	Support Count
{A, B}	4
{B, D}	3
{A, D}	3

Association Rule Mining : Example

- STEP 3. The item sets containing 3 items. We only take the item sets from the previous phase whose support is 60% or more.

2-ItemSet	Support Count
{A, B}	4
{B, D}	3
{A, D}	3

3-ItemSet	Support Count
{A, B, D}	3

Association Rule Mining : Example

STEP2: Rule Generation

- Lets now form the rules and calculate their confidence (c). We only take the item sets from the previous phases whose support is 60% or more. ***Minimum confidence (i.e. accuracy): 80%***

MACHINE INTELLIGENCE

Association Rule Mining : Example

Trans_id	Itemlist
T1	{K, A, D, B}
T2	{D, A C, E, B}
T3	{C, A, B, E}
T4	{B, A, D}

FREQUENT ITEM SETS

{A, B }

{A, D }

{B, D }

{A,B,D}

BINARY PARTITIONING:

A -> B

B -> A

A -> D

D -> A

B -> D

D -> B

AB -> D

D -> AB

AD -> B

B - > AD

BD -> A

A -> BD

A -> B **P(B|A) = |B∩A| / |A| = 4/4, c: 100%**

B -> A **c: 100%**

A -> D **c: 75%**

MACHINE INTELLIGENCE

Association Rule Mining : Example

A -> B	P(B|A) = |B∩A| / |A| = 4/4, |c: 100%
B -> A	c: 100%
A -> D	c: 75%
D -> A	c: 100%
B -> D	c: 75%
D -> B	c: 100%
AB -> D	c: 75%
D -> AB	c: 100%
AD -> B	c: 100%
B - > AD	c: 75%
BD -> A	c: 100%
A -> BD	c: 75%

The rules with a confidence measure of 75% are pruned, and we are left with the following rule set:

A -> B
B -> A
D -> A
D -> B
D -> AB
AD-> B
DB-> A

Minimum confidence (i.e. accuracy): 80%

MACHINE INTELLIGENCE

Association Rule Mining : Example 2

PES UNIVERSITY ONLINE

MACHINE INTELLIGENCE

Association Rule Mining : Example2

Student	Grade	Income	Buys
CS	High	Low	Milk
CS	High	High	Bread
Math	Low	Low	Bread
CS	Medium	High	Milk
Math	Low	Low	Bread

MACHINE INTELLIGENCE

Association Rule Mining : Example2

CONVERTED DATA

Student	Grade	Income	Buys
CS	High	Low	Milk
CS	High	High	Bread
Math	Low	Low	Bread
CS	Medium	High	Milk
Math	Low	Low	Bread

Student = CS (I1)	Student =math (I2)	Grade = high (I3)	Grade =medium (I4)	Grade =low (I5)	Income =high (I6)	Income =low (I7)	Buys=milk (I8)	Buys =bread (I9)
+	-	+	-	-	-	+	+	-
+	-	+	-	-	+	-	-	+
-	+	-	-	+	-	+	-	+
+	-	-	+	-	+	-	+	-
-	+	-	-	+	-	+	-	+

MACHINE INTELLIGENCE

Summary

- Association Rule Mining Task
- Frequent Item Set Generation : Apriori Algorithm
- Factors Affecting Complexity

MACHINE INTELLIGENCE

Resources

- http://www2.ift.ulaval.ca/~chaib/IFT-4102-7025/public_html/Fichiers/Machine_Learning_in_Action.pdf
- http://wwwusers.cs.umn.edu/~kumar/dmbook/.
- ftp://ftp.aw.com/cseng/authors/tan
- http://web.ccsu.edu/datamining/resources.html

THANK YOU

Dr. N MEHALA

Department of Computer Science and Engineering

mehala@pes.edu