

## Overview

The goal of these introductory exercises is to review the basic concepts of association rule mining as well as on sequence mining.

## 1 Association Rule Mining: Basics

### 1.1 Confidence, Support and Interest

TID	Items
1	1 2 4 9
2	3 4 5 9 10
3	1 2 4 9
4	3 4 5 9 10
5	1 3 4 5
6	1 4 5 6
7	1 2 4 9
8	1 3 4 5 6 9 10
9	3 4 5 9 10
10	3 4 5 9 10

Answer the following questions:

1. What is the support of  $\{5, 9\}$ ?
2. What is the support of  $\{1, 3, 4, 5\}$ ?
3. What is the confidence of  $\{5\} \Rightarrow \{9\}$ ?
4. What is the confidence of  $\{3, 4, 5\} \Rightarrow \{1\}$ ?
5. What is the interest of  $\{5\} \Rightarrow \{9\}$ ?
6. What is the interest of  $\{3, 4, 5\} \Rightarrow \{1\}$ ?

For the itemset  $\{3, 4, 5, 9, 10\}$ , do the following:

1. Write one association rule based on this itemset.
2. What is its support?
3. What is its confidence?
4. Write another association rule based on this itemset.
5. What is its confidence?

### 1.2 Lift

The lift of an association rule  $A \Rightarrow B$  is defined as follows:  $lift(A \Rightarrow B) = \frac{conf(A \Rightarrow B)}{supp(B)}$

Use this definition to calculate the lift for the following problems.

**Situation 1** A school has 500 students in it. Out of these students, 300 take machine learning (ML), 200 take data mining (DM) and 50 take both classes. Calculate the lift of the rule  $ML \Rightarrow DM$ .

**Situation 2** A party has 1000 confirmed guests. Out of the guests, 600 drink Hoegaarden (H), 300 drink Kriek (K) and 200 drink both. Calculate the lift of the rule  $H \Rightarrow K$ .

## 2 Closed and Maximal Itemsets

In the following table, which itemsets are frequent, frequent closed and frequent maximal for the min.support threshold 10.

Itemset	support	frequent	closed	maximal
A	15			
B	20			
C	33			
D	25			
AB	15			
AC	12			
AD	15			
BC	18			
BD	5			
CD	25			
ABC	10			
ABD	2			
ACD	12			
BCD	3			
ABCD	1			

## 3 Apriori: Join and Prune

Given the following set of frequent 3-itemsets,  $F_3 = \{\{1, 3, 5\}, \{1, 5, 8\}, \{1, 3, 10\}, \{2, 3, 4\}, \{2, 3, 5\}, \{3, 8, 9\}, \{3, 8, 10\}\}$ :

1. Generate all legal candidates of the next level of Apriori's search.
2. Perform pruning on this candidate set.

## 4 Sequence mining

For the transactional database below:

1. Convert it to a customer-sequence database.
2. Mine the  $a$ -projected database and the  $c$ -projected database using FREESpan with  $minsup = 2$ .
3. Perform the same task using PREFIXSpan.

Date	CID	Items
01/12	1	a b
01/12	2	b h
01/12	3	i
01/12	4	b c
02/12	1	c
02/12	2	c
02/12	3	b g f
02/12	4	d
03/12	1	a d
03/12	2	a
03/12	3	i
03/12	4	c
04/12	1	b
04/12	2	g
04/12	3	j
04/12	4	f
05/12	1	g
05/12	2	f
05/12	3	c
05/12	4	a
06/12	1	f
06/12	2	a c
06/12	3	b g
06/12	4	b c
07/12	2	h
07/12	4	a
08/12	2	c d
08/12	4	d
09/12	4	c d