Data Science

Prof. M.-S. Chen & Prof. M.-L. Lo

TA Office hours

- Every Monday 15:00-16:00 @ 博理603
- Contact
  - email: ntuds2021@gmail.com

DS HW4 - Association rules

繳交方式

- Deadline: **2021/12/03 Fri. 23:59**
- 書寫方式:
  - 打字 or 手寫拍照
- 上傳 : ceiba
- 格式 : PDF

HW4 扣分規則

- 遲交
  - 此次作業0分
- 檔案格式錯誤
  - 此次作業 -10分

Problem 1 (50%)

1) Please use **Apriori** algorithm to mine frequent itemsets and then generate association rules for the database as the following

minimum support count: **3** transactions (**50%**)

minimum confidence: **66%**.

TID	Items
1	ACDE
2	CDE
3	AC
4	BE
5	ABDE
6	ABCDE

Problem 1 (50%)

1) Please use **Apriori** algorithm to mine frequent itemsets and then generate association rules for the database

   minimum support count: **3** transactions (**50%**)

   minimum confidence: **66%**.

   a) (10%) Do the frequent itemset generation. (as p.16 of 11/17 class notes)

   b) (10%) List ***all*** frequent itemsets. (ex. {A}, {B, C})

   c) (10%) Draw the lattice as in p.26 of 11/17 class notes in which indicate all transaction IDs in each itemset.

   d) (10%) Derive ***any* 5** association rules. (ex. {A}→{B})

   e) (10%) Show ***one*** association rule whose **lift** is less than 1.

Problem 2 (30%)

2. Given the following transaction data sets (**dark cells** indicate **presence** of an item in a transaction) and a support threshold of **20%**, answer the following questions

T1

TID \ Item	A	B	C	D	E	F	G	H	I	J
1	■	■	■	■	■					
2	■	■	■	■	■					
3	■	■	■	■	■					
4	■	■	■	■	■					
5	■	■	■	■	■					
6						■	■	■	■	■
7						■	■	■	■	■
8						■	■	■	■	■
9						■	■	■	■	■
10						■	■	■	■	■

T2

TID \ Item	A	B	C	D	E	F	G	H	I	J
1	■	■								
2	■	■								
3	■	■								
4	■	■		■	■					
5	■	■								
6	■	■					■		■	■
7	■	■					■			
8	■	■								
9	■	■								
10	■	■								

T3

TID \ Item	A	B	C	D	E	F	G	H	I	J
1	■	■	■							
2										
3	■	■	■							
4										
5		■	■	■						
6								■	■	■
7			■	■						
8										
9				■	■	■	■	■	■	■
10										

Problem 2 (30%)

2. Given the following transaction data sets (**dark cells** indicate **presence** of an item in a transaction) and a support threshold of **20%**, answer the following questions

(a) (10%) Which dataset will produce **the longest frequent itemset**?

(b) (10%) What is the *number* of **maximal frequent itemsets** for ***each*** dataset?

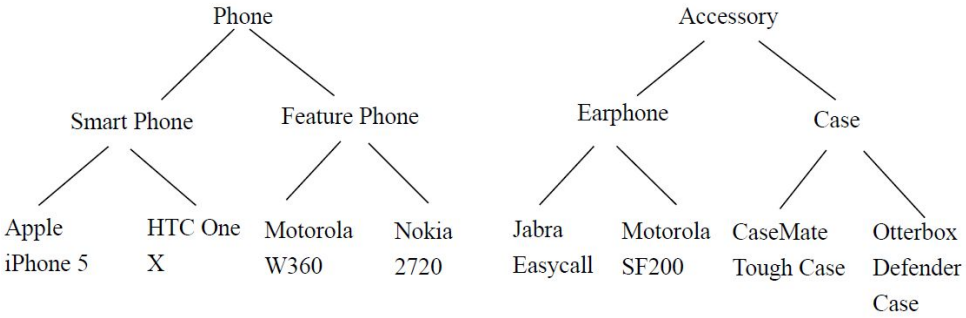
(c) (10%) What is the *number* of **closed frequent itemsets** for ***each*** dataset?

Problem 3 (20%)

3. Given the database below,

(a) (10%) What are the **confidence** and **support** of rule: “Earphone→Phone”?

(b) (10%) What are the **confidence** and **support** of rule: “Case→Earphone”?



Tx	Items bought
100	Apple iPhone 5, Jabra Easycall, CaseMate Tough Case
200	Motorola W360, Motorla SF200, CaseMate Tough Case
300	Nokia2720, Jabra Easycall
400	HTC One X, Jabra Easycall
500	Motorola W360, Motorla SF200, HTC One X
600	Apple iPhone 5, Otterbox Defender Case
700	HTC One X, Otterbox Defender Case
800	Nokia2720
900	Apple iPhone 5, CaseMate Tough Case
1000	Motorola W360, Motorola SF200