

Erasable Itemset Mining Using Multiple Maximum Thresholds Under Different Constraints

👤 Advisor: Dr. Tzung-Pei Hong

👤 Student: Yi-Chen Chang

113 / 04 / 10



Outline



Introduction



Related Work



Method



Experiment



Conclusion



0 1

PART ONE

Introduction

- ✓ Erasable Itemset mining



Introduction



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

- **Erasable itemset mining**
- **Product database**
- **Control loss**



Trouble



Funding decrease



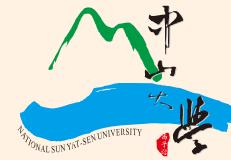
Insufficient storage space



Limited logistics



Introduction



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

- Erasable itemset
- Low loss



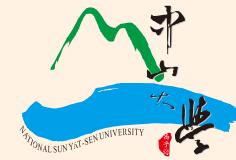
1 2

PART TWO

Related work



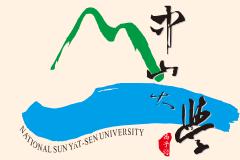
Erasable itemset mining



PID	Items	Profit
P1	ABC	200
P2	D	200
P3	AD	100
P4	AB	400
P5	BCE	100

META algorithm

- [Deng et al., 2009]
- Apriori-base method
- Downward closure



Erasable itemset mining

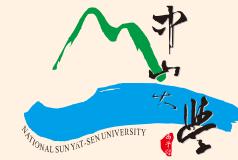
PID	Items	Profit
P1	ACD	200
P2	B	300
P3	AB	100
P4	AD	400
P5	CDE	400

Gain

Losses when certain items cannot purchase

$$\rightarrow \text{Gain (A)} = 200 + 100 + 400 = 700$$

Erasable itemset mining



PID	Items	Profit
P1	ABC	200
P2	D	200
P3	AD	100
P4	AB	400
P5	BCE	100

META algorithm

Gain \leq Maximum gain threshold (MGT)

- Total profit = $200 + 200 + 100 + 400 + 100 = 1000$
- Threshold = 0.4
- Maximum gain threshold = $1000 * 0.4 = 400$



Erasable itemset mining

META algorithm

PID	Items	Profit
P1	ABC	200
P2	D	200
P3	AD	100
P4	AB	400
P5	BCE	100

MGT = 400



Candidate 1-itemsets	
A	700
B	700
C	300
D	300
E	100

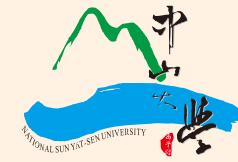


Erasable 1-itemsets	
C	300
D	300
E	100

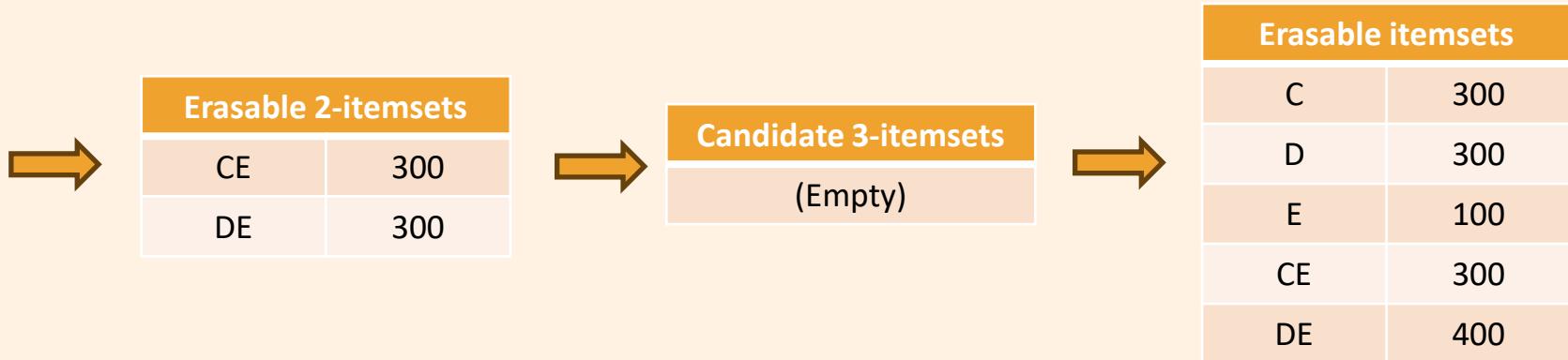
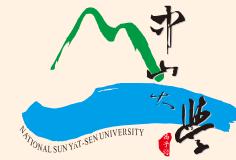


Candidate 2-itemsets	
CD	600
CE	300
DE	400

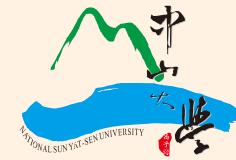
Downward
closure



Erasable itemset mining

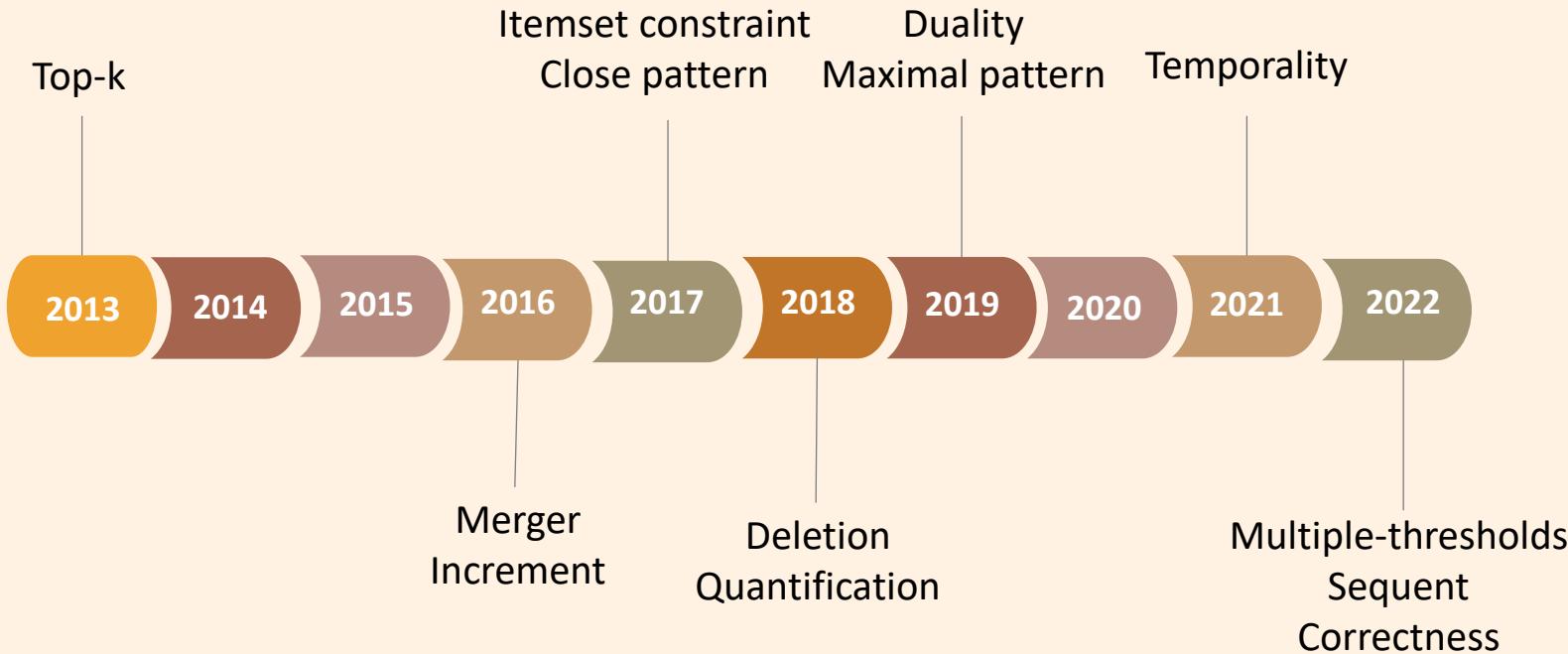
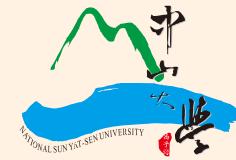


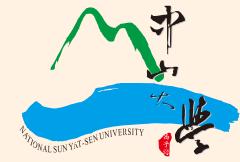
Performance



Algorithm	Year	Author
VME	2010	Deng et al.
MERIT	2012	Deng et al.
MERIT+	2013	Le et al.
dMERIT+	2013	Le et al.
MEI	2014	Le and Bo
BERM	2021	Hong and Huang et al.

Various themes



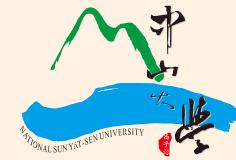


Multiple-threshold itemset mining

- Single threshold → the same standard for each item
- Considering other attributes of the item:
 1. Volume and weight
 2. Quantity
 3. Storage method

Item	A	B
λ	0.4	0.2





Problem of multiple-thresholds

Itemset

$$\lambda(A) = 0.4$$

One item



$$\lambda(B) = 0.2$$

More than one item



Item	A	B
λ	0.4	0.2



Constraints of MGT

Item	A	B
λ	0.4	0.2



Constraint 1: minimum constraint

$$\lambda_{\min}(AB) = \min (0.4, 0.2) = 0.2$$

Constraint 2: maximum constraint

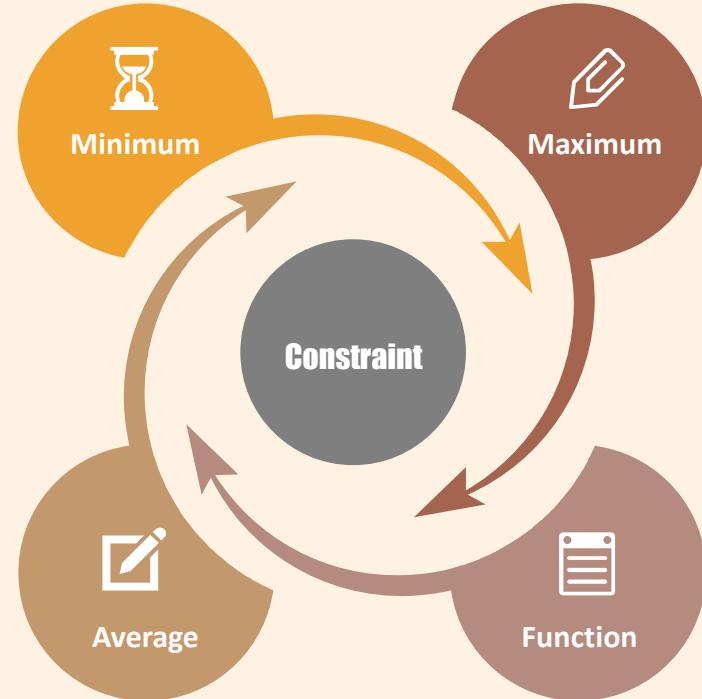
$$\lambda_{\max}(AB) = \max (0.4, 0.2) = 0.4$$

Constraint 3: average constraint

$$\lambda_{\text{avg}}(AB) = (0.4+0.2) / 2 = 0.3$$

Constraint 4: function constraint

$$\lambda_{\text{fun}}(AB) = \text{function} (0.4, 0.2)$$





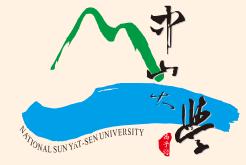
1 3

PART THREE

Method

- ✓ Minimum Constraint
- ✓ Maximum Constraint
- ✓ Average Constraint
- ✓ Function Constraint





01

Minimum Constraint



Multiple-threshold erasable itemset mining



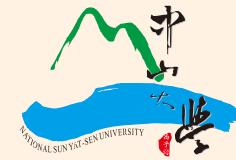
PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

- **Minimum constraint**
- **Downward closure**

Item	A	B	C	D	E	F	G
λ	0.6	0.1	0.7	0.9	0.5	0.8	0.4



Step 1



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Scan database

(1) Total profit:

$$200+300+100+200+400+500+100+200 = 2000$$

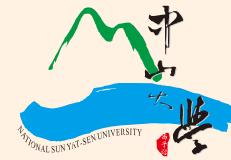
(2) Gain of each item:

Item	A	B	C	D	E	F	G
Gain	700	600	1100	800	1300	1000	1300

Item	A	B	C	D	E	F	G
λ	0.6	0.1	0.7	0.9	0.5	0.8	0.4



Step 2



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Calculate MGT of each item

Total profit = 2000

$$\text{MGT (A)} = 2000 * 0.6 = 1200$$

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800

Item	A	B	C	D	E	F	G
λ	0.6	0.1	0.7	0.9	0.5	0.8	0.4



Step 3

PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Verify candidate 1-itemsets

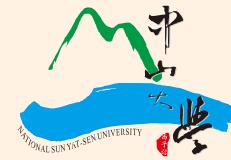
$\text{Gain} \leq \text{MGT}$

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800
Gain	700	600	1100	800	1300	1000	1300

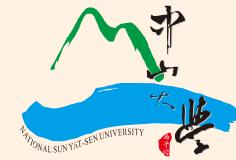


Erasable 1-itemsets			
A	C	D	F

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 4



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Generate candidate 2-itemsets

Erasable 1-itemsets

A	C	D	F
---	---	---	---

↓

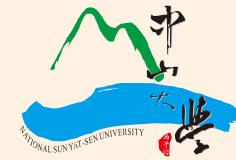
Candidate 2-itemsets

AC	AD	AF
CD	CF	DF

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 5



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Scan database to calculate gain

Item	AC	AD	AF	CD	CF	DF
Gain	700	600	1100	800	1300	1000

Calculate MGT_{min}

$$MGT_{\min}(AC) = \min(MGT(A), MGT(C))$$

$$= \min(1200, 1400) = 1200$$

Item	AC	AD	AF	CD	CF	DF
MGT _{min}	1200	1200	1200	1400	1400	1600

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 6



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Verify candidate 2-itemsets

$\text{Gain} \leq \text{MGT}_{\min}$

Itemset	AC	AD	AF	CD	CF	DF
MGT_{\min}	1200	1200	1200	1400	1400	1600
Gain	1200	1100	1600	1200	1900	1000



Erasable 2-itemsets

AC	AD	CD
----	----	----

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 7



PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Generate candidate 3-itemsets

Erasable 2-itemsets		
AC	AD	CD



Candidate 3-itemsets	
ACD	

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 8

PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Verify candidate 3-itemsets

$\text{Gain} \leq \text{MGT}_{\min}$

Itemset	ACD
MGT_{\min}	1200
Gain	1300

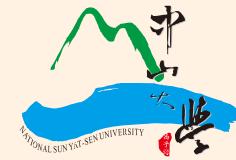


Erasable 3-itemsets
(Empty)

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



Step 9



Output all erasable itemsets

PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Erasable itemsets
A
C
D
F
AC
AD
CD

Item	A	B	C	D	E	F	G
MGT	1200	200	1400	1800	1000	1600	800



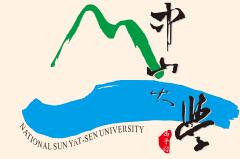
01

Minimum
Constraint

02

Maximum
Constraint

Multiple-threshold erasable itemset mining



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

- Maximum constraint
- No downward closure property
- Sorted closure property

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7

Step 1

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Sort the thresholds in descending order

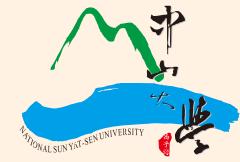
Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7



Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Step 2



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Scan database

(1) Total profit:

$$200+300+100+200+400+500+100+200 = 2000$$

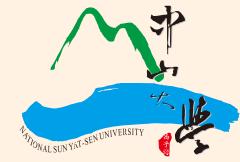
(2) Gain of each item:

Item	G	D	F	C	A	E	B
Gain	1500	600	400	500	500	900	1300

Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Step 3



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Calculate MGT of each item

Total profit = 2000

Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 4

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Find the first erasable itemset

$\text{Gain} \leq \text{MGT}$



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200
Gain	1500	600	400	500	500	900	1300

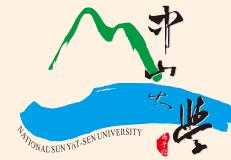


Candidate 1-itemsets

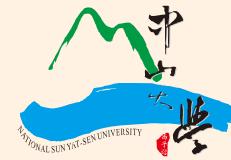
D

$$\text{MGT}(D) = 1200$$

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 5



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Verify itemset sorted after {D}

$\text{Gain} \leq \text{MGT}(D)$

Item	D	F	C	A	E	B
MGT(D)	1200	1200	1200	1200	1200	1200
Gain	600	400	500	500	900	1300



Candidate 1-itemsets

D	F	C	A	E
---	---	---	---	---

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 6



Verify candidate 1-itemsets

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900

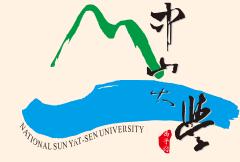


Erasable 1-itemsets				
D	F	C	A	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 2-itemset

Candidate 1-itemsets				
D	F	C	A	E

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200

Generate candidate 2-itemset



Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900

$\text{Gain}(F) \leq \text{MGT}(D)$



Candidate 2-itemsets

DF



Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 2-itemset



Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900



Candidate 2-itemsets			
DF	DC	DA	DE

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 7



Generate candidate 2-itemset

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900



$\text{Gain}(E) > \text{MGT}(A)$

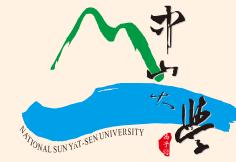
Candidate 2-itemsets

AE



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200

Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 2-itemsets

Candidate 1-itemset

D	F	C	A	E
---	---	---	---	---

↓

Candidate 2-itemsets

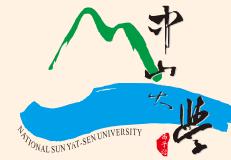
DF	DC	DA	DE	FC
FA	FE	CA	CE	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 8

Verify candidate 2-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	Gain
DF	1200	900
DC	1200	800
DA	1200	1100
DE	1200	1100
FC	1200	900
FA	1200	600
FE	1200	1200
CA	1000	800
CE	1000	1100

Gain \leq MGT_{max}

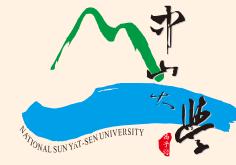


Erasable 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 9



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 3-itemset

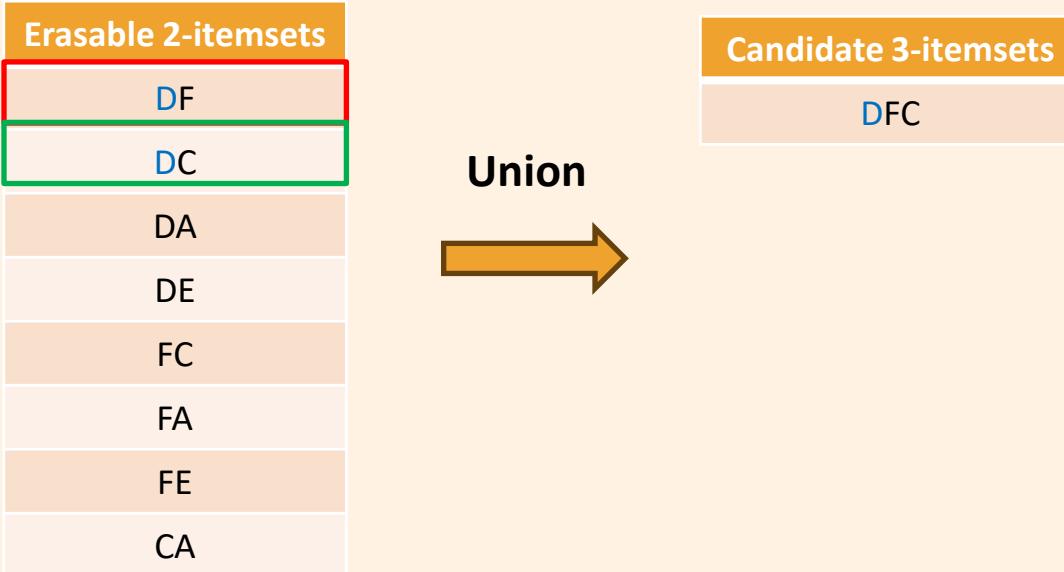
Erasable 2-itemsets			
DF	DC	DA	DE
FC	FA	FE	CA

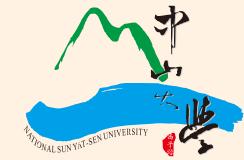
Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



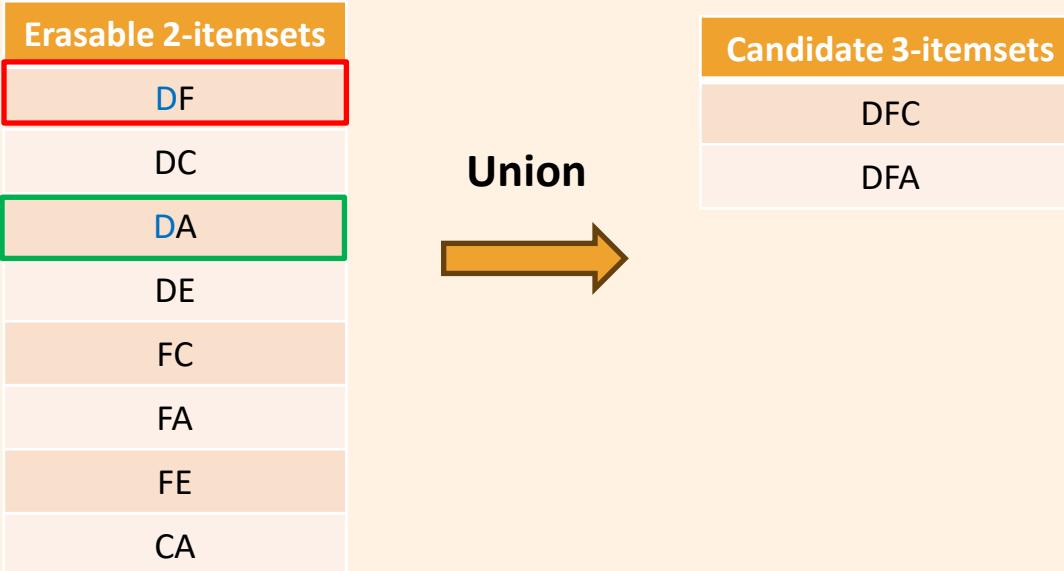


Step 9-1 Union of any two itemsets with the same front (k-1) items.





Step 9-1 Union of any two itemsets with the same front (k-1) items.

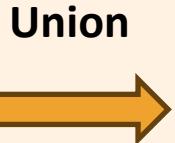


Step 9-1

Union of any two itemsets with the same front ($k-1$) items

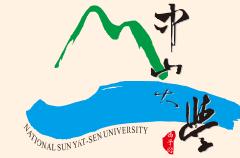


Erasable 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA



Candidate 3-itemsets
DFC
DFA
DFE
DCA
DCE
DAE
FCA
FCE





Step 9-2 Prune candidate itemsets

Case 1: $MGT(item_1) = MGT(item_2)$

Candidate 3-itemsets
DFC
DFA
DFE
DCA
DCE
DAE
FCA
FCE

$MGT(D)=MGT(F)$

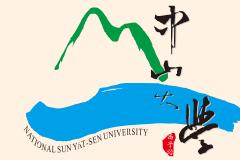


All (k-1)-subsets

2-subset
DF
DC
FC

Erasable 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 9-2 Prune itemsets

Case 2: $MGT(item_1) \neq MGT(item_2)$

Candidate 3-itemsets
DFC
DFA
DFE
DCA
DCE
DAE
FCA
FCE

(k-1)-subsets contain item₁

$MGT(F) \neq MGT(C)$

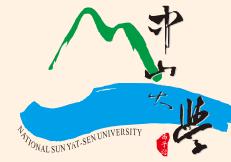


2-subset
FC
FE
CE X

Erasable 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200

Step 9



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 3-itemset

Erasable 2-itemsets			
DF	DC	DA	DE
FC	FA	FE	CA



Candidate 3-itemsets				
DFC	DFA	DFE	DCA	DCE
DAE	FCA	FCE	FAE	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 10

Verify candidate 3-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	Gain
DFC	1200	1100
DFA	1200	1100
DFE	1200	1400
DCA	1200	1100
DCE	1200	1300
DAE	1200	1600
FCA	1200	900
FCE	1200	1400
FAE	1200	1400

Gain \leq MGT_{max}

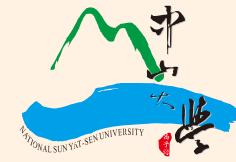


Erasable 3-itemsets
DFC
DFA
DCA
FCA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 11



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 4-itemsets

Erasable 3-itemsets			
DFC	DFA	DCA	FCA



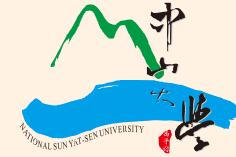
Candidate 4-itemsets	
DFCA	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 12

Verify candidate 4-itemset



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	Gain
DFCA	1200	1100

$\text{Gain} \leq \text{MGT}_{\text{max}}$

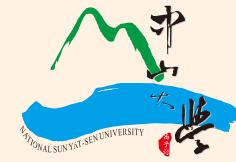


Erasable 4-itemset
DFCA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 13



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 5-itemsets

Erasable 4-itemsets
DFCA



Candidate 5-itemsets
(Empty)

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 14



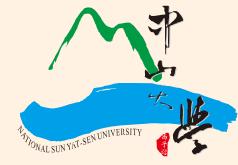
Output all erasable itemsets

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200

Erasable itemsets	
D	FA
F	FE
C	CA
A	DFC
DF	DFA
DC	DCA
DA	FCA
DE	DFCA
FC	





01

Minimum
Constraint

02

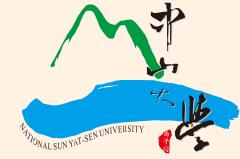
Maximum
Constraint

03

Average
Constraint



Multiple-threshold erasable itemset mining

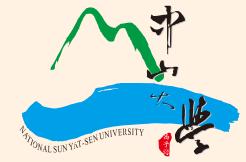


PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

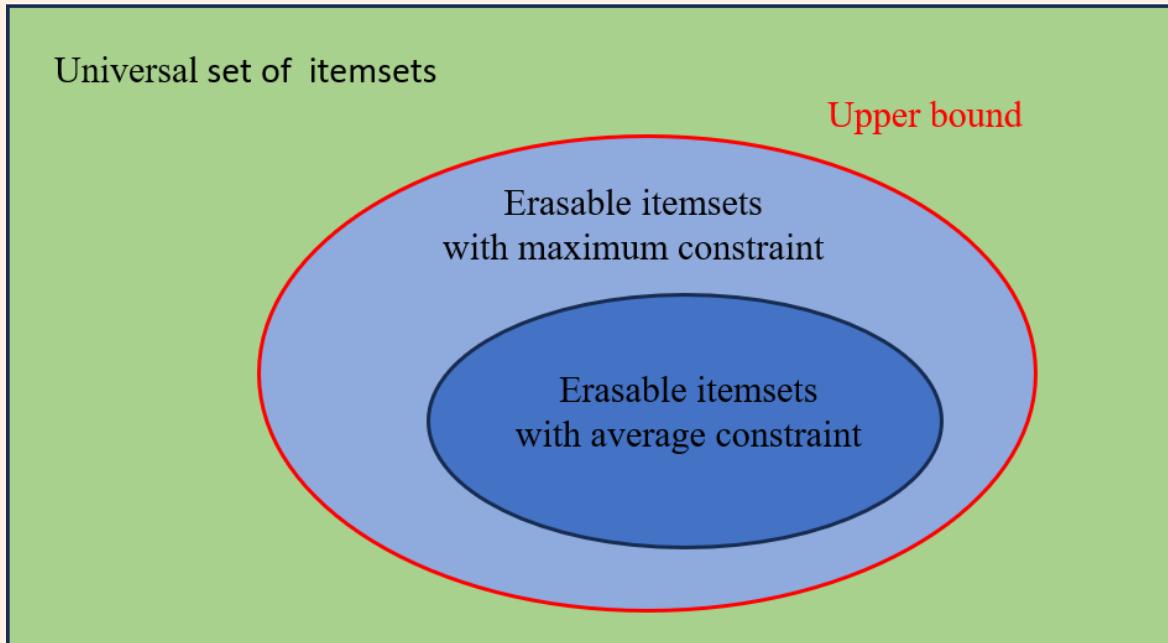
- Average constraint
- No downward closure property
- Upper bound strategy

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7





Upper bound strategy



Step 1

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Sort the thresholds in descending order

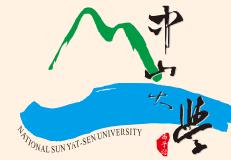
Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7



Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Step 2



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Scan database

(1) Total profit:

$$200+300+100+200+400+500+100+200 = 2000$$

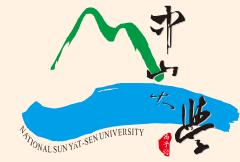
(2) Gain of each item:

Item	G	D	F	C	A	E	B
Gain	1500	600	400	500	500	900	1300

Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Step 3



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Calculate MGT of each item

Total profit = 2000

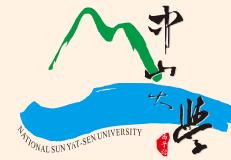
Item	G	D	F	C	A	E	B
λ	0.7	0.6	0.6	0.5	0.4	0.3	0.1



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 4



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Find the first erasable itemset

→

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200
Gain	1500	600	400	500	500	900	1300



Candidate 1-itemsets

D

$$MGT(D) = 1200$$

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 5



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Verify itemset sorted after {D}

Item	D	F	C	A	E	B
MGT(D)	1200	1200	1200	1200	1200	1200
Gain	600	400	500	500	900	1300



Candidate 1-itemsets				
D	F	C	A	E

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 6



Verify candidate 1-itemsets

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900

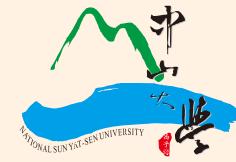


Erasable 1-itemsets				
D	F	C	A	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200

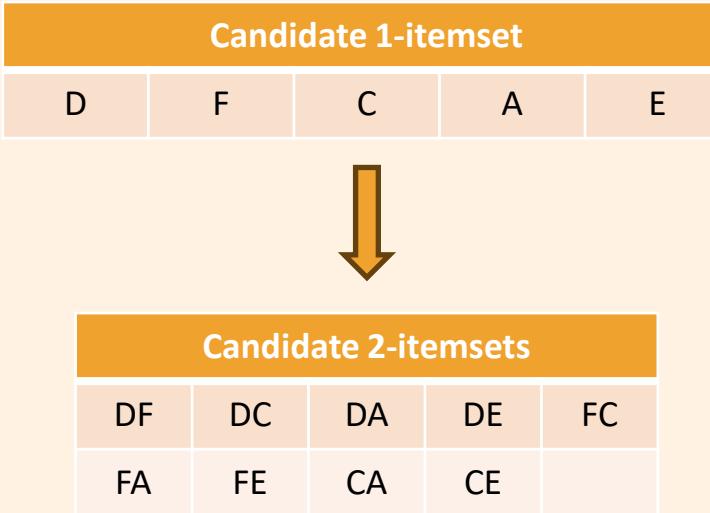


Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 2-itemsets



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 8

Verify candidate 2-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	MGT _{avg}	Gain
DF	1200	1200	900
DC	1200	1100	800
DA	1200	1000	1100
DE	1200	900	1100
FC	1200	1100	900
FA	1200	1000	600
FE	1200	900	1200
CA	1000	900	800
CE	1000	800	1100

Case 1:

$\text{Gain} \leq \text{MGT}_{\text{max}}$ & $\text{Gain} \leq \text{MGT}_{\text{avg}}$

Add to the erasable itemsets



Erasable 2-itemsets
DF
DC
FC
FA
CA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 8

Verify candidate 2-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	MGT _{avg}	Gain
DF	1200	1200	900
DC	1200	1100	800
DA	1200	1000	1100
DE	1200	900	1100
FC	1200	1100	900
FA	1200	1000	600
FE	1200	900	1200
CA	1000	900	800
CE	1000	800	1100

Case 2:

$\text{Gain} \leq \text{MGT}_{\text{max}}$ & $\text{Gain} > \text{MGT}_{\text{avg}}$

Do nothing

Ex: {D, A}, {D, E}, {F, E}

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 8

Verify candidate 2-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	MGT _{avg}	Gain
DF	1200	1200	900
DC	1200	1100	800
DA	1200	1000	1100
DE	1200	900	1100
FC	1200	1100	900
FA	1200	1000	600
FE	1200	900	1200
CA	1000	900	800
CE	1000	800	1100

Case 3:

Gain > MGT_{max}

Prune the candidate itemset

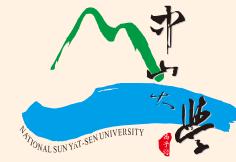
Ex: {C, E}

Candidate 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 9



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 3-itemset

Candidate 2-itemsets			
DF	DC	DA	DE
FC	FA	FE	CA



Candidate 3-itemsets				
DFC	DFA	DFE	DCA	DCE
DAE	FCA	FCE	FAE	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 10

Verify candidate 3-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	MGT _{avg}	Gain
DFC	1200	1133.3	1100
DFA	1200	1066.6	1100
DFE	1200	1000	1400
DCA	1200	1000	1100
DCE	1200	933.3	1300
DAE	1200	866.6	1600
FCA	1200	1000	900
FCE	1200	933.3	1400
FAE	1200	866.6	1400

Candidate 3-itemsets
DFC
DFA
DCA
FCA



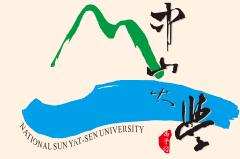
Erasable 3-itemsets
DFC
FCA



Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 11



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 4-itemsets

Candidate 3-itemsets			
DFC	DFA	DCA	FCA



Candidate 4-itemsets	
DFCA	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 12

Verify candidate 4-itemset



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _{max}	Gain
DFCA	1200	1100

Candidate 4-itemset

DFCA



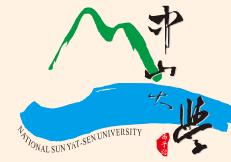
Erasable 4-itemset

(Empty)

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 13



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 5-itemsets

Candidate 4-itemset
DFCA

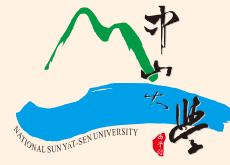


Candidate 5-itemset
(Empty)

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200



Step 14



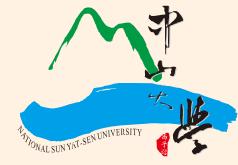
Output all erasable itemsets

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Erasable itemsets	
D	FC
F	FA
C	CA
A	DFC
DF	FCA
DC	

Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200





01

Minimum
Constraint

02

Maximum
Constraint

03

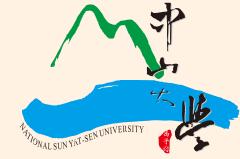
Average
Constraint

04

Function
Constraint



Multiple-threshold erasable itemset mining



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

- Function constraint
- No downward closure property
- Upper bound strategy
- Extend from average constraint

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7

Idea

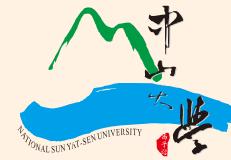


- $0 \leq \lambda \leq 1$
- Function : sum of λ
- $\lambda_{\text{sum}}(\text{ABD}) = 0.4 + 0.1 + 0.6 = 1.1 > 1$

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7



Idea



- Highest threshold $\lambda_h = 0.7$
- $\lambda_{\text{sum}}(\text{ABD}) = 1.1 > \lambda_h \rightarrow 0.7$
- $\lambda_{\text{sum}}(\text{AB}) = 0.5 < \lambda_h \rightarrow 0.5$

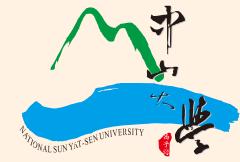
Downward closure

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7

$$f_{\text{fun}}(A) = \begin{cases} \lambda_h & , f(A) > \lambda_h \\ f(A) & , 0 \leq f(A) \leq \lambda_h \\ 0 & , f(A) < 0 \end{cases}$$



Step 1



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Scan database

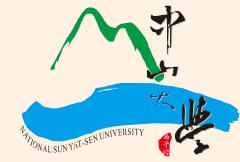
(1) Total profit:

$$200+300+100+200+400+500+100+200 = 2000$$

(2) Gain of each item:

Item	A	B	C	D	E	F	G
Gain	500	1300	500	600	900	400	1500

Step 2



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Calculate MGT of each item

Total profit = 2000, $\lambda_h = 0.6$

Item	A	B	C	D	E	F	G
λ	0.4	0.1	0.5	0.6	0.3	0.6	0.7



Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400

$$MGT_h = 2000 * 0.6 = 1200$$



Step 3



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Verify itemset with MGT_h

$Gain \leq MGT_h$

Item	A	B	C	D	E	F	G
MGT_h	1200	1200	1200	1200	1200	1200	1200
Gain	500	1300	500	600	900	400	1500



Candidate 1-itemsets

A	C	D	E	F
---	---	---	---	---

Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400



Step 4



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Verify candidate 1-itemsets

$\text{Gain} \leq \text{MGT}$

Item	A	C	D	E	F
MGT	800	1000	1200	600	1200
Gain	500	500	600	900	400

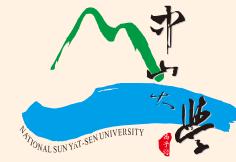


Erasable 1-itemsets			
A	C	D	F

Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400

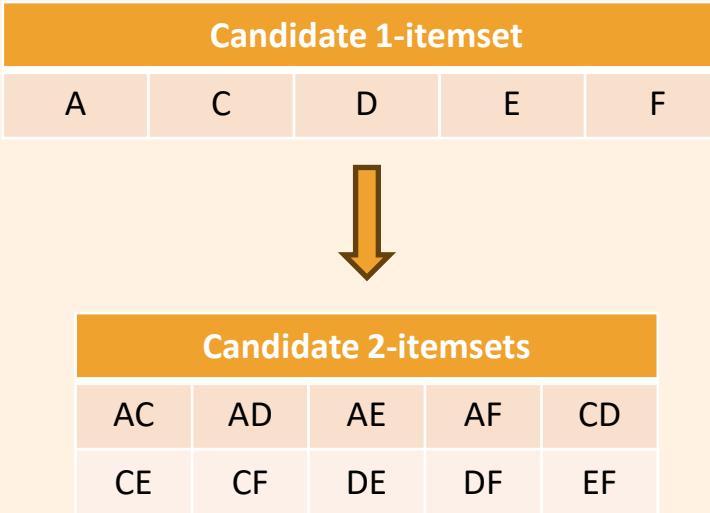


Step 5



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 2-itemsets



Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400



Step 6

Verify candidate 2-itemsets

PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _h	MGT _{avg}	Gain
AC	1200	900	800
AD	1200	1000	1100
AE	1200	700	1400
AF	1200	1000	600
CD	1200	1100	800
CE	1200	800	1100
CF	1200	1100	900
DE	1200	900	1100
DF	1200	1200	900
EF	1200	900	1200

Candidate 2-itemsets

AC

AD

AF

CD

CE

CF

DE

DF

EF

Erasable 2-itemsets

AC

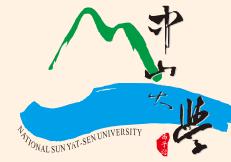
AF

CD

CF

DF

Step 7



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Generate candidate 3-itemset

Candidate 2-itemsets				
AC	AD	AF	CD	CE
CF	DE	DF	EF	



Candidate 3-itemsets				
ACD	ACF	ADF	CDE	CDF
CEF	DEF			

Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400



Step 8

Verify candidate 3-itemsets



PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Itemset	MGT _h	MGT _{avg}	Gain
ACD	1200	1000	1100
ACF	1200	1000	900
ADF	1200	1066.6	1100
CDE	1200	933.3	1300
CDF	1200	1133.3	1100
CEF	1200	933.3	1300
DFG	1200	1000	1400

Candidate 3-itemsets
ACD
ACF
ADF
CDF



Erasable 3-itemsets
ACF
CDF

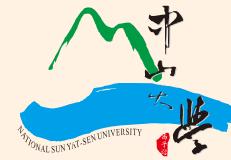


Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400



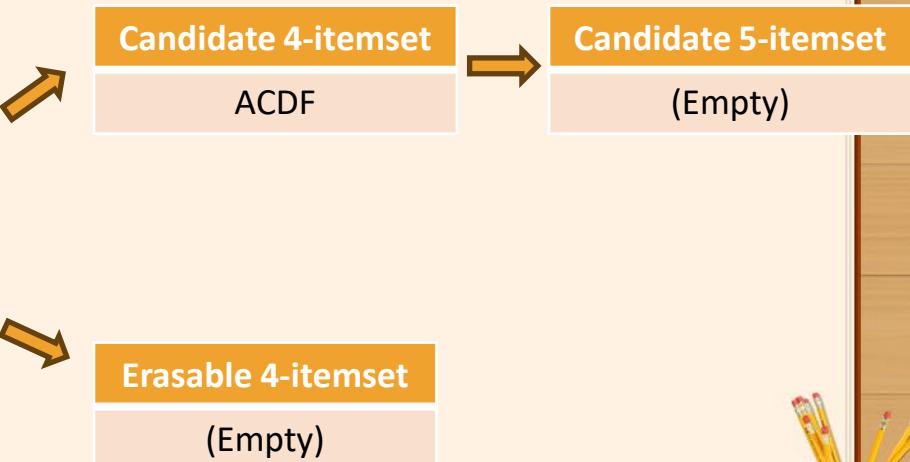
Step 9

Verify candidate 3-itemsets



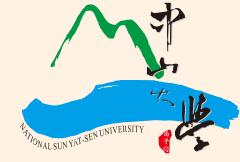
PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Candidate 3-itemsets
ACD
ACF
ADF
CDF



Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400

Step 10



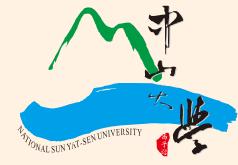
PID	Items	Profit
P1	ABCG	200
P2	CDE	300
P3	AF	100
P4	AFG	200
P5	BG	400
P6	BEG	500
P7	DEF	100
P8	BDG	200

Output all erasable itemsets

Erasable itemsets	
A	CD
C	CF
D	DF
F	ACF
AC	CDF
AF	

Item	A	B	C	D	E	F	G
MGT	800	200	1000	1200	600	1200	1400





01

Minimum
Constraint

02

Maximum
Constraint

03

Average
Constraint

04

Function
constraint

05

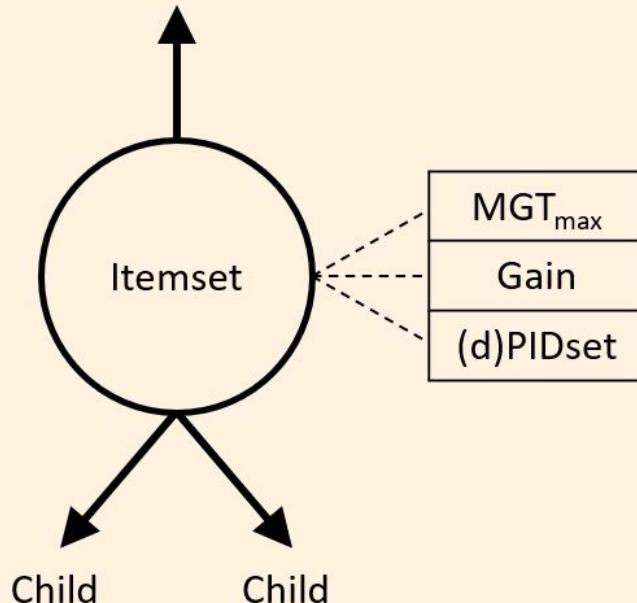
Efficient
maximum
constraint



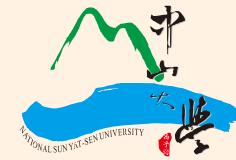
Node structure



Additional information :



PIDset



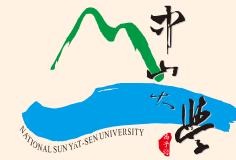
PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

PIDset (A)= {1, 5, 6}

PIDset (B)= {1, 2, 4, 5}



New method for calculating gain



PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Itemset {A, B}

$$\begin{aligned}d\text{PIDset}(AB) &= \text{PIDset}(B) - \text{PIDset}(A) \\&= \{1, 2, 4, 5\} - \{1, 5, 6\} \\&= \{2, 4\}\end{aligned}$$

$$\begin{aligned}\text{Gain}(AB) &= \text{Gain}(A) + (\text{P2.profit} + \text{P4.profit}) \\&= 400 + (200 + 300) \\&= 900\end{aligned}$$

Item	A	B
Gain	400	800
PIDset	1, 5, 6	1, 2, 4, 5



Step 1

PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Sort the thresholds in descending order

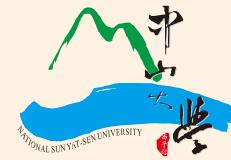
Item	A	B	C	D	E	F
λ	0.5	0.3	0.4	0.4	0.6	0.2



Item	E	A	C	D	B	F
λ	0.6	0.5	0.4	0.4	0.3	0.3



Step 2



PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Scan database

(1) Total profit:

$$200+200+100+300+100+100 = 1000$$

(2) Gain and PIDset of each item:

Item	E	A	C	D	B	F
Gain	700	400	400	100	600	300
PIDset	2, 5, 6	1, 5, 6	1, 3, 6	3	1, 2, 4, 5	4

Item	E	A	C	D	B	F
λ	0.7	0.6	0.6	0.5	0.4	0.3



Step 2

PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Item	E	A	C	D	B	F
λ	0.7	0.6	0.6	0.5	0.4	0.3

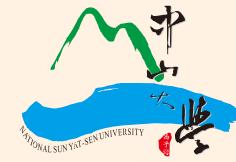
Scan database

(3) The profit hash table:

Key	Value
P1	200
P2	200
P3	100
P4	300
P5	100
P6	100



Step 3



PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Calculate MGT of each item

Total profit = 1000

Item	E	A	C	D	B	F
λ	0.6	0.5	0.4	0.4	0.3	0.3



Item	E	A	C	D	B	F
MGT	600	500	400	400	300	300



Step 4

PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Find the first erasable itemset

Gain \leq MGT



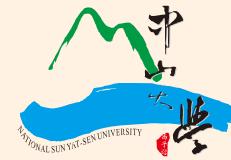
Item	E	A	C	D	B	F
MGT	600	500	400	400	300	300
Gain	700	400	400	100	600	300



Candidate 1-itemsets

A

$$MGT(A) = 500$$



Step 5

PID	Items	Profit
P1	ABC	200
P2	BE	200
P3	CD	100
P4	BEF	300
P5	ABE	100
P6	ACE	100

Verify itemset sorted after {A}

$\text{Gain} \leq \text{MGT}(A)$ 

Item	A	C	D	B	F
MGT(A)	500	500	500	500	500
Gain	400	400	100	600	300



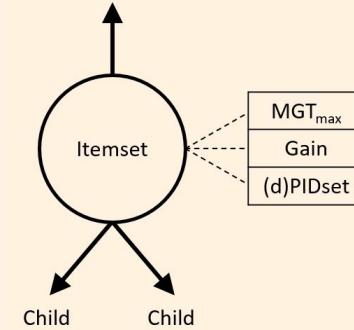
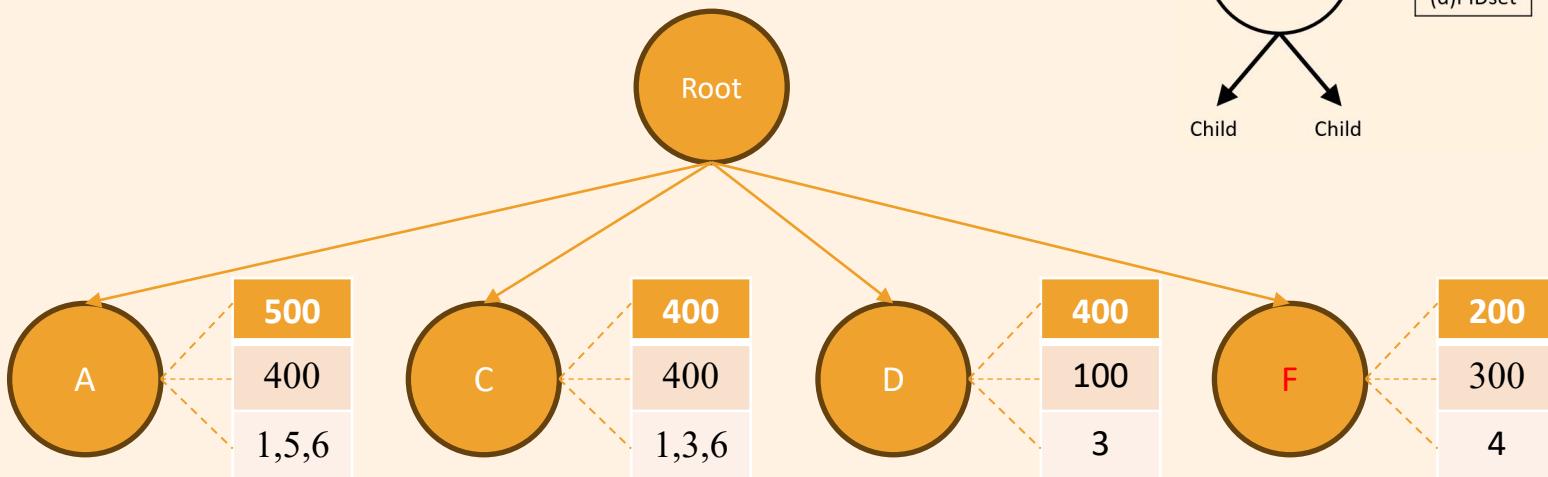
Candidate 1-itemsets			
A	C	D	F



Step 6 Insert candidate 1-itemsets

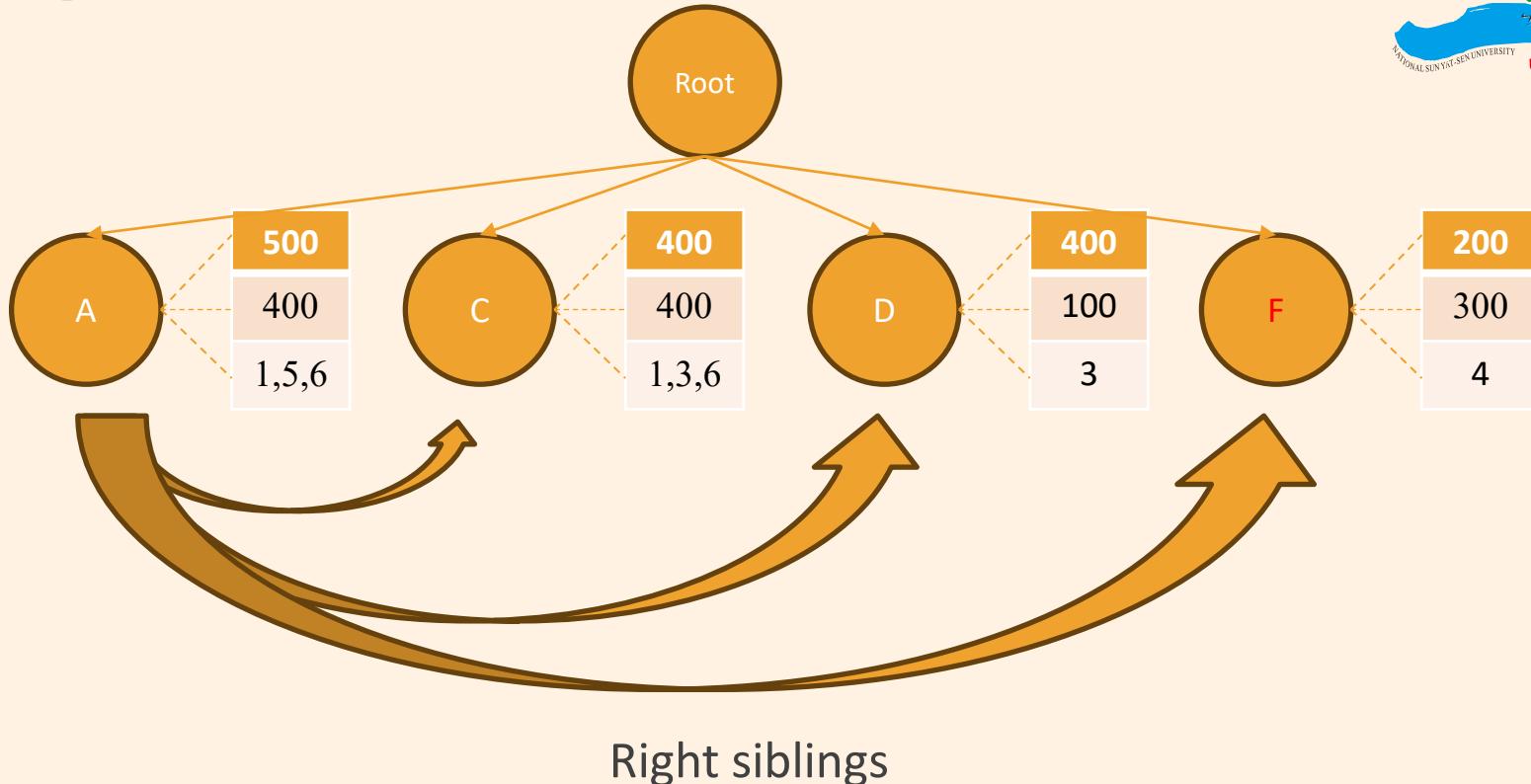
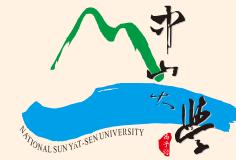


Item	A	C	D	F
MGT	500	400	400	200
<i>Gain</i>	400	400	100	300
PIDset	1, 5, 6	1, 3, 6	3	4

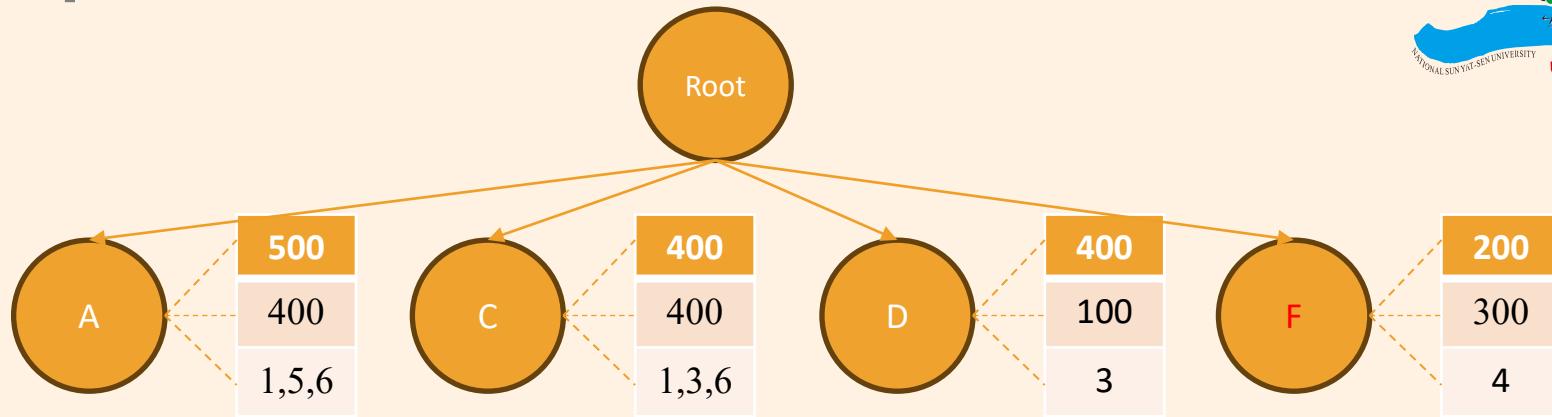
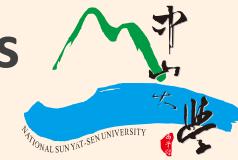


Step 7

Generate candidate itemsets for {A}



Step 8 Calculate information of candidate itemsets

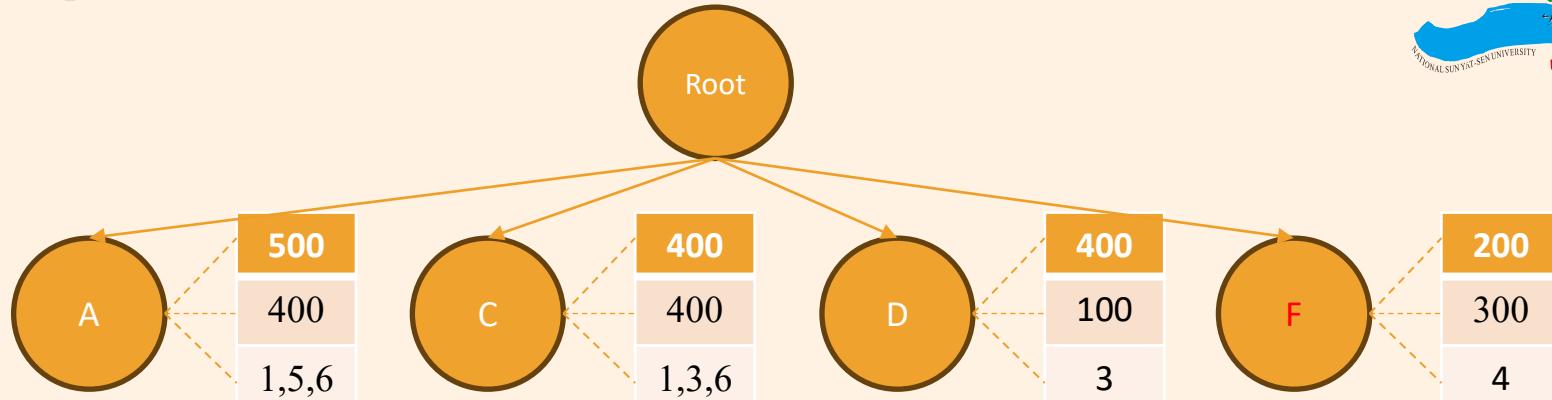


$$\begin{aligned} \text{MGT}_{\max}(\text{AC}) &= \text{MGT}_{\max}(\text{AC.parent}) \\ &= \text{MGT}_{\max}(\text{A}) \\ &= 500 \end{aligned}$$

Candidate itemsets		
AC	AD	AF



Step 8 Calculate information of candidate itemsets



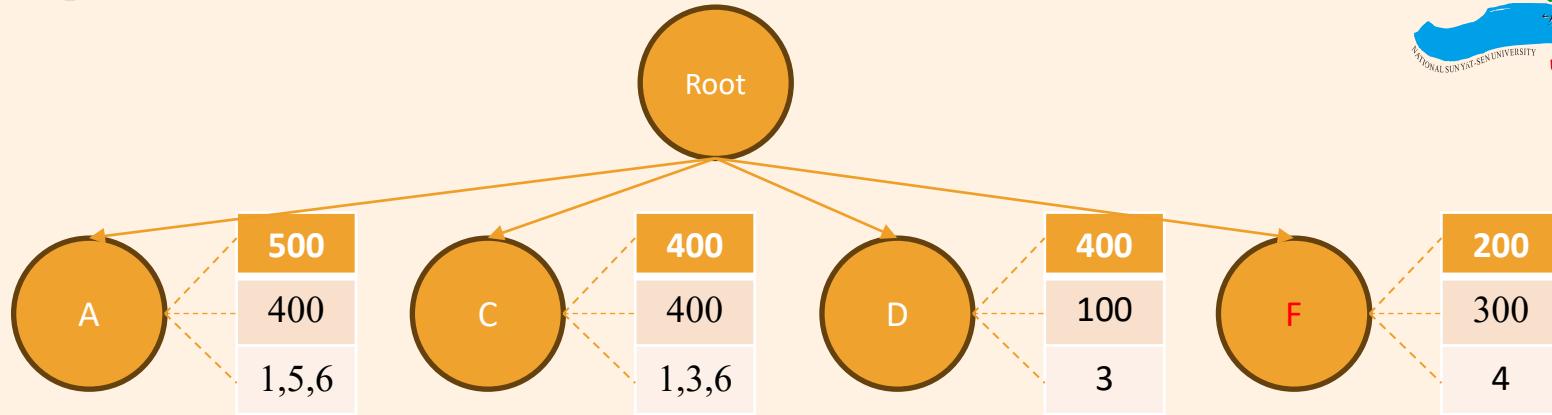
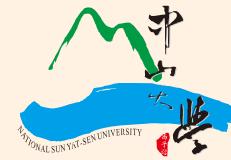
A table titled "Candidate itemsets" showing the items AC, AD, and AF.

Candidate itemsets	AC	AD	AF
	AC	AD	AF

A table summarizing information for items AC, AD, and AF.

Item	AC	AD	AF
MGT _{max}	500	500	500
Gain	500	500	700
dPIDset	3	3	4

Step 9 Verify candidate itemsets for {A}



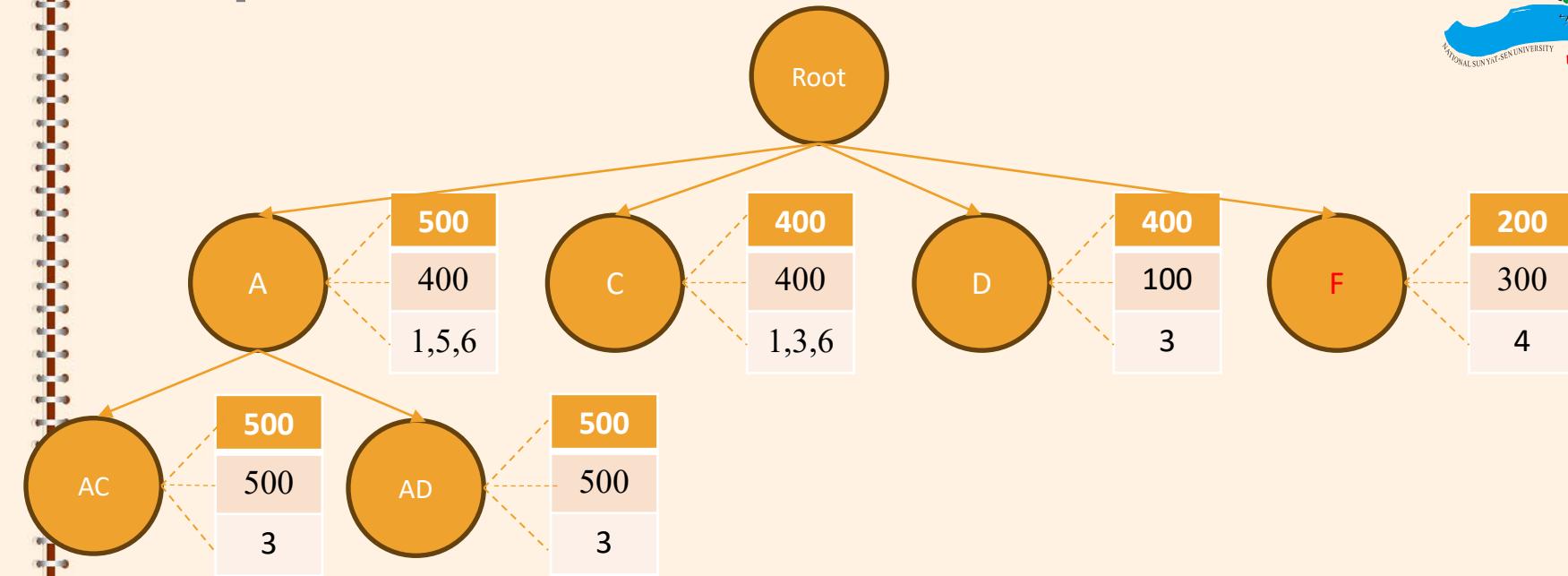
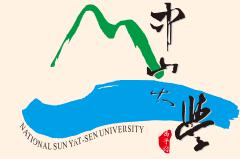
Item	AC	AD	AF
MGT_{max}	500	500	500
Gain	500	500	700
dPIDset	3	3	4



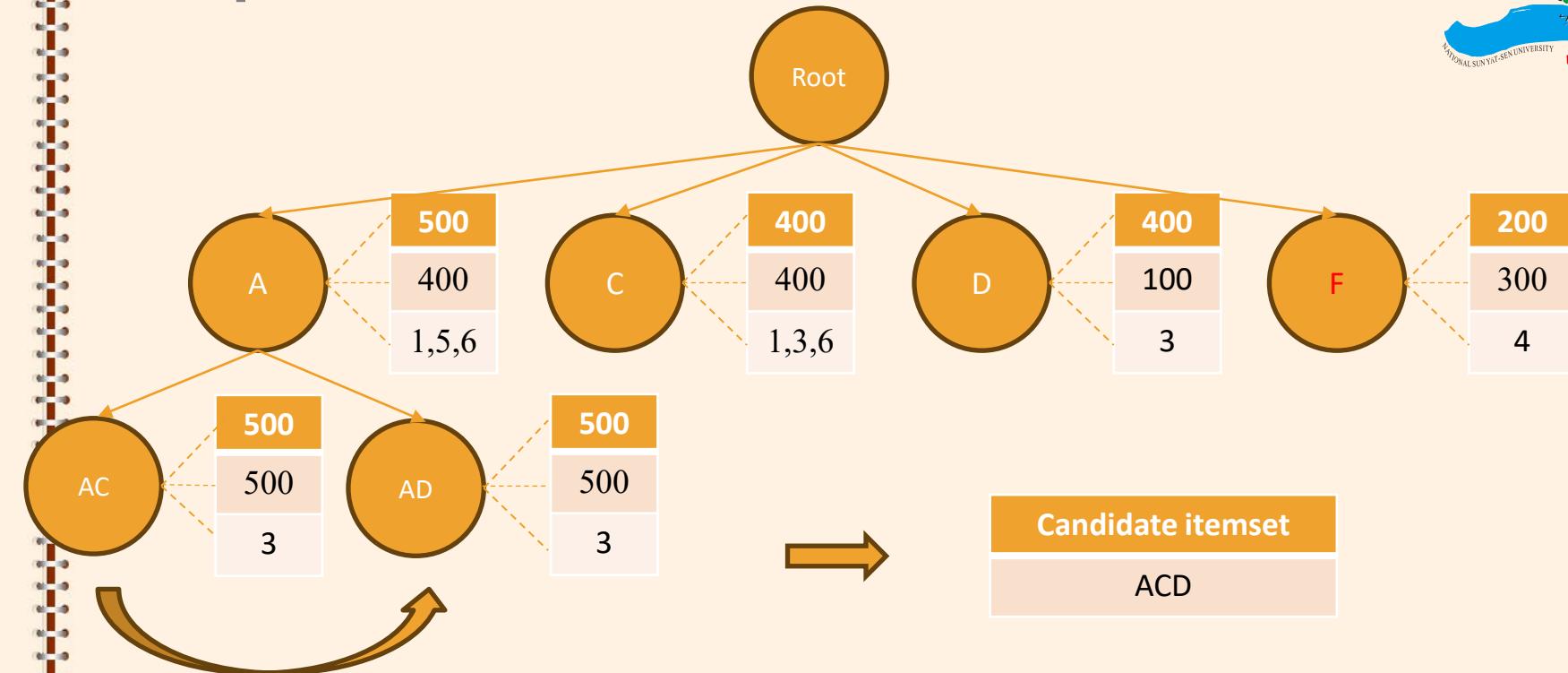
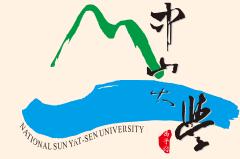
Erasable itemsets	
AC	AD



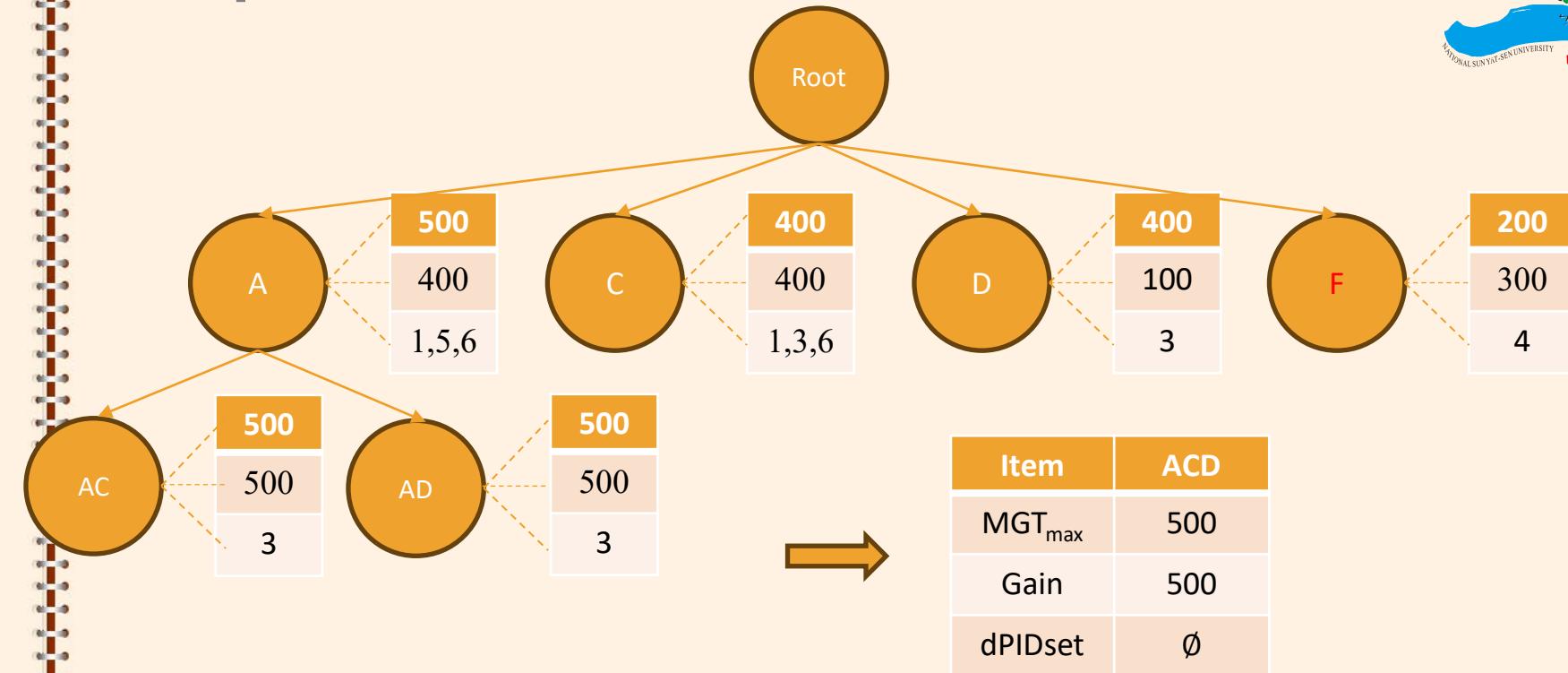
Step 10 Insert erasable itemsets for {A}



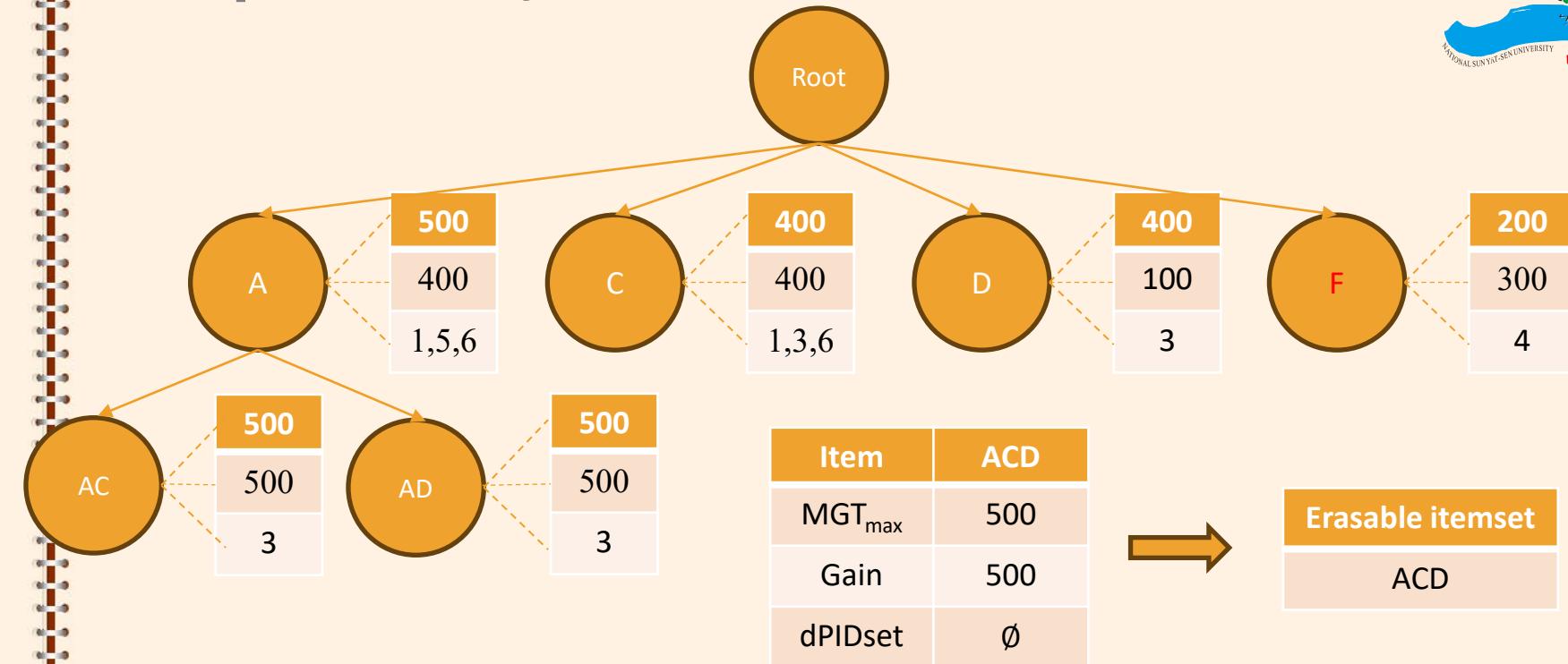
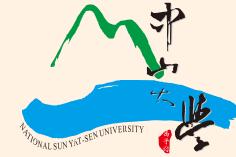
Step 11 Generate candidate itemsets for {A, C}



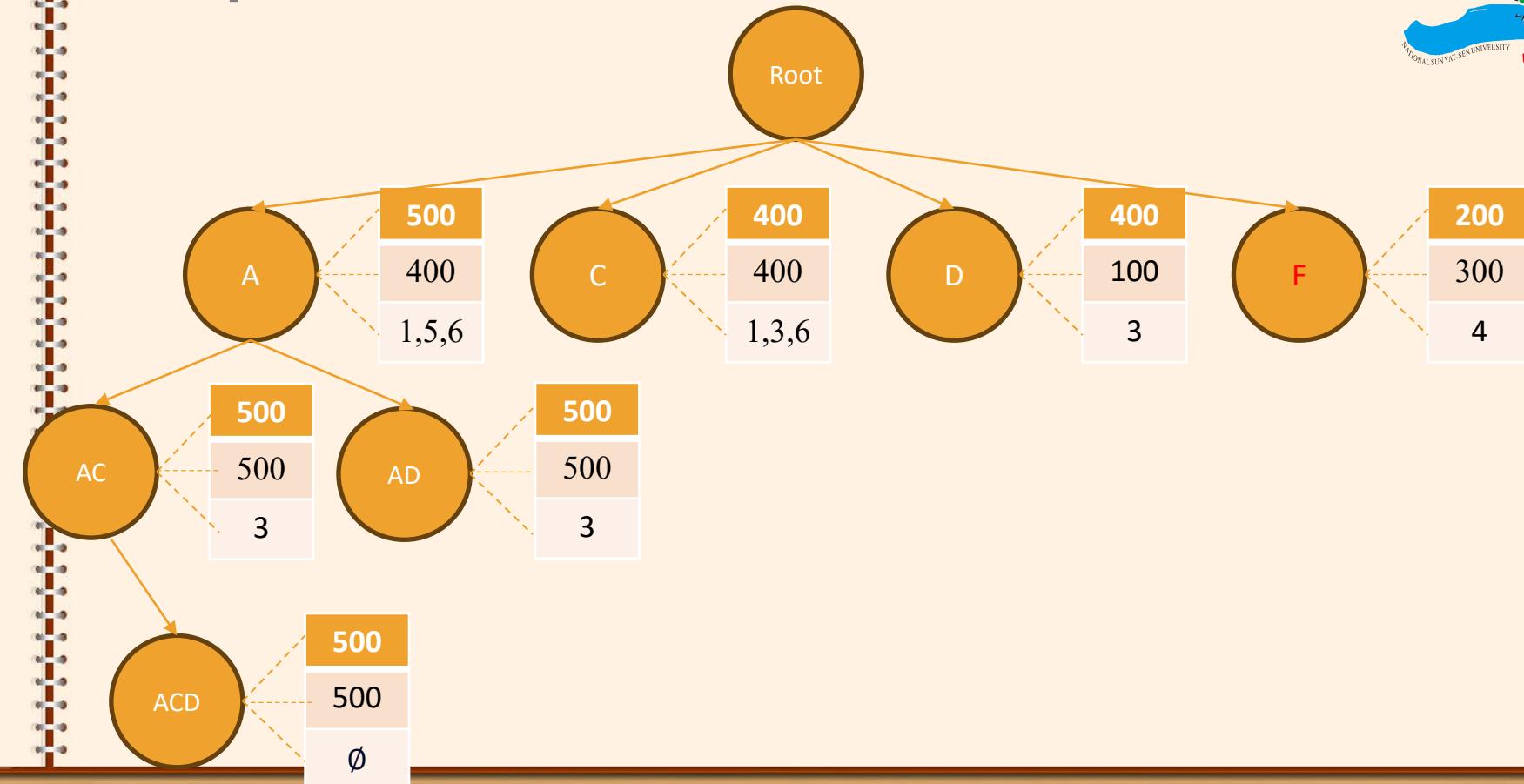
Step 12 Calculate candidate itemsets information



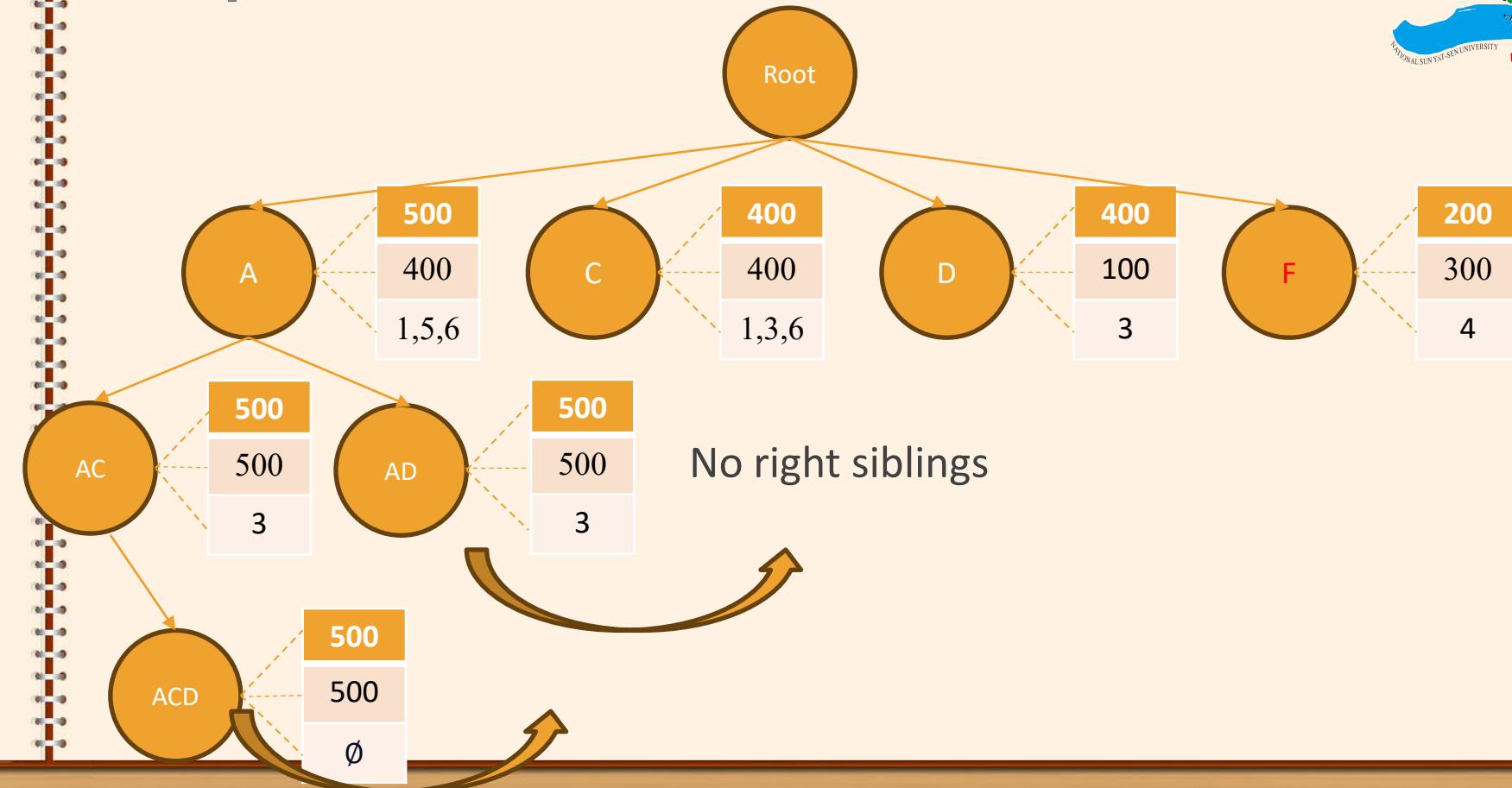
Step 13 Verify candidate itemsets for {A, C}



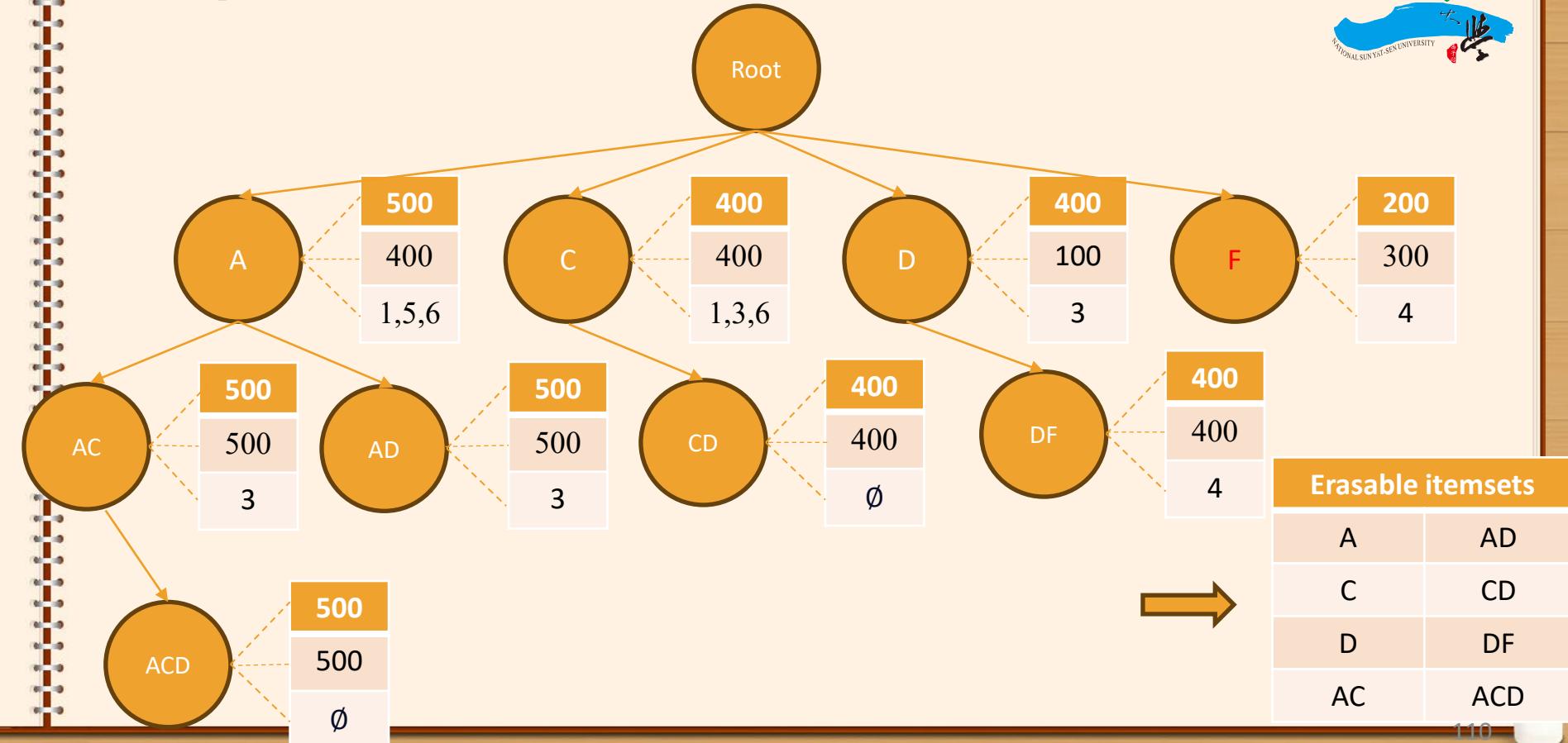
Step 14 Insert erasable itemsets for {A, C}

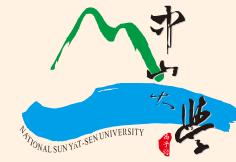


Step 15 Generate candidate itemsets for {A, D}

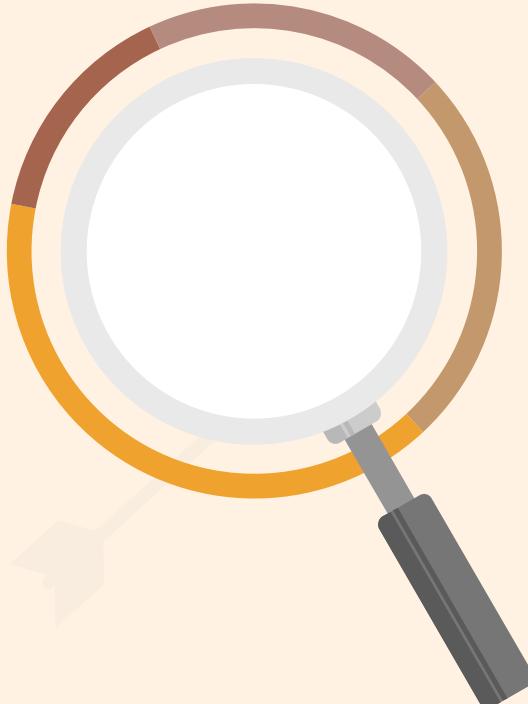


Step 16 Final result





The key for improving speed



- 01 Single database scan
- 02 Hash table for profit access
- 03 Calculate gain with dPIDset
- 04 Getting MGT_{max} from parent node



4

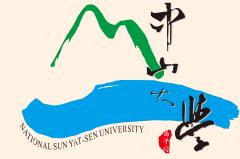
PART FOUR

Experiment

- ✓ Multiple threshold vs. Single threshold
- ✓ Apriori-base vs. MEI-base



Environment



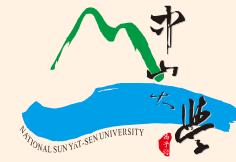
- Language: JDK 19.0.1
- CPU: Intel Core i7-9750H@2.60GHz
- Memory: 8G



Multiple-Thresholds versus Single-threshold



Synthetic datasets

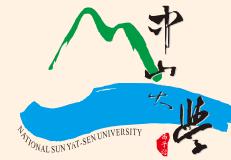


IBM IBM data generator

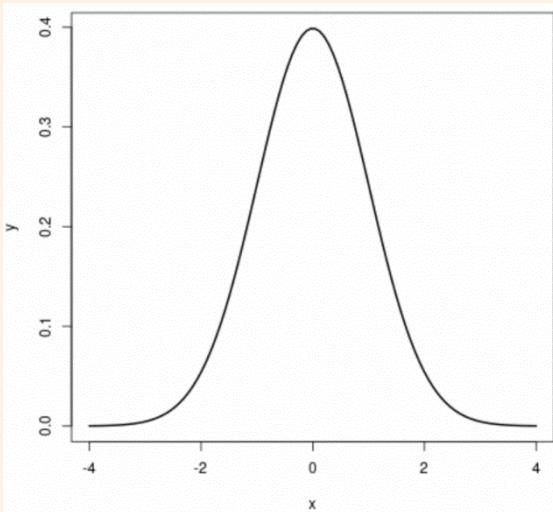
Parameter	Description
P	Number of products
I	Number of items
A	Average items per product



Additional data



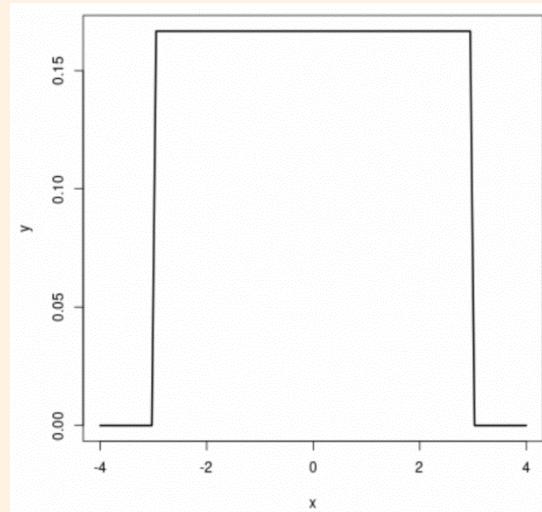
Profit



Normal distribution

$$N(100, 20)$$

Thresholds



Uniform distribution

$$U(L, H)$$



Compare method

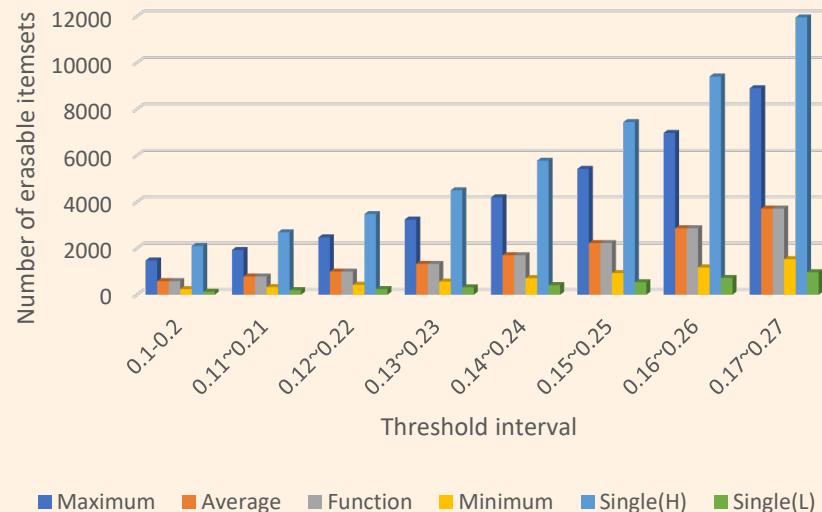
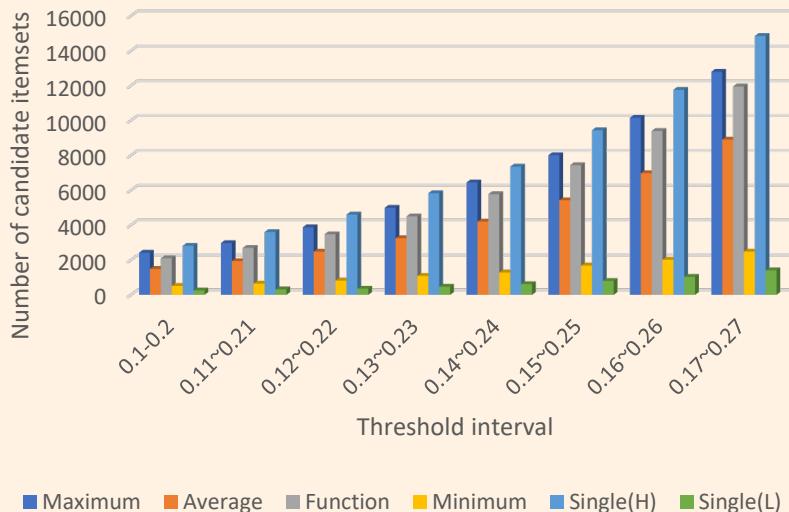
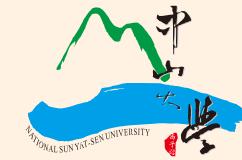


$U(L, H)$

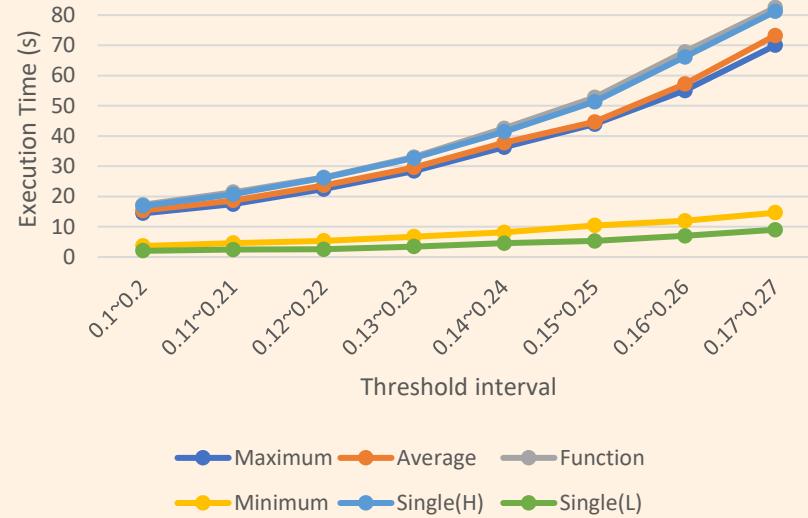
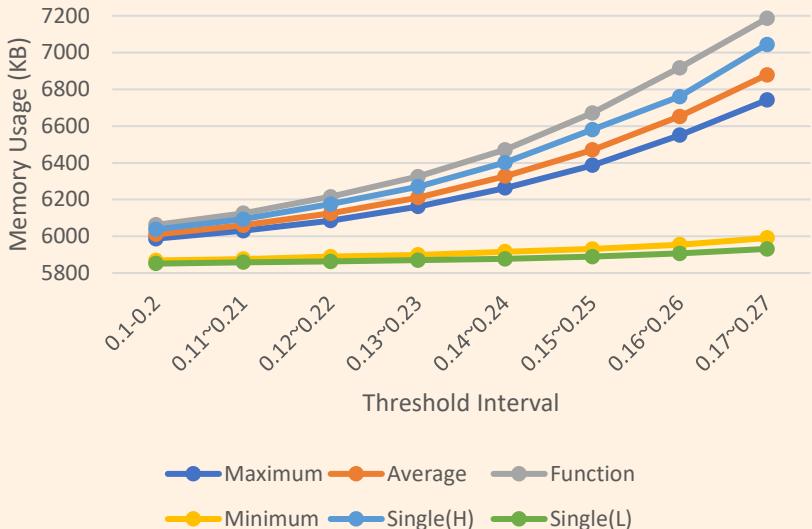
- 01 Minimum constraint
- 02 Maximum constraint
- 03 Average constraint
- 04 Function constraint (Average, $\lambda_h = H$)
- 05 Single(L) and Single(H)



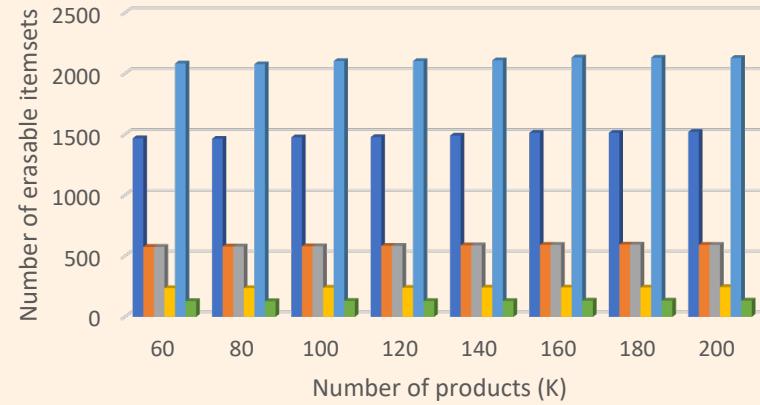
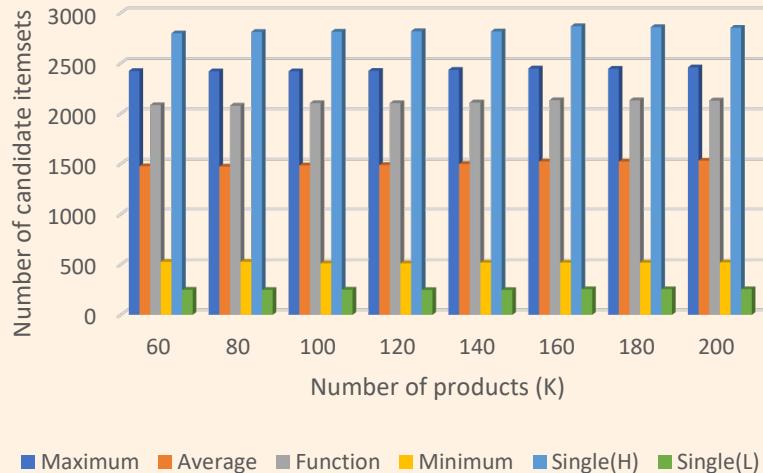
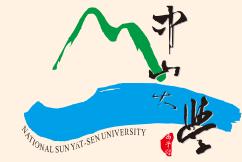
Different threshold intervals



Different threshold intervals



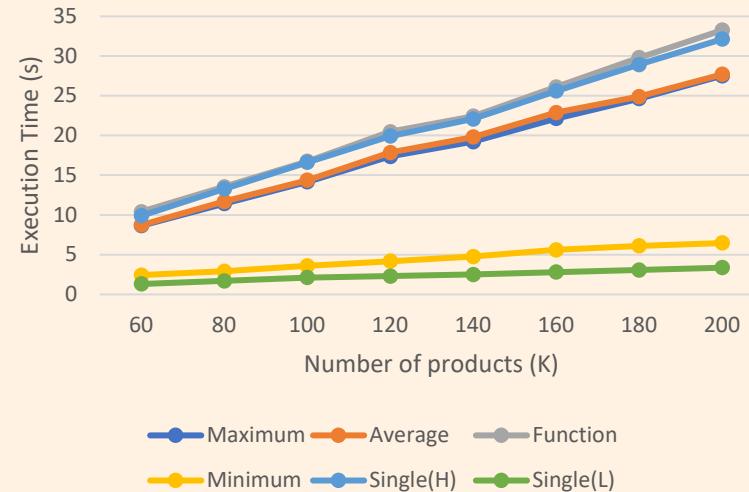
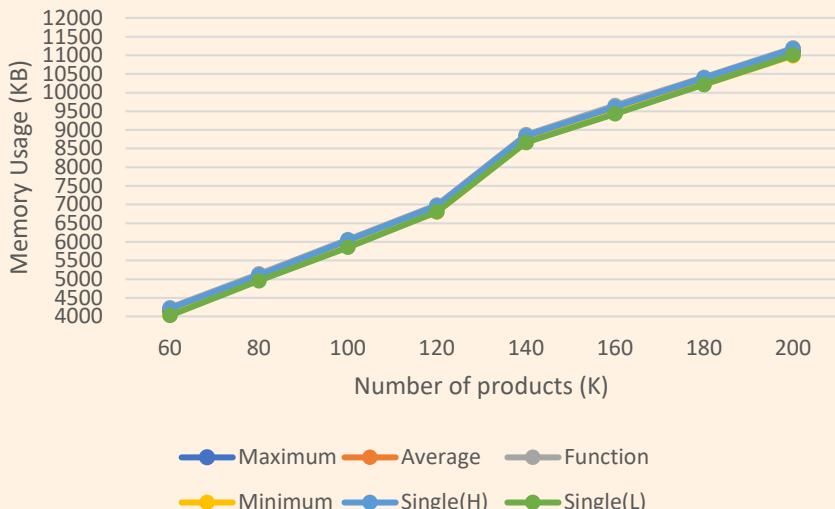
Different data sizes



Threshold interval: $U(0.1, 0.2)$



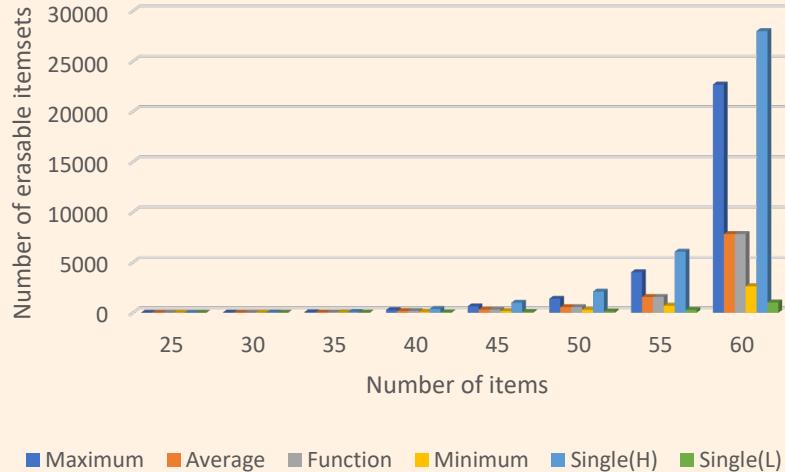
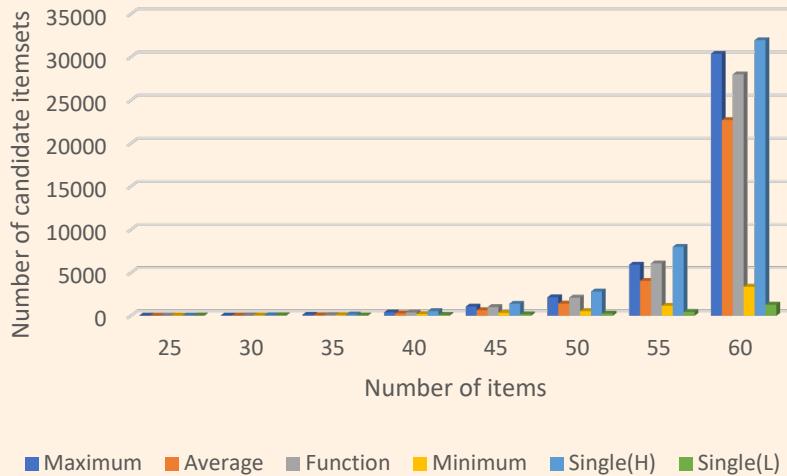
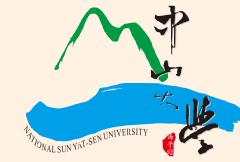
Different data sizes



Threshold interval: U(0.1, 0.2)



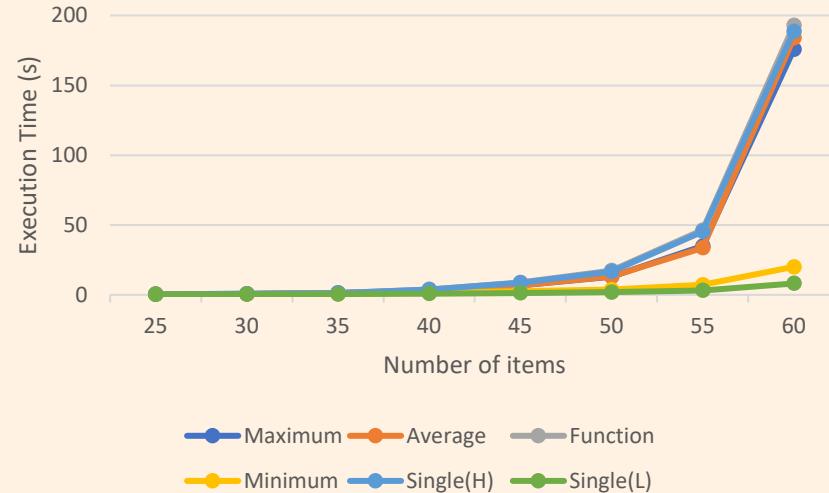
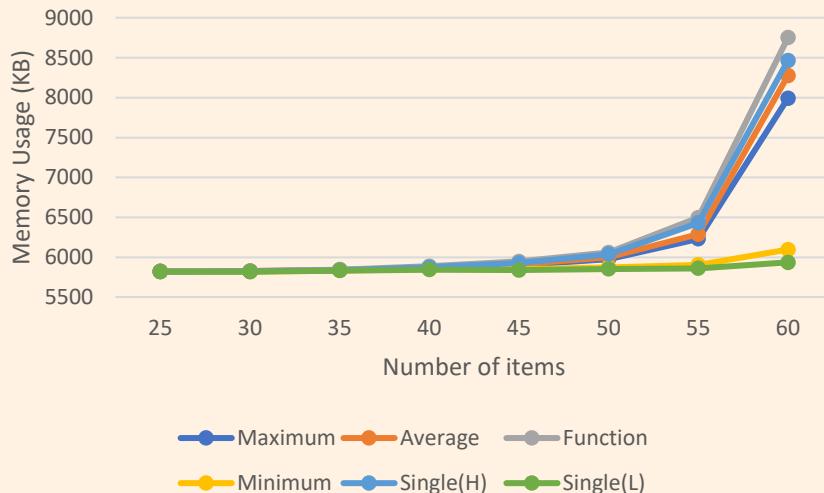
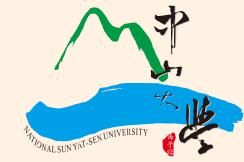
Different numbers of items



Threshold interval: U(0.1, 0.2)



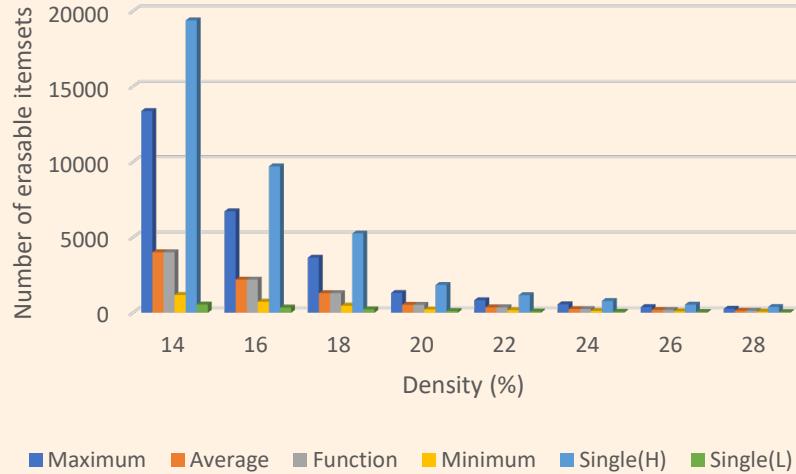
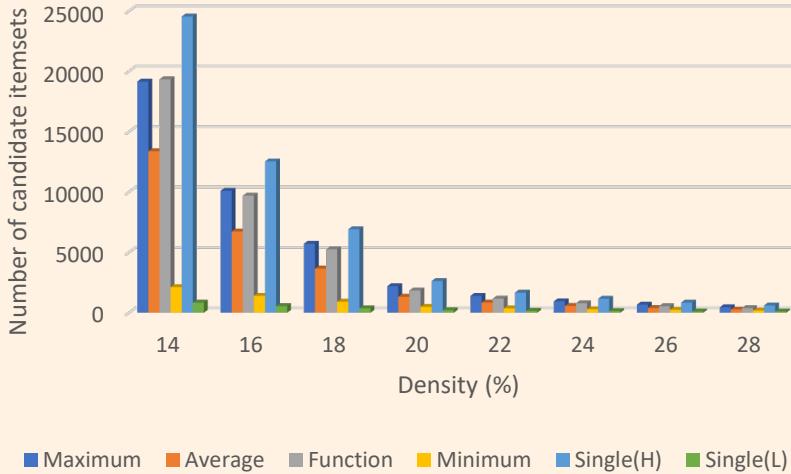
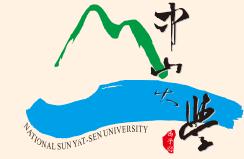
Different numbers of items



Threshold interval: U(0.1, 0.2)



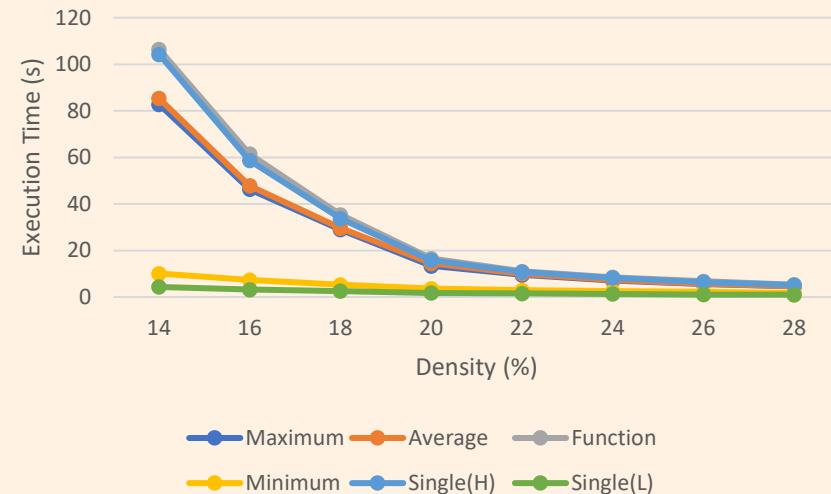
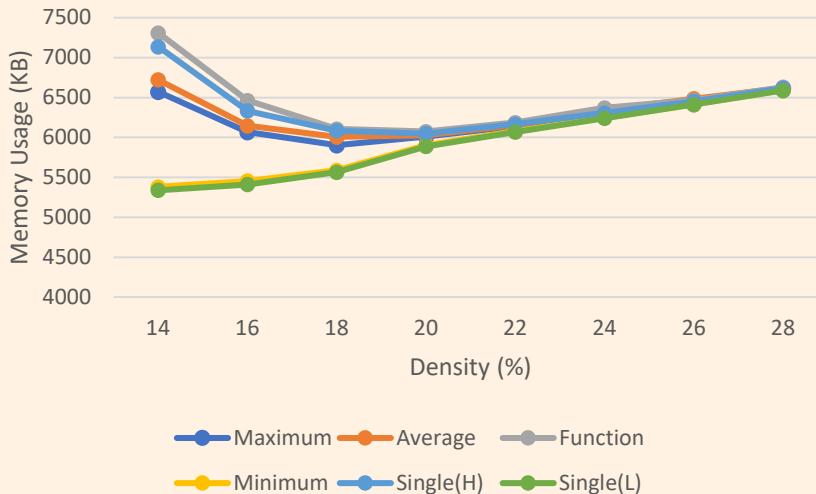
Different densities



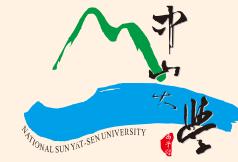
Threshold interval: U(0.1, 0.2)



Different densities



Threshold interval: U(0.1, 0.2)



Apriori-base versus MEI-base



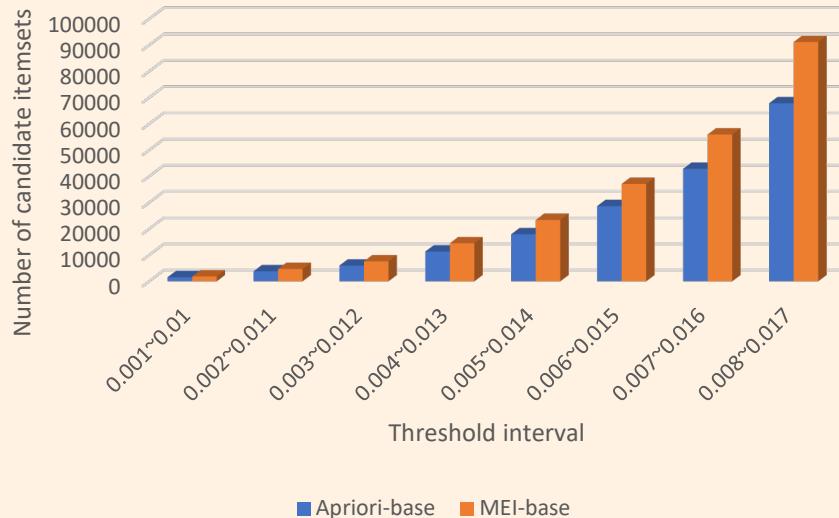
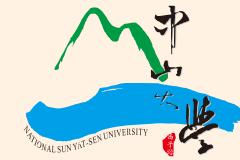
Real-life datasets



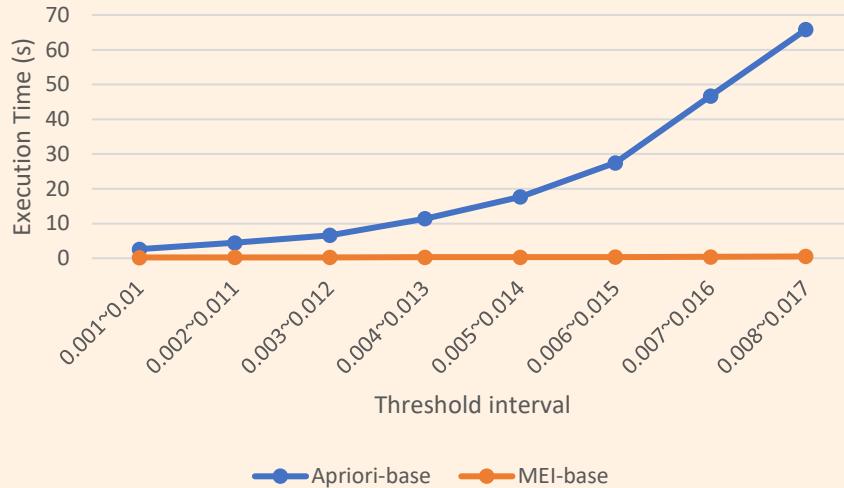
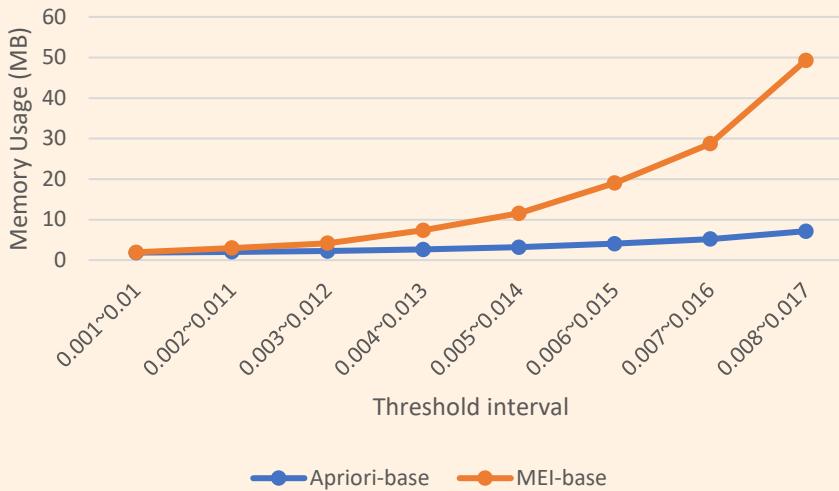
	Product	item	Average	Density
Mushrooms	8416	119	119	19.33%
Chess	3196	75	75	49.33%
Connect	67557	129	129	33.33%



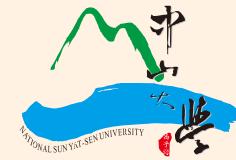
Mushrooms



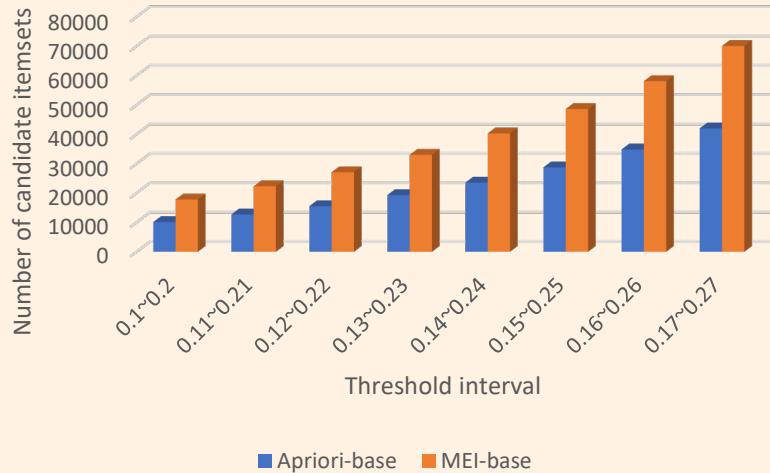
Mushrooms



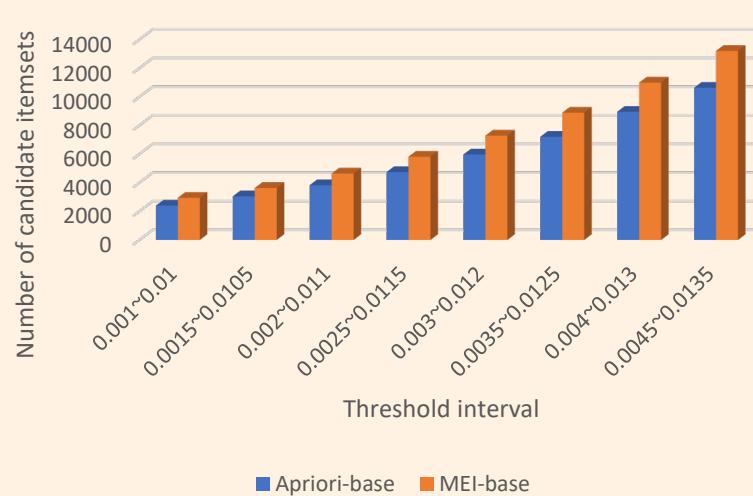
Strategy comparison



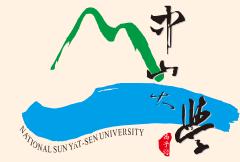
Chess



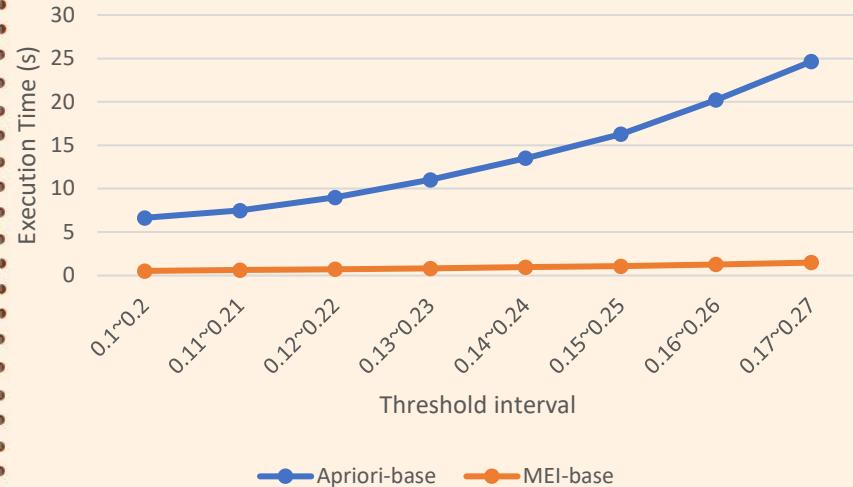
Connect



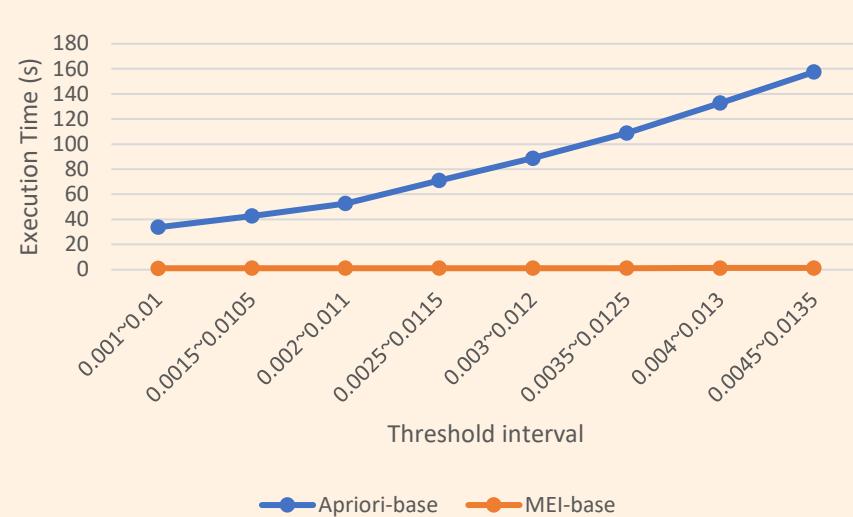
Strategy comparison



Chess



Connect



	Product	item	Average	Density
Mushrooms	8416	119	119	19.33%
Chess	3196	75	75	49.33%
Connect	67557	129	129	33.33%



1 5

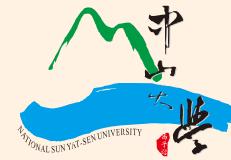
PART FIVE

Conclusion

✓ Future work

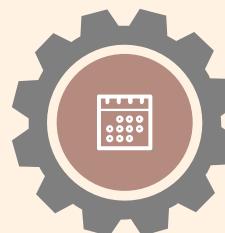


Conclusion



Maximum constraint

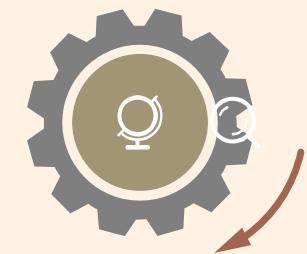
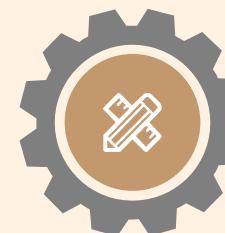
Sorted closure



Function constraint

Upper-bound

Downward closure



Minimum constraint

Downward closure

Average constraint

Upper-bound

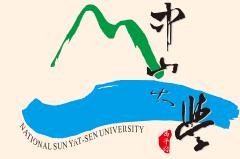
Efficient maximum constraint

Tree structure

Hash table



Future work



Better pruning skill



Reduce memory usage



Smaller upper bound



Appropriate threshold





Thank you

Q & A

