ACCSE 2024
Mining Erasable Itemsets with Multiple Thresholds under the Loose Constraint

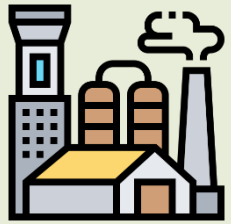
Tzung-Pei Hong, Yi-Chen Chang,
Chun-Ho Wang, and Wei-Ming Huang

Reporter: Yi-Chen Chang
Email: 4a0g0902@stust.edu.tw

CONTENTS

01 Introduction

Erasable itemset mining



PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

- Production management
- Product database
- Sudden trouble

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

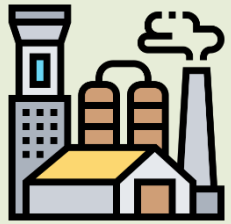
Erasable itemset mining

Funding decrease

Insufficient storage space

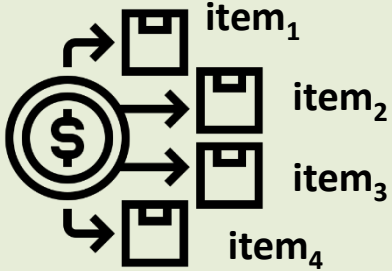
Limited logistics

Erasable itemset mining



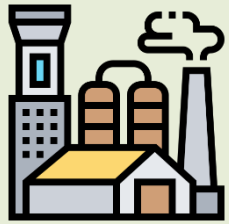
PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

- Control loss
- Find combinations of items that result in low losses



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

Erasable itemset mining



PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

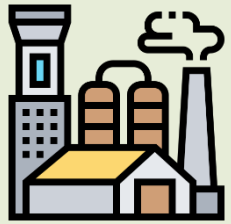
Gain

The loss incurred when certain items cannot be stocked.

→ Gain(A) = 200 + 100 + 100 = 400

Item	A	B	C	D	E	F
λ	0.6	0.5	0.4	0.7	0.3	0.8

Erasable itemset mining



PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

Maximum gain threshold (MGT)

Actual threshold value

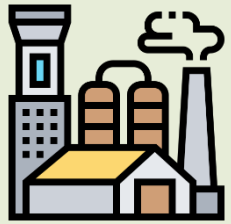
MGT(i) = Total profit * λ(i)

Total profit = 200+200+100+100+300+100 = 1000

MGT(A) = 1000 * 0.6 = 600

Item	A	B	C	D	E	F
λ	0.6	0.5	0.4	0.7	0.3	0.8

Erasable itemset mining



$$\lambda(x) = \max(\lambda(I) \mid I \in X)$$

Loose constraint

PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

Itemset

One item → λ(A) = 0.6

More than one item →

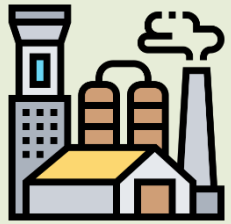
λ (AB) = max(λ(A), λ(B))

= max(0.6, 0.5)

= 0.6

Item	A	B	C	D	E	F
λ	0.6	0.5	0.4	0.7	0.3	0.8

Erasable itemset mining



PID	Items	Profit
P1	ABE	200
P2	DEF	200
P3	BCE	100
P4	ADF	100
P5	BF	300
P6	ACDF	100

Lack of downward closure property

Itemset{D, E} is an erasable itemset

Gain(DE) =700 ≤ MGT(DE) = 700

→ Subset {D} and {E} are also erasable itemsets

Gain(E)=500 >MGT(E) = 300 ✖

Item	A	B	C	D	E	F
λ	0.6	0.5	0.4	0.7	0.3	0.8

02 Proposed algorithm

Step 1

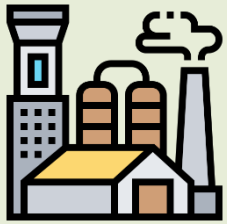
PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Sort the thresholds in descending order

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

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	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 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

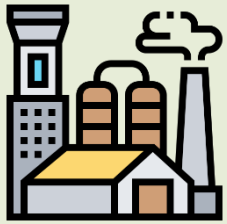
Calculate MGT of each item

Total profit = 2000

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

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

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

Find the first erasable itemset

Gain ≤ MGT

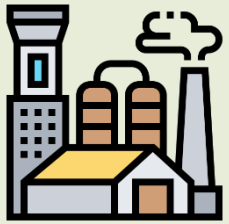
Item	G	D	F	C	A	E	B
MGT	1400	1200	1200	1000	800	600	200
Gain	1500	600	400	500	500	900	1300

Candidate 1-itemsets
D

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

Verify itemset sorted after {D}

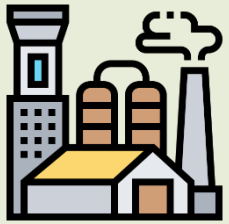
Gain ≤ MGT(D)

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



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

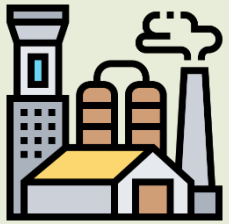
Gain ≤ MGT

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



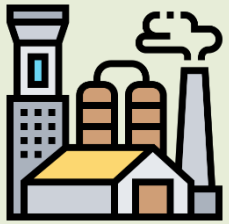
PID	Items	Profit
P1	ABC	200
P2	CDE	300
P3	AF	100
P4	BFG	200
P5	ACDEG	400
P6	EFG	500
P7	DE	100
P8	BCFG	200

Generate candidate 2-itemsets

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	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

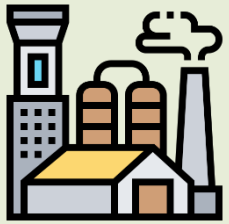
Item	D	F	C	A	E
MGT	1200	1200	1000	800	600
Gain	600	400	500	500	900

Gain(F) ≤ MGT(D)

Candidate 2-itemsets
DF

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

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

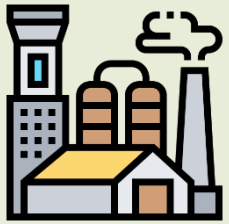
Generate candidate 2-itemsets

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



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

Candidate 1-itemsets				
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

$Gain \leq MGT_{max}$

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

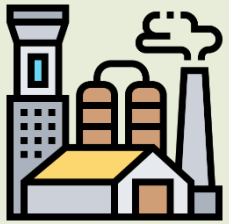
Itemset	MGT_{max}	Gain
DF	1200	900
DC	1200	800
DA	1200	1100
DE	1200	1100
FC	1200	900
FA	1200	600
FE	1200	1200
CA	1000	800
CE	1000	1100

$Gain \leq MGT_{max}$

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	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			
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.

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

Union

Candidate 3-itemsets
DFC

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

Union

Candidate 3-itemsets
DFC
DFA

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

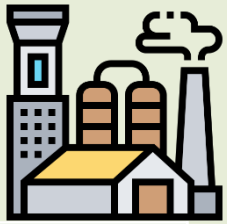
Union

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

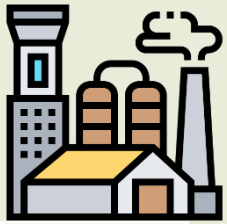
Erasable 2-itemsets
DF
DC
DA
DE
FC
FA
FE
CA

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

Step 9-2

Prune candidate itemsets

Case 1: MGT($item_1$) ≠ MGT($item_2$)



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

MGT(F) ≠ MGT(C)

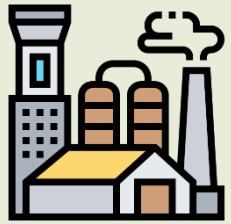
(k-1)-subsets contain $item_1$

2-subset
FC
FA
CA

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	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

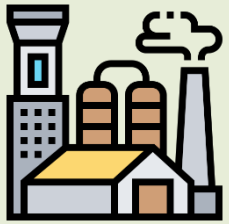
Erasable 2-itemsets			
DF	DC	DA	DE
FC	FA	FE	CA

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

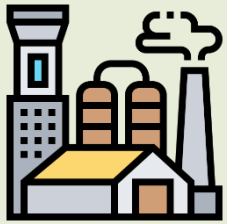
$Gain \leq MGT_{max}$

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	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 4-itemsets

Erasable 3-itemsets			
DFC	DFA	DCA	FCA

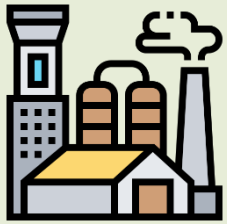
Gain ≤ MGT_{max}

Candidate 4-itemsets
DFCA

Erasable 4-itemset
DFCA

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

Step 12



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

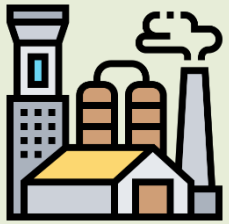
Generate candidate 5-itemsets

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 13



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

Output all erasable itemsets

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

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

03 Experiment

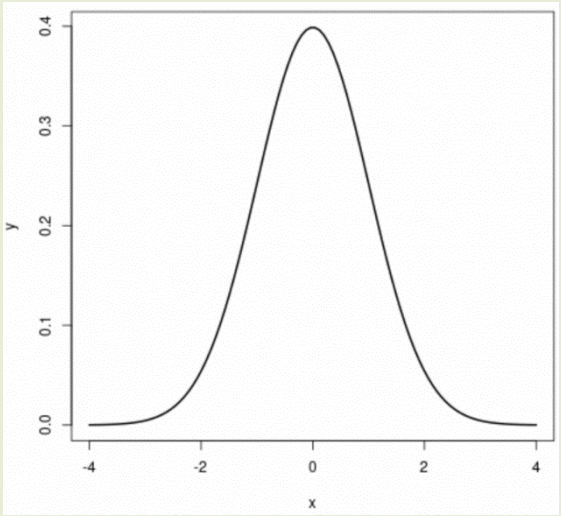
Synthetic datasets

IBM data generator

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

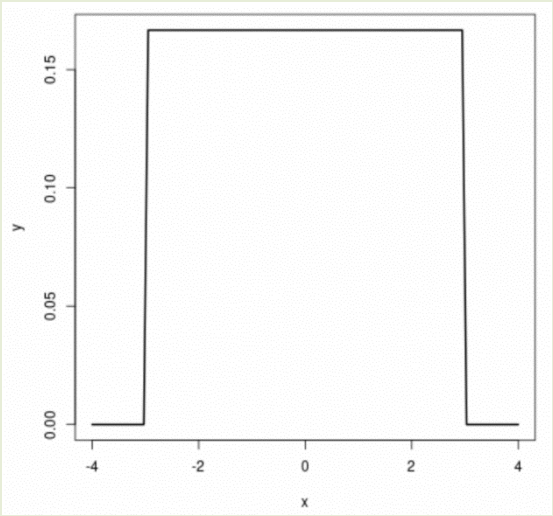
Additional data

Profit



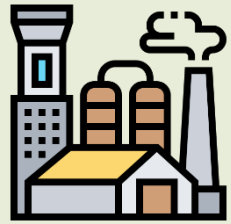
Normal distribution
N(100, 20)

Thresholds



Uniform distribution
U(L, H)

Compare



Multiple thresholds

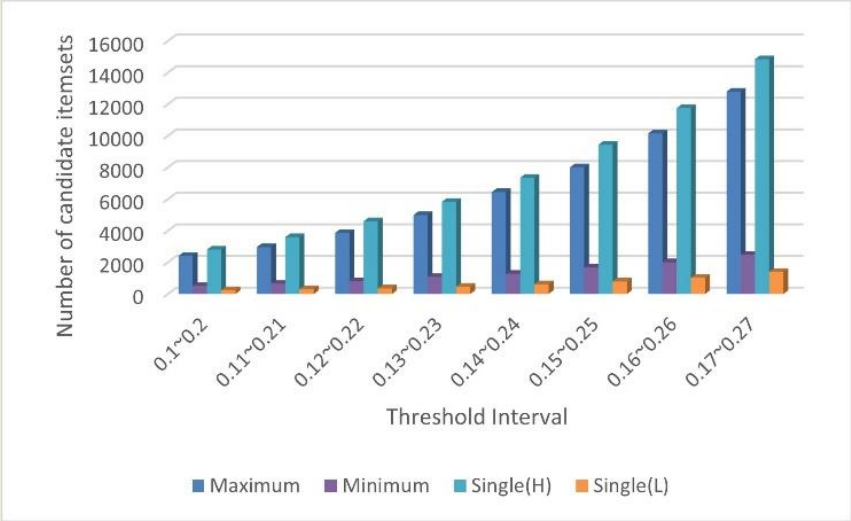
- Tight constraint
- Loose constraint

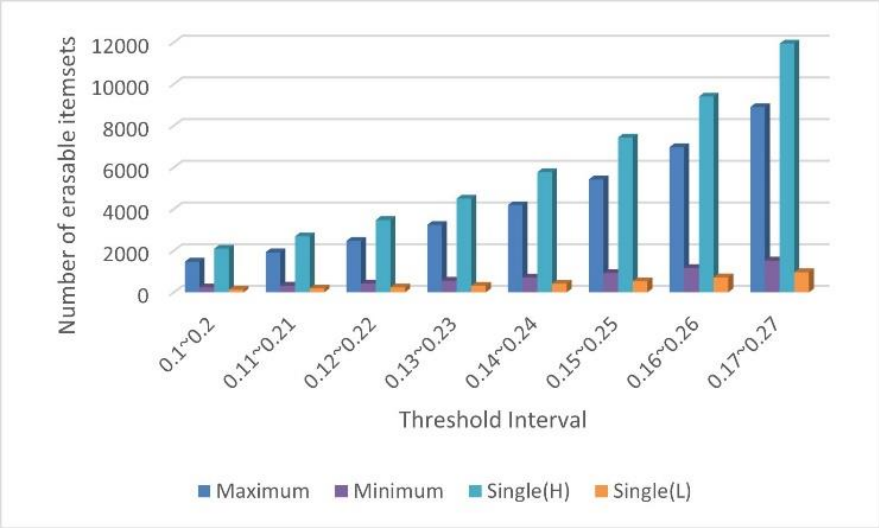
Single threshold

- Lowest threshold
- Highest threshold

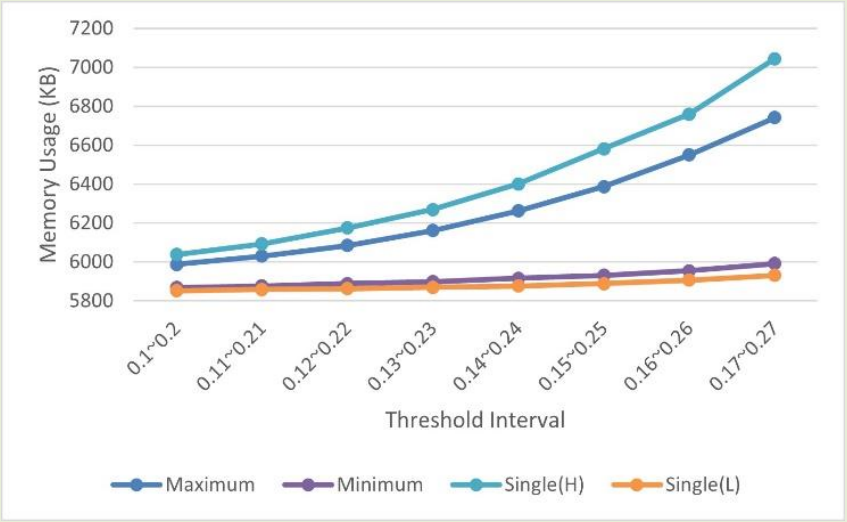
Different threshold intervals

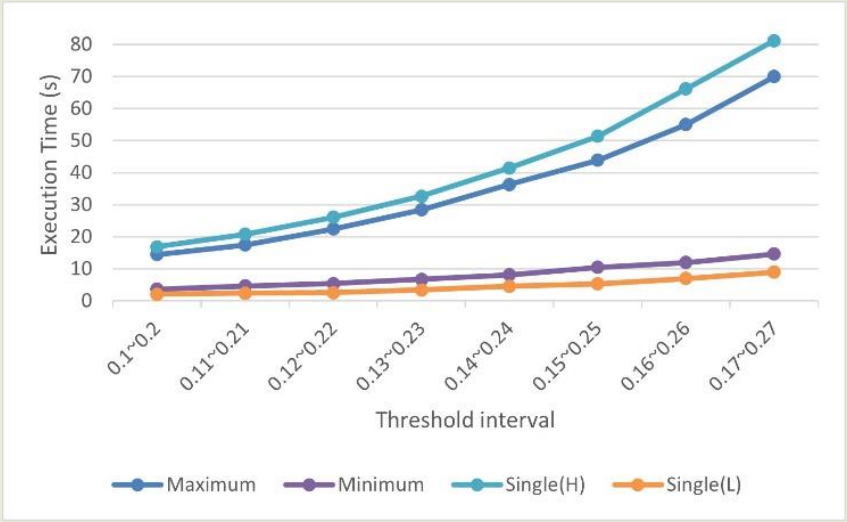
Dataset: P100KI0.05KA10





Different threshold intervals





04 Conclusion

Conclusion and future work

Solve the lack of downward closure

Better performance

Different constraint

Thanks for listening