CONTENTS

Association Rule Mining

F(

	Transaction database
1.	milk, bread, cereal
2.	milk, bread, sugar, eggs
3.	milk, bread, butter
…..	……..
n.	sugar, eggs

,Ts,Tc) → (milk, bread) → (eggs)

(bread) → (butter)

Association Rule Mining

Association Rule Mining is generally done in two steps

1. Find all frequent itemsets
2. Generate association rules from frequent itemsets

Apriori and FP-Growth are traditional algorithms for finding frequent itemsets.

Why MASP algorithm?

- No need to search entire lattice of item combinations
- No pruning step required
- Less computational resources
- K-items patterns can be obtained easily

MASP algorithm

MASP algorithm generates rules in two steps

1. Generate MASP tree
2. Generate rules from MASP tree

MASP tree generation

	C1	C2	C3	C4	C5
1	1	2	3	4	5
2	1	5	3	4	2
3	8	1	9	2	3
4	9	2	3	1	7
5	1	9	3	8	7
6	1	8	3	2	7

Threshold Support = 5%
Threshold Confidence = 10%

MASP tree generation

	C1	C2	C3	C4	C5
1	1	2	3	4	5
2	1	5	3	4	2
3	8	1	9	2	3
4	9	2	3	1	7
5	1	9	3	8	7
6	1	8	3	2	7

Item	Freq	Item	Freq	Item	Freq
C1=8	1	C2=8	1	C4=1	1
C1=9	1	C2=9	1	C4=8	1
C1=1	4	C3=3	5	C5=2	1
C2=2	2	C3=9	1	C5=3	1
C2=5	1	C4=4	2	C5=5	1
C2=1	1	C4=2	2	C5=7	3

MASP tree generation

	C1	C2	C3	C4	C5
1	1	2	3	4	5
2	1	5	3	4	2
3	8	1	9	2	3
4	9	2	3	1	7
5	1	9	3	8	7
6	1	8	3	2	7

Item	Freq	Item	Freq	Item	Freq
C1=8	1	C2=8	1	C4=1	1
C1=9	1	C2=9	1	C4=8	1
C1=1	4	**C3=3**	**5**	C5=2	1
C2=2	2	C3=9	1	C5=3	1
C2=5	1	C4=4	2	C5=5	1
C2=1	1	C4=2	2	C5=7	3

MASP tree generation

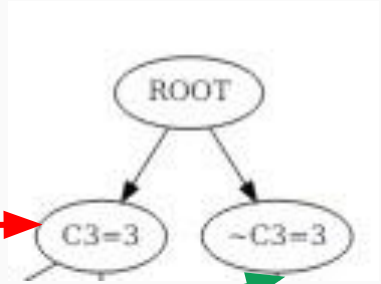
	C3	C1	C2	C4	C5
1	3	1	2	4	5
2	3	1	5	4	2
4	3	9	2	1	7
5	3	1	9	8	7
6	3	1	8	2	7
3	9	8	1	2	3

Support(C3=3) = 5/6 >= Threshold Support
Confidence(C3=3) = 5/6 >= Threshold Confidence

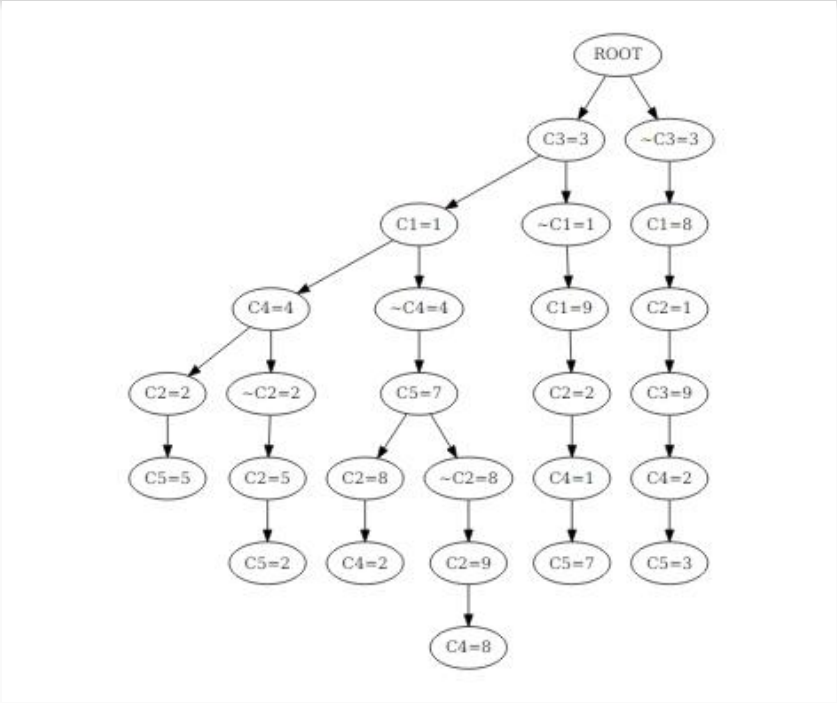
Support(~C3=3) = 1/6 >= Threshold Support
Confidence(~C3=3) = 1/6 >= Threshold Confidence

MASP tree generation

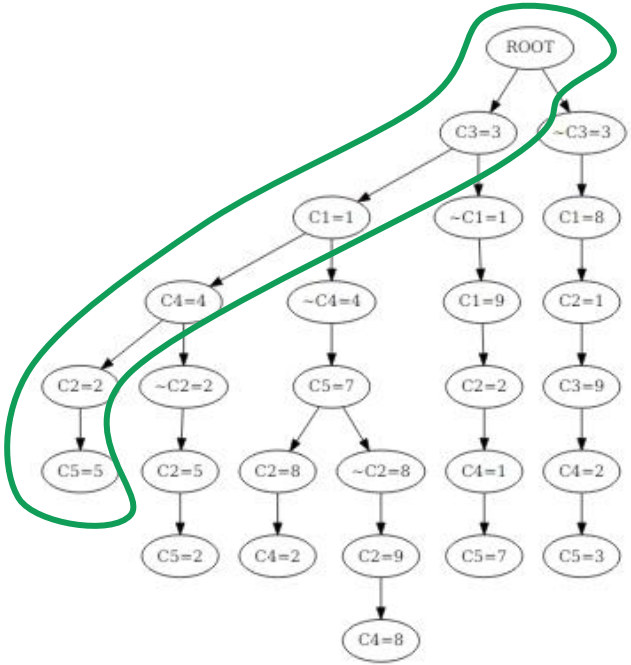
	C3	C1	C2	C4	C5
1	3	1	2	4	5
2	3	1	5	4	2
4	3	9	2	1	7
5	3	1	9	8	7
6	3	1	8	2	7
3	9	8	1	2	3



MASP tree generation



Rules generation



PATH

C3=3,C1=1,C4=4,C2=2,C5=5

RULES

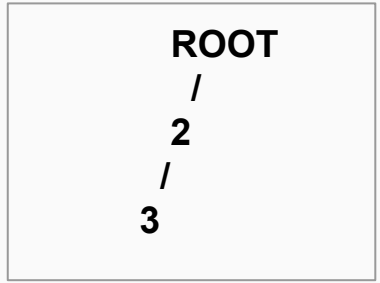
(C3=3) → (C1=1)
(C3=3,C1=1) → (C4=4)
(C3=3,C1=1,C4=4) → (C2=2)
(C3=3,C1=1,C4=4,C2=2) → (C5=5)

Drawbacks of MASP algorithm

- It depends on the order of items in transactions

1	2	3	4
2	4	3	1
3	2	4	1
4	2	3	1

Ts = Tc = 50% →

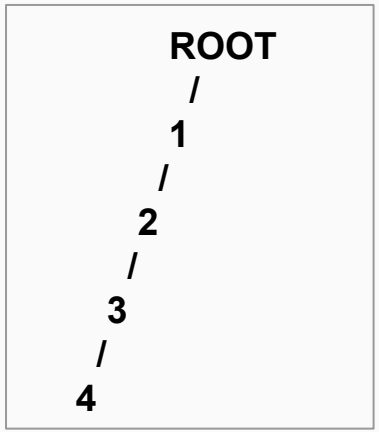


Drawbacks of MASP algorithm

- It depends on the order of items in transactions

1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4

Ts = Tc = 50% →



Drawbacks of MASP algorithm

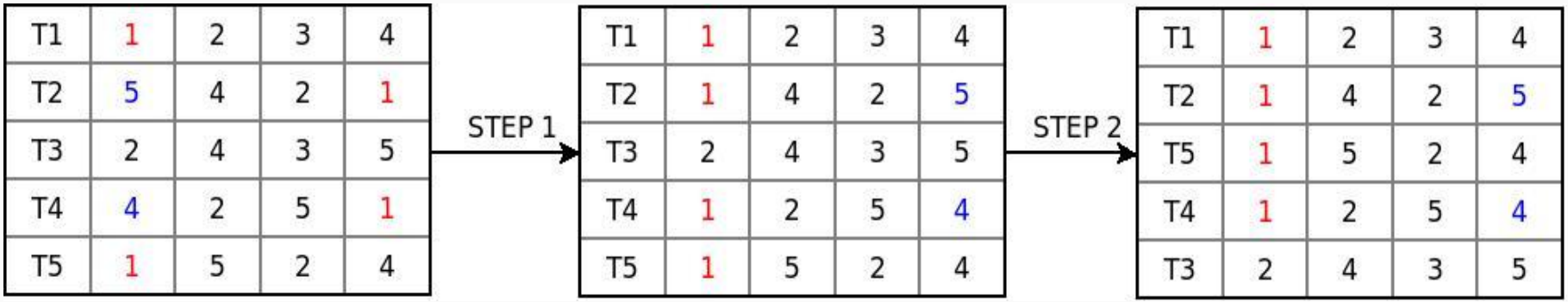
- It can generate contradictory rules
- Examples
  - (A, B, ~C, D, E) → (~E)
  - (A, B, ~C, D, E) → (C)

OIMASP algorithm

- Order independent
- Non-contradictory rules

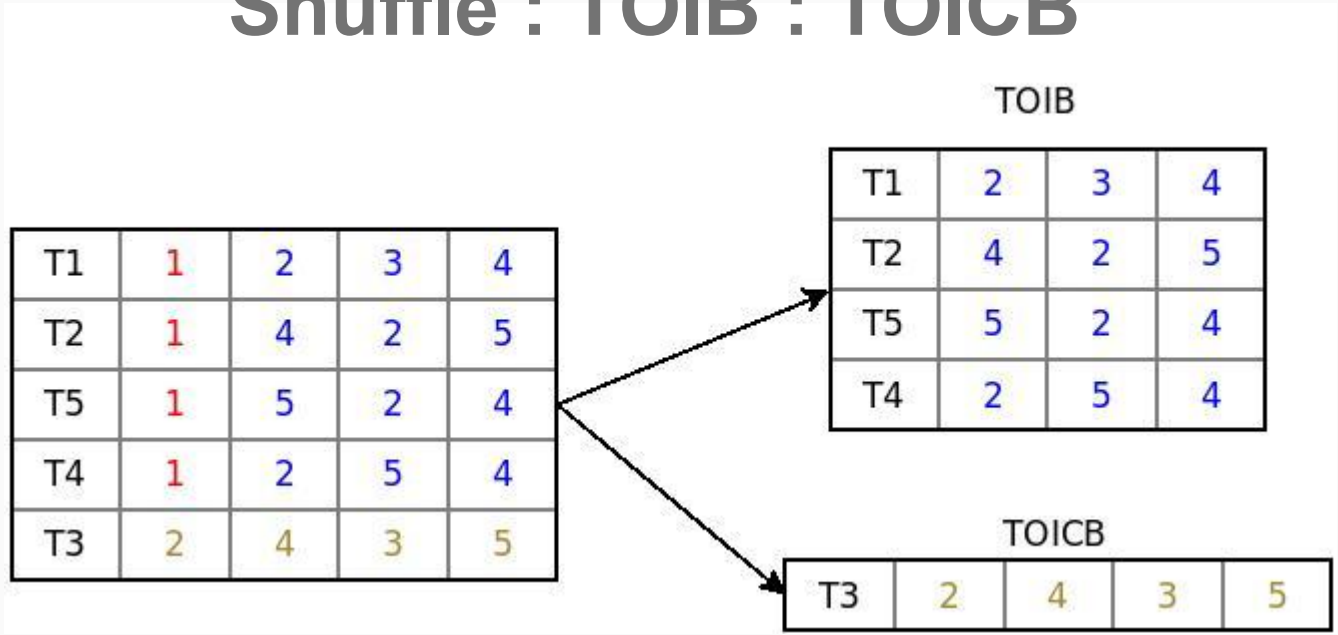
OIMASP algorithm

Shuffle : TOIB : TOICB



OIMASP algorithm

Shuffle : TOIB : TOICB



OIMASP algorithm

Threshold Support = 20%
Threshold Confidence = 30%

Transaction Table

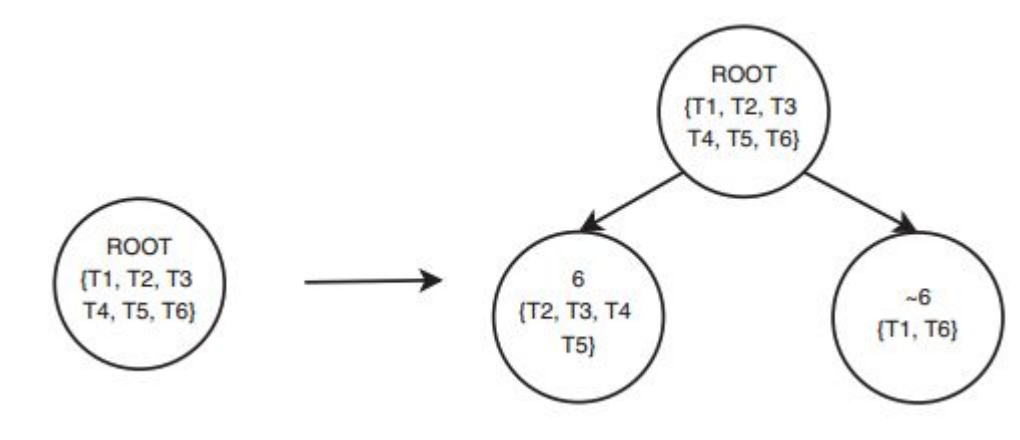
T1	1	12	3	4	5
T2	1	5	6	4	12
T3	8	6	9	12	5
T4	9	2	3	6	7
T5	6	9	10	8	7
T6	1	8	3	2	7

Frequency Table

Item	Count
1	3
10	1
12	3
2	2
3	3
4	2
5	3
6	4
7	3
8	3
9	3

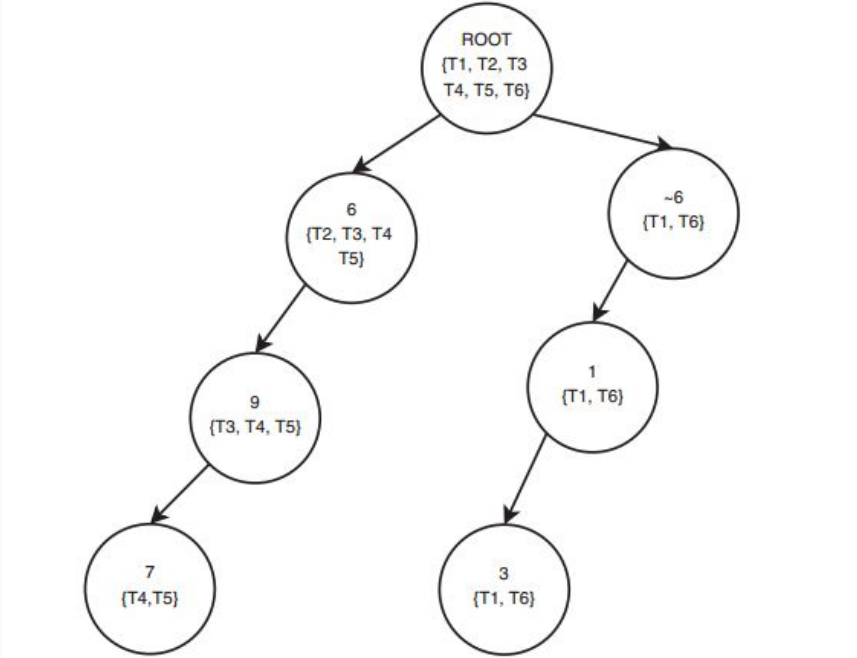
OIMASP algorithm

Threshold Support = 20%
Threshold Confidence = 30%



OIMASP algorithm

OIMASP tree



OOIMASP = **O**rigin + OIMASP

- Take into consideration the origins of items in the transaction dataset

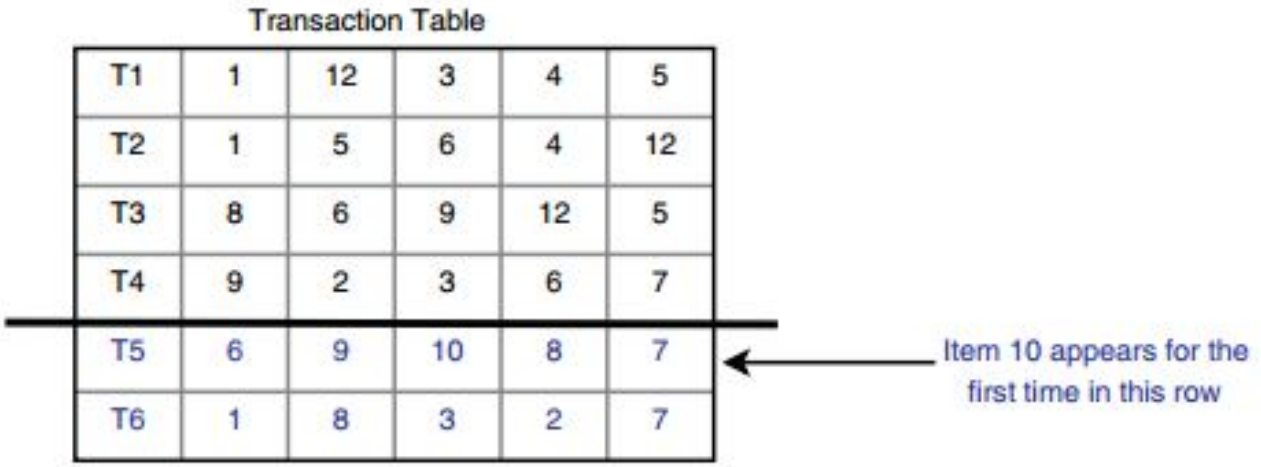
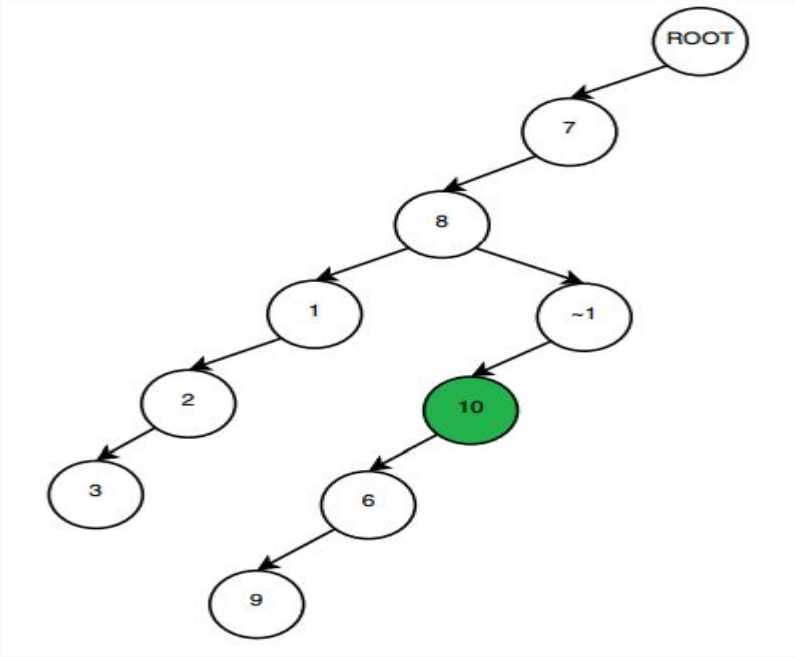


Fig. 9. Partition of dataset based on the origin of item 10.

OOIMASP = **O**rigin + OIMASP



(7, 8, ~1) => (10)
(7, 8, ~1, 10) => (6)
(7, 8, ~1, 10, 6) => (9)

OOIMASP vs MASP

			MASP		OOIMASP	
Dataset	Min-Support	Min-Confidence	total rules	max-rule-size	total rules	max-rule-size
A	0.30	0.60	1	2	4	4
B	0.30	0.60	0	0	23	10
C	0.30	0.60	0	0	31	17
D	0.30	0.60	0	0	81	51
E	0.30	0.60	0	0	262	132
Connect	0.30	0.60	29	30	55	32
Mushroom	0.30	0.60	6	7	11	14
Chess	0.30	0.60	19	20	20	20
A	0.01	0.30	14	7	24	7
B	0.01	0.30	0	0	85	21
C	0.01	0.30	0	0	133	41
D	0.01	0.30	0	0	201	101
E	0.01	0.30	0	0	1690	846
Connect	0.01	0.30	50	38	192	45
Mushroom	0.01	0.30	144	23	282	24
Chess	0.01	0.30	50	33	91	33
A	0.05	0.10	24	7	28	7
B	0.05	0.10	394	27	493	22
C	0.05	0.10	45	12	207	37
D	0.05	0.10	0	0	151	86
E	0.05	0.10	0	0	920	461
Connect	0.05	0.10	56	35	211	45
Mushroom	0.05	0.10	85	19	129	20
Chess	0.05	0.10	53	27	83	27

Conclusions

- OOIMASP generates same association rules irrespective of the order of items
- OOIMASP will not generate contradictory rules
- OOIMASP outperforms MASP in terms of both the metrics i.e. total rules and length of longest rule
- OOIMASP requires more computational resources as compared to MASP algorithm