What's so special is proposed in the paper ?

Mining of **structured data** to find **useful patterns** by **Association Rule Mining**. Traditional method is available for Association Rule Mining (**T-ARM)**. A new approach is proposed in the paper to extract **Mostly Associated Sequential Patterns.**

Why new approach ?

- The proposed approach requires **less computational resources** in terms of **time** and **memory** requirements.
- **No need to search the entire lattice of item combinations.**
- **Pruning process** is **not required**.
- **K-item patterns** can be obtained taking the advantage of **sub-lattice property of MASPs**.
- **Tree (MASP)** formation take place which helps in the **visual analysis of data**.

What does ARM do ?

Discover hidden rules among **enormous pattern combination** based on **individual** and **conditional** frequencies.

T-ARM

- Generate **all possible patterns** from data while **pruning out non-frequent ones**.
- **Produce rules** from **frequent patterns**.
- Apply some **interesting measures** to obtain **interesting rules** that can be used in **decision making**.

General ARM Workflow

- Data
- Preprocess - localize, summarize and transform data
- C-Generator - finds candidate patterns
- Pruning - throw less frequent patterns
- R-Generator - generate k-item rules
- R-Filter - find interesting rules

(ARM algorithms varies because of constraints)

Default constraints: minimum support and minimum confidence
If **only default constraint** is used then **exhaustive search** will be required to **find interesting rules**.

MASP

In case of **MASPs no** need to search **entire lattice** because **one more constrain**t other than the **default** has been imposed.

MASPs imposes **interestingness** constraint on patterns to **detect highest co-occurence** without searching all possible combinations. During search process, **MASP tree** will be formed.

MASP has **sub-lattice** property that reveals **k-item rules** from MASP.

MASP+

T-ARM can be conducted within small dataset of each MASP. The combination of rules obtained from **T-ARM** and **MASP** give rise to **MASP+** approach.

Terminology

Transactional Data : A set of **attribute-value** pairs. **e.g.** t = {Age:Young, Day:Sunday}

A rule in the form **A -> B** is an association rule.

Support : Probability of occurence of **AUB** together in the complete data set.

Confidence : The conditional probability P(B|A) i.e. probability of getting B when A is given.

Steps to detect interesting rules

Data must be in the form

A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**
V11	V21	V31	V41	V51	V61	V71	V81	V91
V11	V22	V32	V41	V51	V62	V71	V81	V92

Where each cell store a value of the corresponding attribute.

How to generate MASP tree ?

Notation

Let X and Y are two sets then
X (Y implies X is a subset of Y
X\Y implies set of elements of X which are not in Y
|X| = number of entries in set X
If D is a data table:
One can say that it is composed of **cells $C_{ri} = (A_i = V_{ij})$** where r = 1, 2, ….., |D| and i = 1,2,....., |A|. J = 1, 2, ,3,....., |unique entries in the ith column|

An item I is unique attribute-value pair **$I = V_{ij}$**

Condition for MASP

A set M = {I_1, I_2, I_3, ……, I_k, ……., I_K} will be **MASP** iff **for all k belongs to** {1, 2, 3, ……., K}

1. $P(I_1 , I_2, I_3, ……., I_k)$ >= (threshold value of support)
2. $P(I_k | I_1 , I_2, I_3, ……., I_{k-1})$ >= (threshold value of confidence)

3. $P(I_k | I_1, I_2, I_3, \ldots\ldots, I_{k-1})$ must be maximum

Block (M) and Counter Block (M)

M = $\{V_{11}, V_{23}, V_{42}\}$
Attr(D) = $A_1, A_2, A_3, A_4, A_5, A_6$
B(M) = {select A3, A5, A6; FROM D; where (A1 = V11 && A2 = V23 && A4 = V42)}
CB(M) = {select A3, A4, A5, A6; FROM D; where (A1 = V11 && A2 = V23 && A4 $\neq$ V42)}

ROOT

A1	A2	A3	A4	A5
V11	***V21***	***V31***	***V41***	***V51***
V11	***V21***	***V32***	***V41***	***V53***
V11	***V22***	***V31***	***V42***	***V52***
V11	***V21***	***V32***	***V42***	***V52***
V11	***V22***	***V32***	***V43***	***V52***
V11	***V22***	***V32***	***V43***	***V52***
V12	**V22**	**V31**	**V44**	**V51**
V12	**V22**	**V31**	**V44**	**V51**
V12	**V23**	**V31**	**V45**	**V51**
V12	**V23**	**V32**	**V46**	**V51**

V11'

A1	A2	A3	A4	A5
V12	V22	V31	V44	V51
V12	V22	V31	V44	V51
V12	V23	V31	V45	V51

V12	V23	V32	V46	V51

V11

A3	**A2**	**A4**	**A5**
V32	***V21***	***V41***	***V53***
V32	***V21***	***V42***	***V52***
V32	***V22***	***V43***	***V52***
V32	***V22***	***V43***	***V52***
V31	**V21**	**V41**	**V51**
V31	**V22**	**V42**	**V52**

V32'

A3	**A2**	**A4**	**A5**
V31	V21	V41	V51
V31	V22	V42	V52

V32

A2	**A4**	**A5**
V21	*V41*	*V53*
V21	*V42*	*V52*
V22	*V43*	*V52*
V22	*V43*	*V52*

How to generate rules using MASP Tree ?

If MASP **M = {I_1, I_2, I_3, …. , I_K}** then rules will be

(I_1, I_2, I_3, …. ,I_{K-1}) -> I_K
(I_1, I_2, I_3, …. ,I_{K-2}) -> I_{K-1}
(I_1, I_2, I_3, …. ,I_{K-3}) -> I_{K-2}
……..
……..
……..
I_1 -> I_2

Additional rule
(I_1, I_2, ……, I_{k-1}) -> (I_k, I_{k+1}, ……, I_K) satisfies the minimum support condition. If it will satisfy the minimum confidence then it will be included in the rule set.

MASP+ patterns

The combination of **MASP and** the **rules** obtained **from its block** is also a **rule**.

MASP **M = {V11, V41, V31}**
Rules from its block {**V21, V61 -> V51; V22 -> V53**}

New rules {((**V11, V41, V31**), **V21, V61**) **-> V51,** ((**V11, V41, V31**), **V22**) **-> V53**}