Multivariate Time Series Association Analysis for Petroleum Big Data

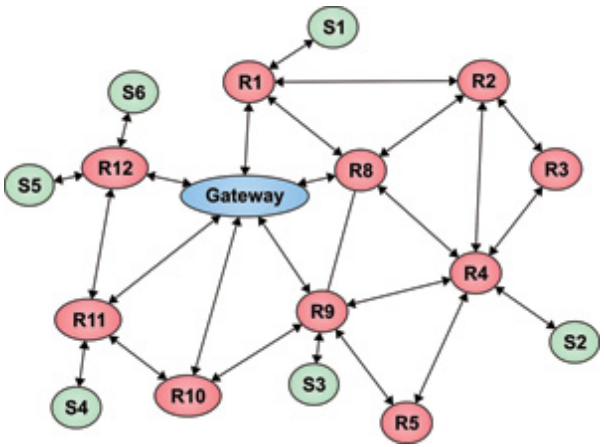
Fan Min (闵帆)

Lab of Machine Learning, Southwest Petroleum University
www.fansmale.com

October 16, 2017

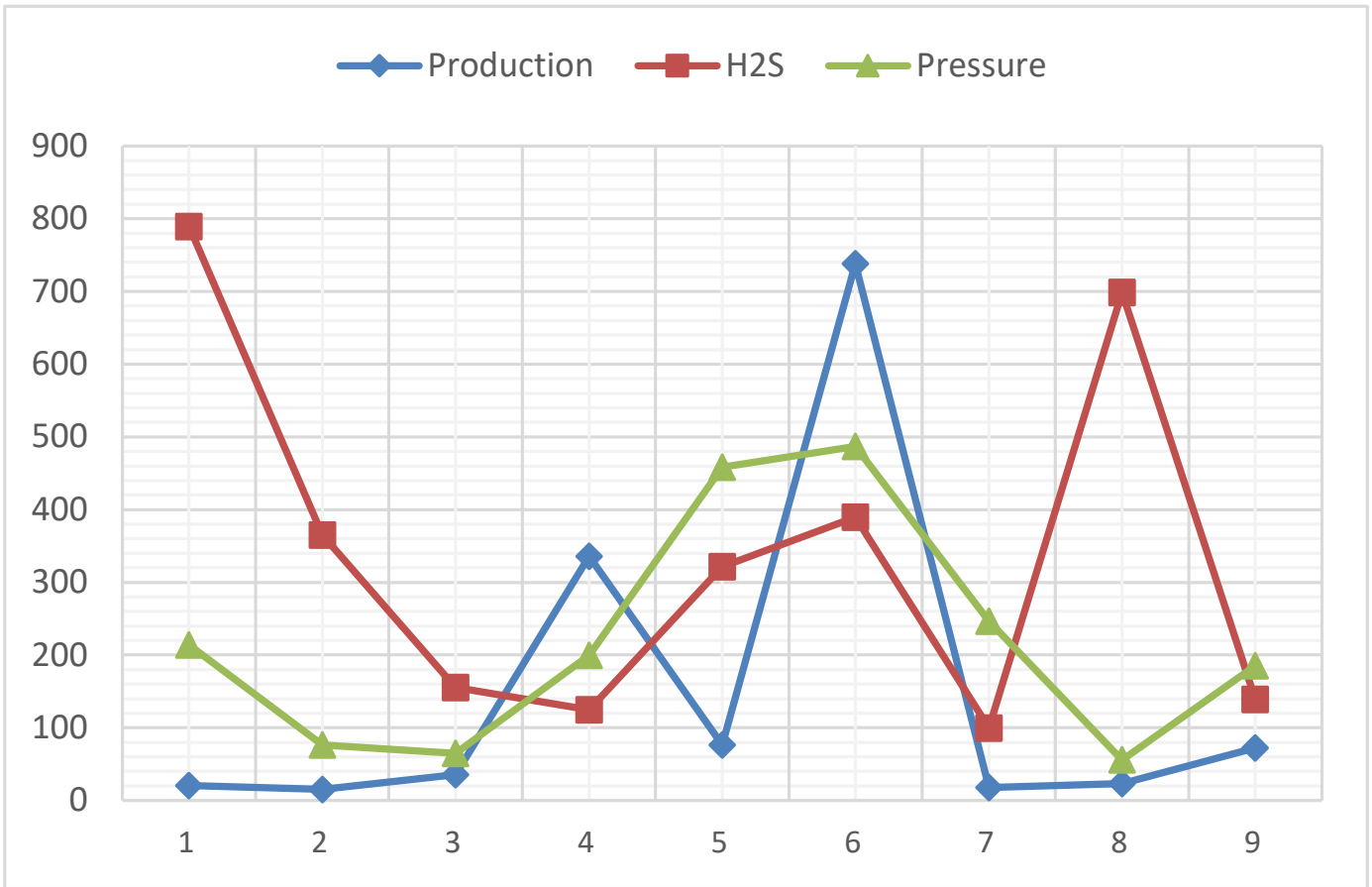
Contents

1.1 Anomaly detection



- Various sensor networks.
- Monitoring data are collected automatically.
- Anomalies should be predicted.

1.2 Multivariate time series



1.3 Multivariate time series information system

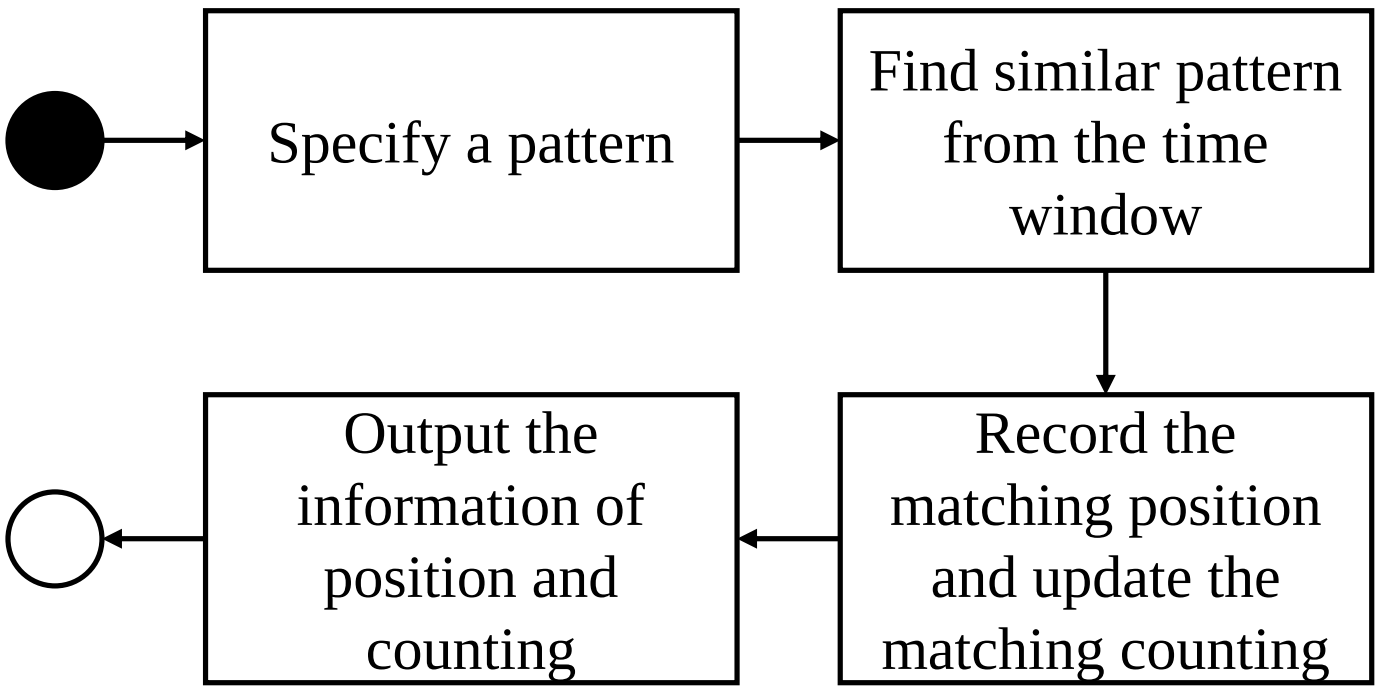
Time series	Production	H_2S	Pressure
20/4/2016	L	H	N
21/4/2016	L	N	L
22/4/2016	N	L	L
23/4/2016	N	L	N
24/4/2016	H	N	H
25/4/2016	H	N	H
26/4/2016	L	L	N
27/4/2016	N	H	L
28/4/2016	H	L	N

2. Pattern matching and discovery

2.1 Problem statement

- Assumption: The time series data is synchronous.
- Input: (1) $S = (T, A, V = \cup V_{a \in A}, f : T \times A \rightarrow V)$; (2) Pattern p; (3) Similarity threshold η; (4) Size of time window L.
- Output: (1) The counting of p occurring; (2) All occurrence position of p in S.

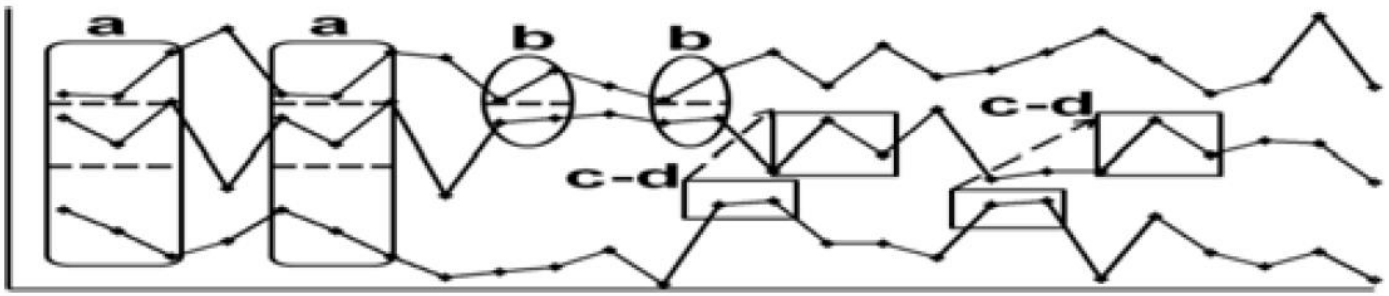
2.1 Pattern matching



2.1 Pattern matching

Time series	Production	H_2S	Pressure	
20/4/2016	L	H	N	
21/4/2016	L	N	L	
22/4/2016	N	L	L	p_3
23/4/2016	N	L	N	p_3
24/4/2016	H	N	H	p_2
25/4/2016	H	N	H	p_2
26/4/2016	L	L	N	
27/4/2016	N	H	L	p_1
28/4/2016	H	L	N	

2.1 Pattern matching



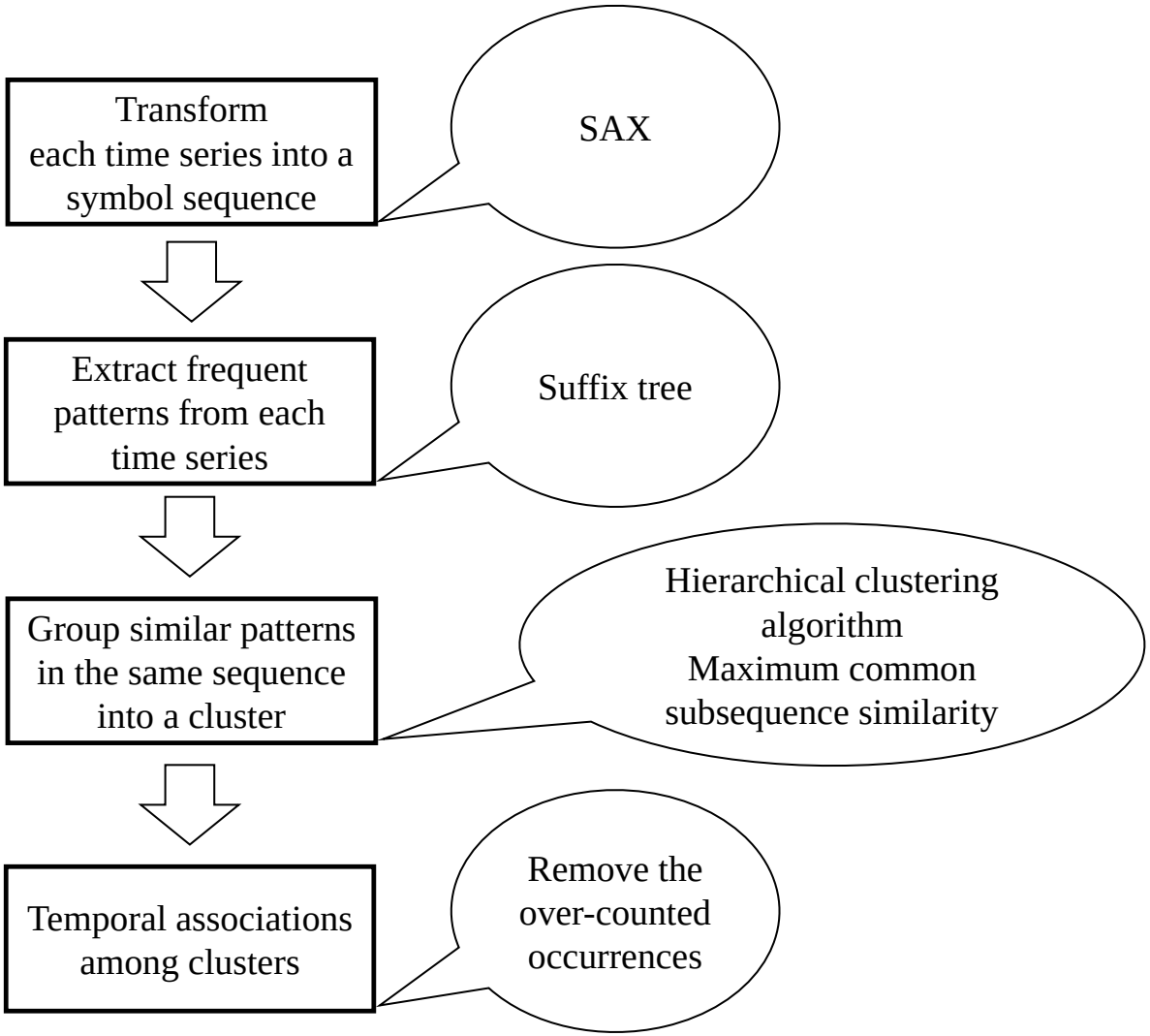
2.1 Pattern matching

- Question:
  What can we do when we cannot specify a pattern?
- Answer:
  Pattern discovery.

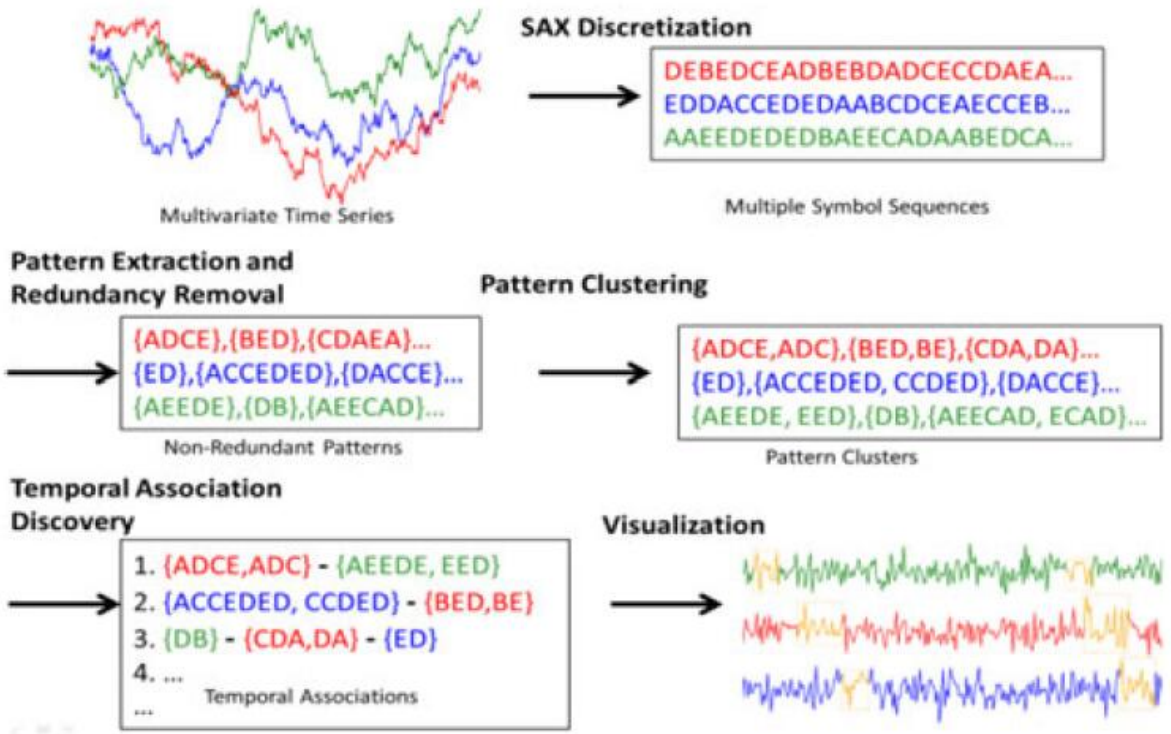
2.2 Pattern discovery

- Assumption: The time series data is synchronous.
- Input: (1) $S = (T, A, V = \cup V_{a \in A}, f : T \times A \rightarrow V)$; (2) Support threshold δ.
- Output: Temporal associations.
- Constraint: The supports of temporal associations are not less than δ.

2.2 Pattern discovery



2.2 Pattern discovery

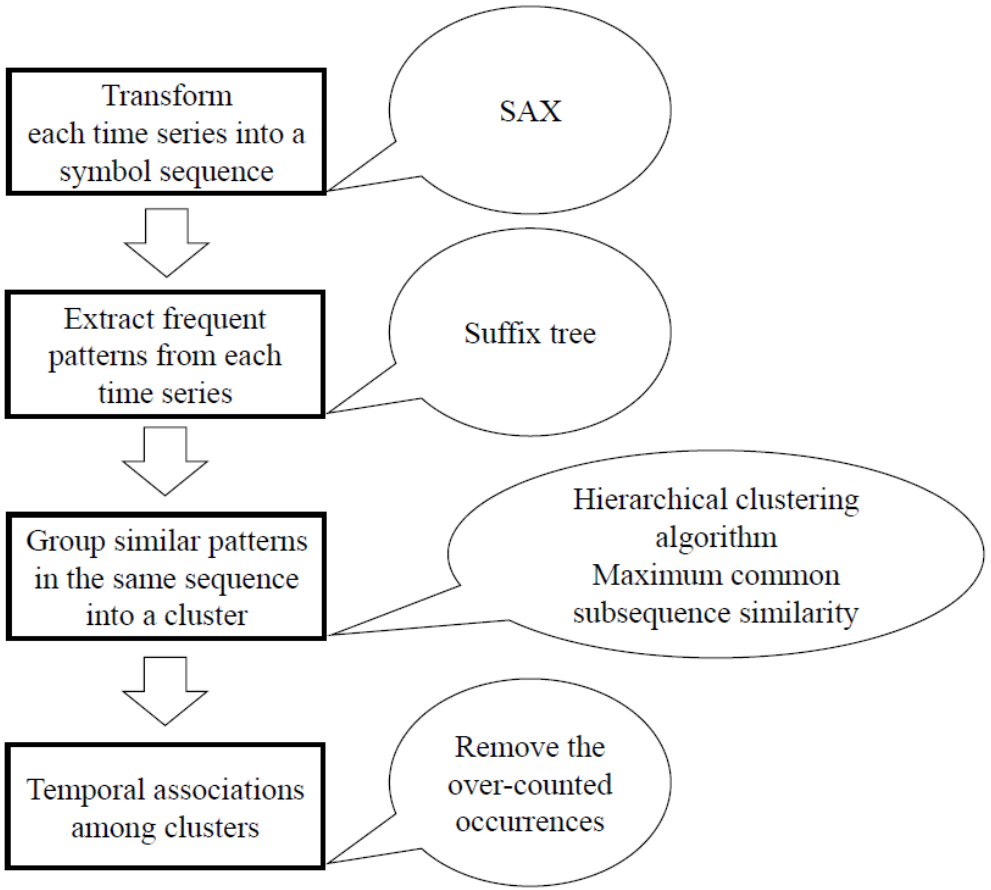


[Zhuang et al.(2014)Zhuang, Li, and Wong]

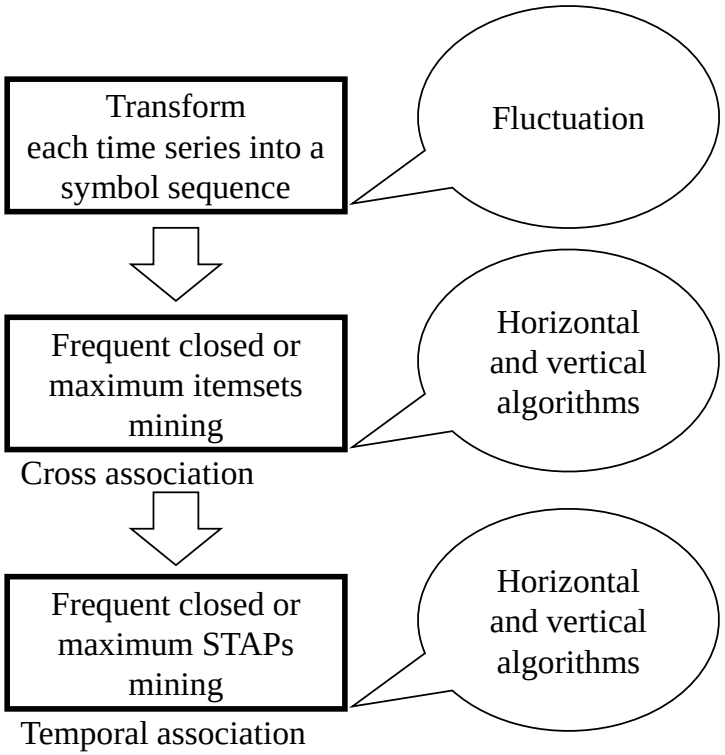
3.1 Problem statement

- Assumption: The time series data is synchronous.
- Input: (1) $S = (T, A, V = \cup V_{a \in A}, f : T \times A \rightarrow V)$; (2) Itemset support threshold δ_w; (3) Sequence support threshold δ_d.
- Output: Frequent STAPs.
- Constraint: (1) The supports of itemsets are not less than δ_w; (2) The supports of sequences are not less than δ_d.

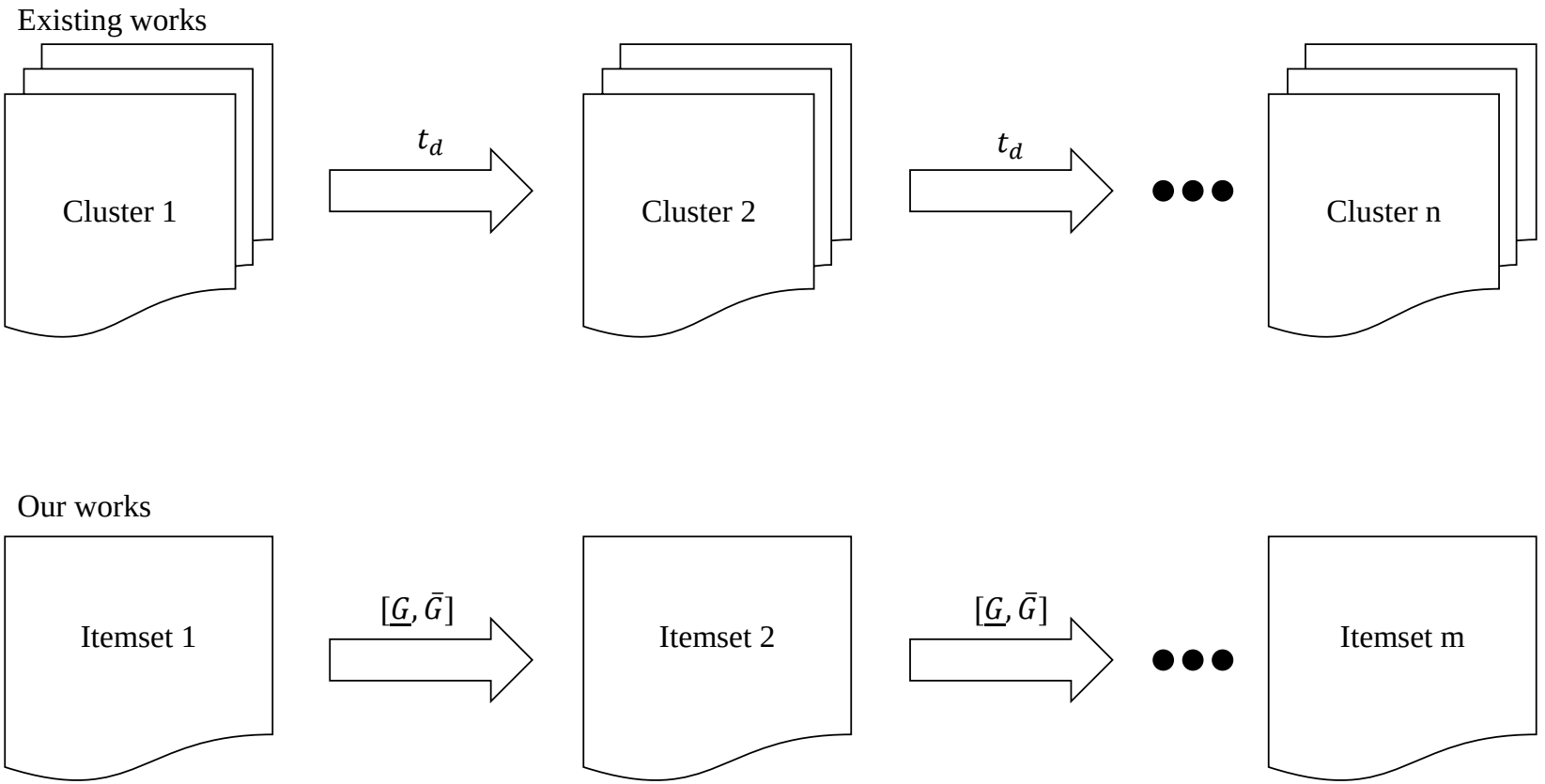
3.2 Framework of STAP



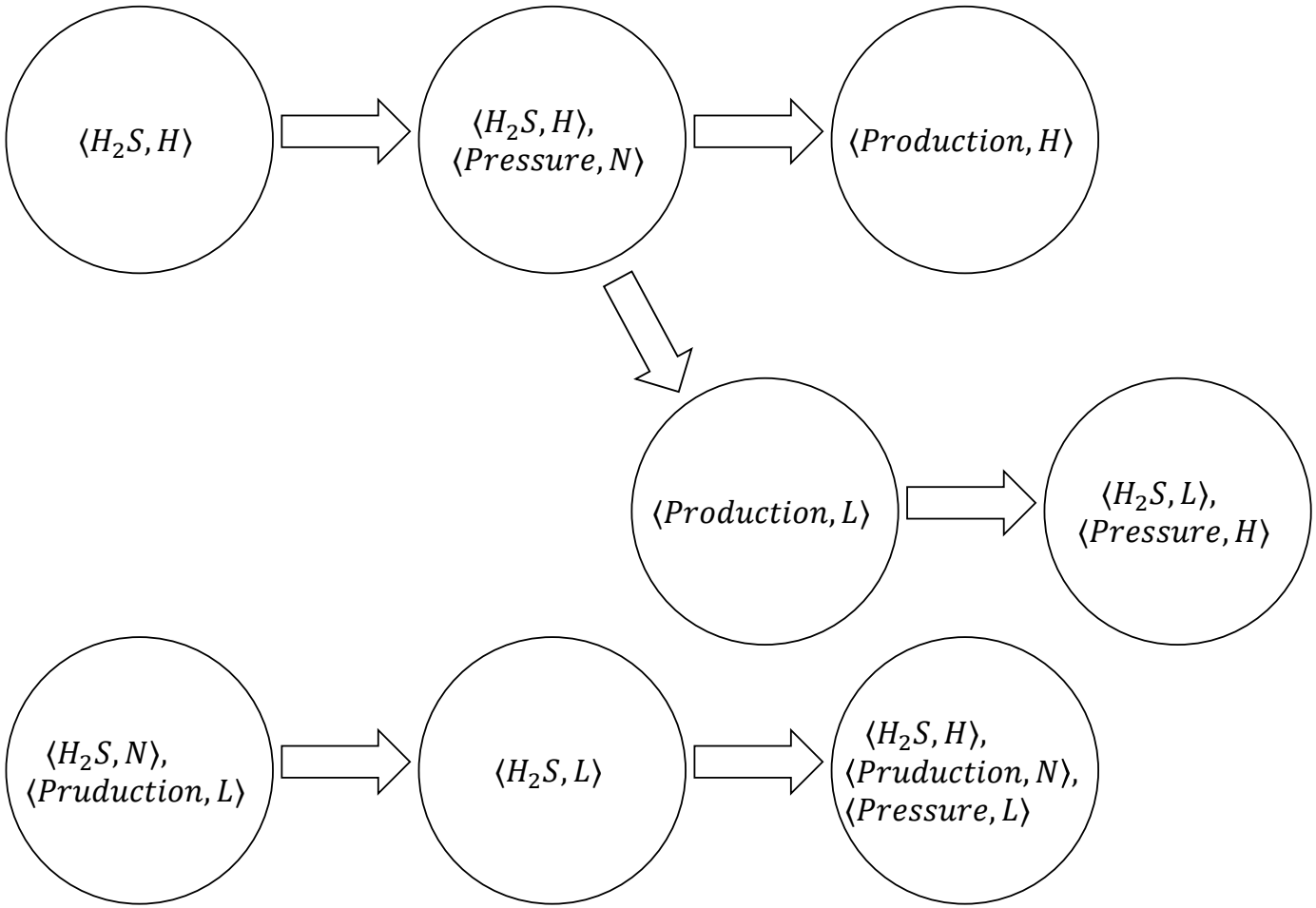
Cross and temporal association



Type of STAP


Existing works
Cluster 1
t_d
Cluster 2
t_d
Cluster n
Our works
Itemset 1
$[\underline{G}, \bar{G}]$
Itemset 2
$[\underline{G}, \bar{G}]$
Itemset m

Example of STAP



Discussions

- Semantics explanation of patterns
- Feature extracting, Occurrence counting

Any question?

Welcome to www.fansmale.com!


D. E. H. Zhuang, G. C. L. Li, A. K. C. Wong, Discovery of Temporal Associations in Multivariate Time Series, IEEE Transactions on Knowledge & Data Engineering 26 (12) (2014) 2969–2982.