UNIVERSITY OF TORONTO | Engineering

Sequence Patterns Mining with Generating unit data

Centre for Maintenance Optimization and Reliability Engineering (C-MORE)

Objective

- There are some huge **data** produced through the **maintenance** of hydroelectric and fossil generating units.

UnitEvent	GDID	StateCode	ForcedOutageTyp	SynchronousConden	CommonMode	OCIDGE	Amplificat	Amplificat	AuxiliaryN	EodIndica	OutageTy	Comments	StartDateTime	FinishDateTime
15811437	HGU0719	21	1	NULL	1	G105200	NULL	NULL	NULL	NULL	NULL	load rejection	6/20/2017	6/20/2017
15811438	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		6/20/2017	6/22/2017
15811439	HGU0719	21	1	NULL	1	G105200	NULL	NULL	NULL	NULL	NULL	line trip cause	6/22/2017	6/24/2017
15811440	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		6/24/2017	6/24/2017
15811441	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		6/24/2017	7/9/2017
15811442	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		7/9/2017	7/9/2017
15811443	HGU0719	24	NULL	NULL	NULL	G129620	NULL	NULL	NULL	NULL	NULL	Penstock drair	7/9/2017	7/13/2017
15811444	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		7/13/2017	8/10/2017
15811445	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/10/2017	8/10/2017
15811446	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/10/2017	8/19/2017
15811447	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/19/2017	8/19/2017
15811448	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/19/2017	9/2/2017
15811449	HGU0719	24	NULL	NULL	NULL	G129620	NULL	NULL	NULL	NULL	NULL	Install Penstoc	9/2/2017	9/12/2017
15811450	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/12/2017	9/16/2017
15811451	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/16/2017	9/16/2017
15811452	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/16/2017	9/20/2017
15811453	HGU0719	24	NULL	NULL	NULL	G142260	NULL	NULL	NULL	NULL	NULL	AVR Inspectio	9/20/2017	9/20/2017
15811454	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/20/2017	9/20/2017
15811455	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/20/2017	12/31/2017
15812121	HGU0693	21	1	NULL	NULL	G142171	NULL	NULL	NULL	NULL	F	Unit remains c	1/1/2017	12/31/2017
15812125	HGU0682	22	NULL	NULL	NULL	G141100	NULL	NULL	NULL	NULL	M	Closed By Yea	1/1/2017	12/31/2017
15812126	HGU0689	21	1	NULL	NULL	G142171	NULL	NULL	NULL	NULL	NULL	Unit remains c	1/1/2017	12/31/2017

Objective

- **Can Predict** the future behavior and future events of generating units and their components?
- Can **understand** the system health better?

- In this presentation:
  - Introduce ideas on applying **predictive and descriptive analytics of Machine Learning** for maintenance practices of generating units.

Case study

We have **continuous records** of the operating and outage data of each unit. (2013-2017)

UnitEvent	GDID	StateCode	ForcedOutageTyp	SynchronousConden	CommonMode	OCIDGE	Amplificat						ishDateTime
15811437	HGU0719	21	1	NULL	1	G105200	NULL						6/20/2017
15811438	HGU0719	11	NULL	NULL	NULL	NULL	NULL						6/22/2017
15811439	HGU0719	21	1	NULL	1	G105200	NULL						6/24/2017
158114													6/24/2017
158114													7/9/2017
158114													7/9/2017
158114													7/13/2017
158114													8/10/2017
158114													8/10/2017
158114								NULL	NULL	NULL	NULL	8/10/2017	8/19/2017
158114								NULL	NULL	NULL	NULL	8/19/2017	8/19/2017
158114								NULL	NULL	NULL	NULL	8/19/2017	9/2/2017
158114								NULL	NULL	NULL	NULL	9/2/2017	9/12/2017
15811450	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	9/12/2017	9/16/2017
15811451	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	9/16/2017	9/16/2017
15811452	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	9/16/2017	9/20/2017
15811453	HGU0719	24	NULL	NULL	NULL	G142260	NULL	NULL	NULL	NULL	NULL	9/20/2017	9/20/2017
15811454	HGU0719	14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	9/20/2017	9/20/2017
15811455	HGU0719	11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	9/20/2017	12/31/2017
15812121	HGU0693	21	1	NULL	NULL	G142171	NULL	NULL	NULL	NULL	F	1/1/2017	12/31/2017
15812125	HGU0682	22	NULL	NULL	NULL	G141100	NULL	NULL	NULL	NULL	M	1/1/2017	12/31/2017
15812126	HGU0689	21	1	NULL	NULL	G142171	NULL	NULL	NULL	NULL	NULL	1/1/2017	12/31/2017

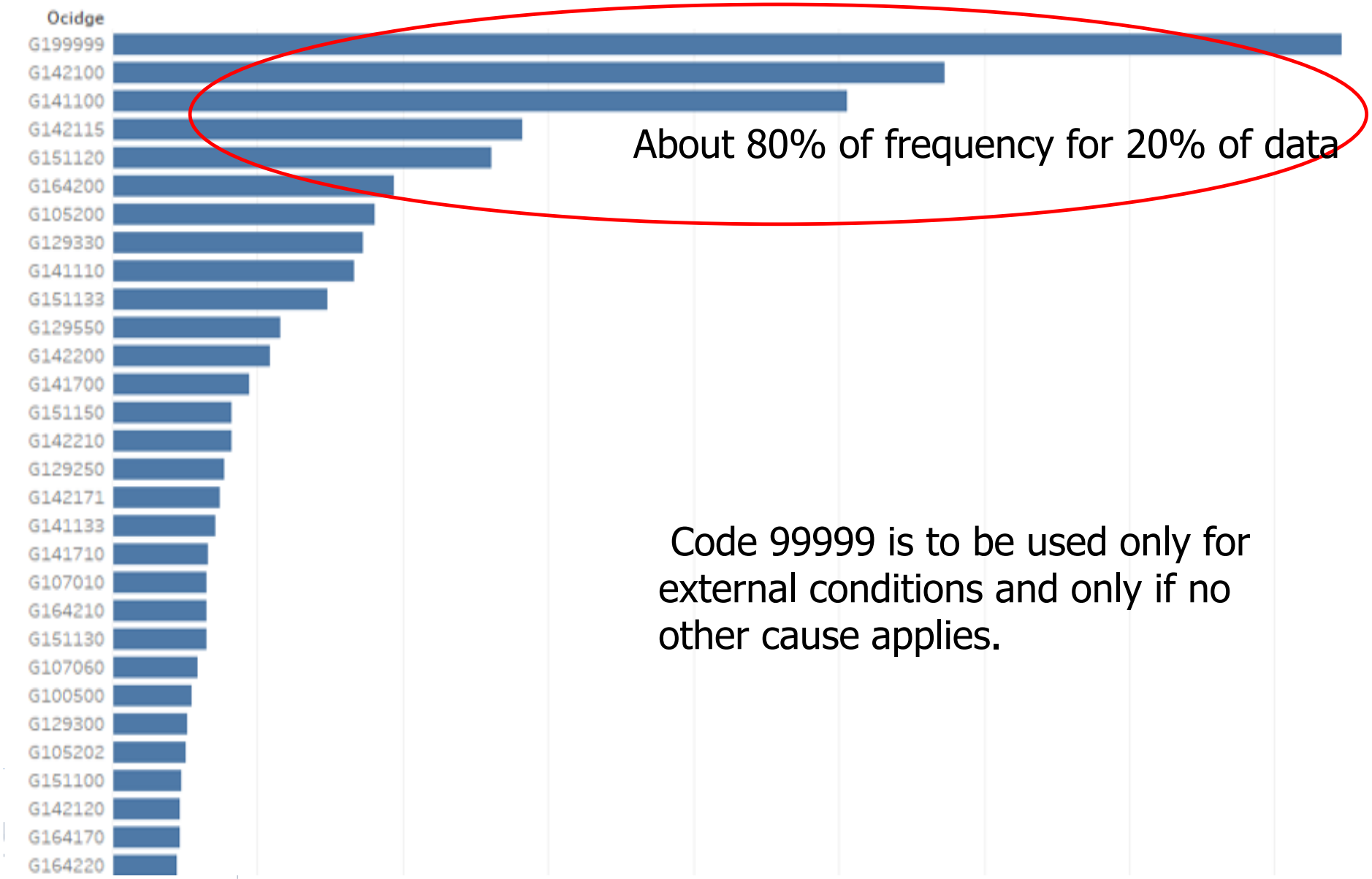
State Code

Identifies the state of the unit during the indicated time interval

- Forced Outage (21)
- Maintenance Outage (24)
- Planned Outage (25)

- **OCIDGE** codes were developed from the System Classification Index (SCI) used to identify **equipment**, **systems**, and conditions.

OCIDGE: Outage Component Codes



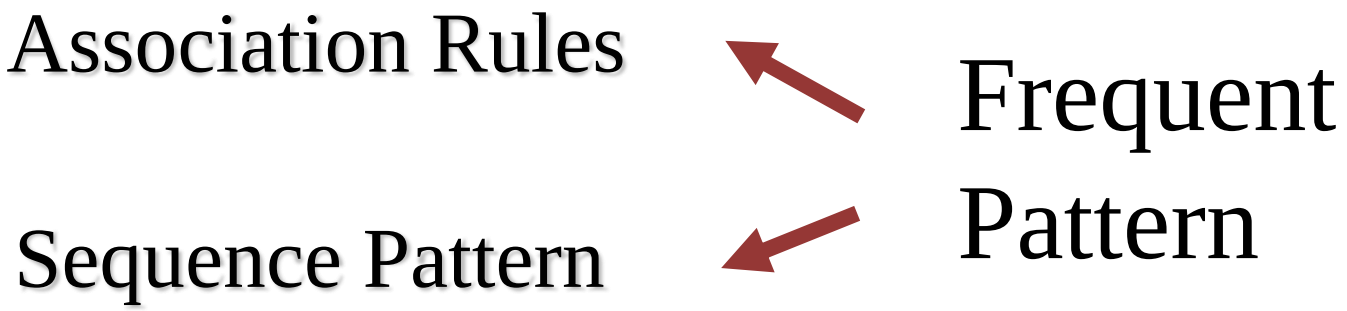
About 80% of frequency for 20% of data

Code 99999 is to be used only for external conditions and only if no other cause applies.

Problem Definition

- Is there any relationship among **Outage Component Codes** with more frequently?

Frequent pattern



What Is Frequent Pattern Analysis?

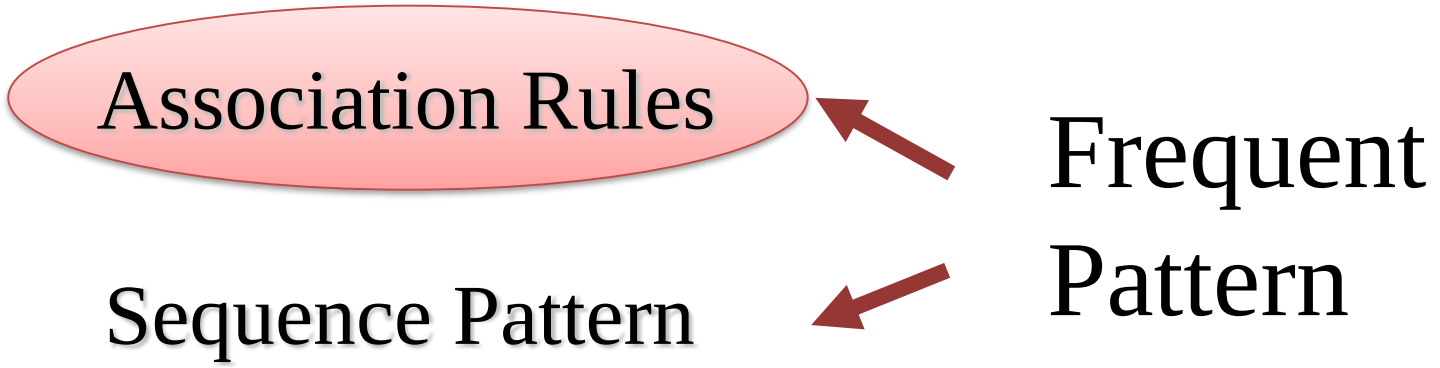
- Frequent pattern: a pattern (a set of items, subsequences, etc.) that occurs **frequently** in a data set.

- Motivation: Finding inherent rules in data

  For example:

  - What products were often purchased **together**?
  - What are the **subsequent purchases** after buying a PC?

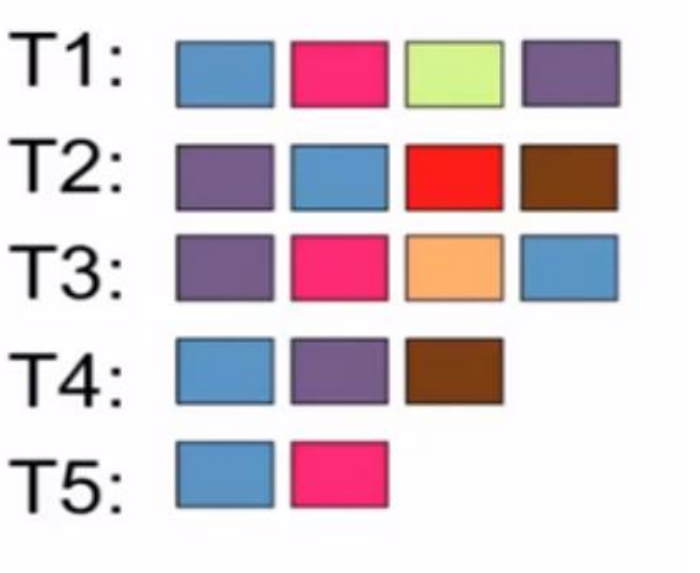
Frequent pattern

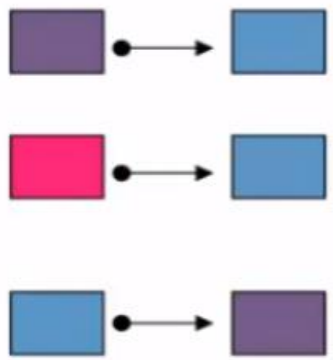


Association rule mining

- Proposed by Agrawal et al in 1993.
- It is an important **Data Mining** model studied extensively by the database and data mining community.
- Initially used for Market Basket Analysis to find how items purchased by customers are related. (frequent pattern)

Support & Confidence





Support: 4/5; Conf: 1

Support: 3/5; Conf: 1

Support: 4/5; Conf: 4/5

support, *s*, probability that a transaction contains X and Y

confidence, *c*, conditional probability that a transaction having X also contains *Y*

Association rules algorithms

- Scalable frequent pattern mining methods
  - Apriori (Candidate generation & test)
  - Projection-based (FPgrowth, CLOSET+, ...)
  - Vertical format approach (ECLAT, CHARM, ...)

Approaches and Data preparation

Consider only these state codes:

- Forced Outage (21) or Maintenance Outage (24) to a Planned Outage (25)

Delete some OCIDGE

- 99999
- 42100(Generator)

Used **Appriori** algorithm

- Number of rules =67
- Maximum number of antecedents: 5
- Minimum antecedent support (%): 20.0
- Minimum rule confidence (%): 70.0

Output

Generator And Auxiliaries →Generator Power Transformers (c=71.5%)

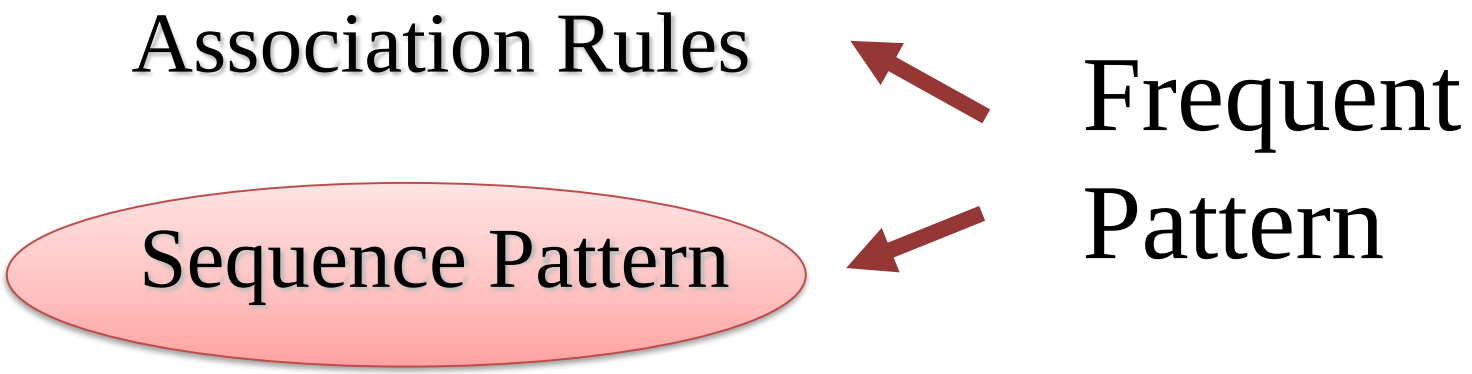
Circuit Breakers - Generator Voltage → Generator Power Transformers (c=80%)

Output

- (Excitation) → Generator And Auxiliaries (c=74%)
- (Excitation) → Generator Power Transformers (c=71%)
- Generator And Auxiliaries →Generator Power Transformers (c=71.5%)
- Circuit Breakers - Generator Voltage → Generator Power Transformers (c=80%)

- (Circuit Breakers - Generator Voltage) and (Generator Power Transformers) → (Brushes And Brush Rigging) (c=72%)
- (Headgates) → Brushes And Brush Rigging (c=73%)

Frequent pattern



Sequence Databases

- A sequence database consists of ordered elements or events
- Transaction databases vs. sequence databases

A transaction database

TID	itemsets
10	a, b, d
20	a, c, d
30	a, d, e
40	b, e, f

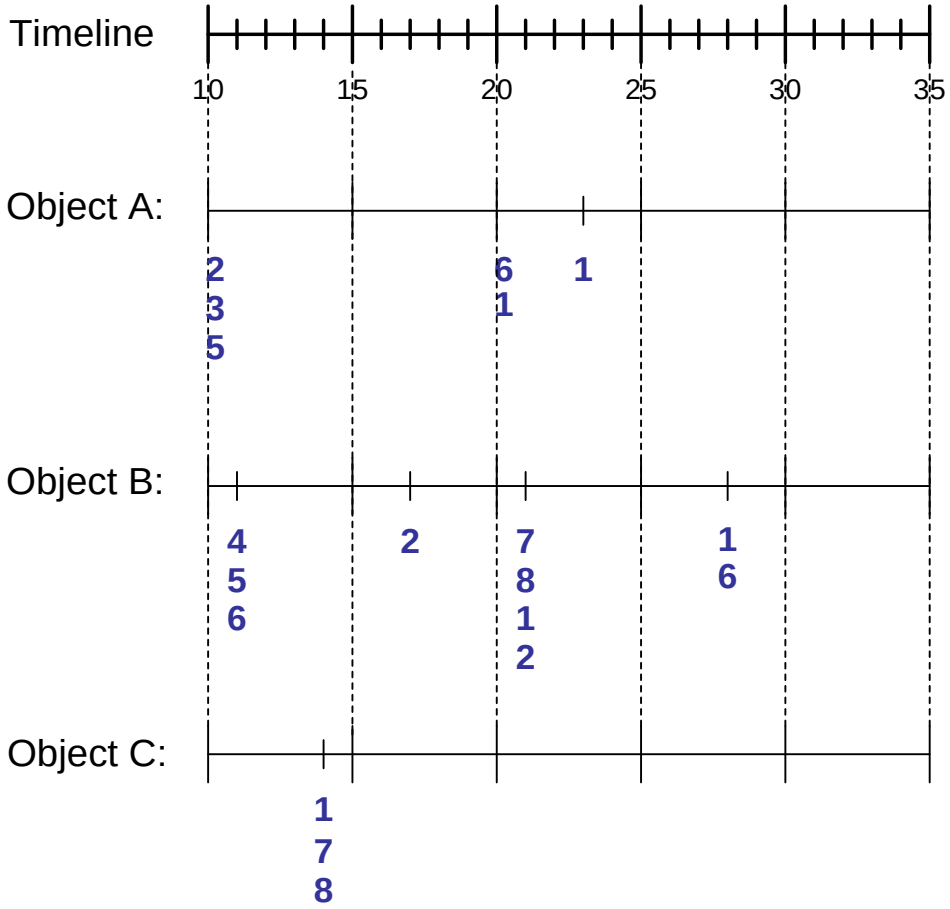
A sequence database

SID	sequences
10	<a(abc)(ac)d(cf)>
20	<(ad)c(bc)(ae)>
30	<(ef)(ab)(df)cb>
40	<eg(af)cbc>

Sequence Data

Sequence Database:

Object	Timestamp	Events
A	10	2, 3, 5
A	20	6, 1
A	23	1
B	11	4, 5, 6
B	17	2
B	21	7, 8, 1, 2
B	28	1, 6
C	14	1, 8, 7



Sequential Pattern Mining:

- Given:
  - a database of sequences
  - a user-specified minimum support threshold, *minsup (minimum support)*

- Task:
  - Find all subsequences with support ≥ *minsup*

Studies on Sequential Pattern Mining

- Apriori-based method: GSP (Generalized Sequential Patterns: Srikant & Agrawal [EDBT'96])
- Pattern-growth methods: FreeSpan & PrefixSpan (Han et al.KDD'00; Pei, et al. [ICDE'01])
- Vertical format-based mining: SPADE (Zaki [Machine Leanining'00])
- Constraint-based sequential pattern mining (SPIRIT: Garofalakis, Rastogi, Shim [VLDB'99]; Pei, Han, Wang [CIKM'02])
- Mining closed sequential patterns: CloSpan (Yan, Han & Afshar [SDM'03])

Approaches and Data preparation

Consider only these state codes:

- Forced Outage (21) or Maintenance Outage (24) to a Planned Outage (25)

Delete some OCIDGE

- 99999
- 42100(Generator)

Number of Rules: 443
Number of Valid Transactions: 476
Minimum Support: 21.218%
Maximum Support: 98.739%
Minimum Confidence: 60.366%
Maximum Confidence: 99.099%

Output

(Antecedent) → (Consequent)

(Brushes And Brush Rigging) and (Circuit Breakers - Generator Voltage) → Brushes And Brush Rigging (c=78%)
(Brushes And Brush Rigging) and (Generator And Auxiliaries) → Brushes And Brush Rigging (c=70%)
(Brushes And Brush Rigging) and (Turbines) ⟶ Brushes And Brush Rigging(c=74%)
(Headgates) → Brushes And Brush Rigging(c=61.9%)
(Brushes And Brush Rigging) and (Generator Power Transformers) → Generator Power Transformers(c=73.9%)

(Generator Power Transformers) → Generator Power Transformers(c=66.9%)
(Brushes And Brush Rigging) → Generator Power Transformers(c=61.3%)
(Brushes And Brush Rigging) and (Maintenance Outage) → Generator Power Transformers(c=61.3%)
(Generator And Auxiliaries) and (Generator Power Transformers) → Generator Power Transformers (c=64.8)
(Circuit Breakers - Generator Voltage) → Generator Power Transformers (c=68.3)
(Turbines) → Generator Power Transformers (c=58.3)
(Transmission Limitations) → (Transmission Limitations) (c=62.4)

Output

(Antecedent) → (Consequent)

(Brushes And Brush Rigging) and (Circuit Breakers - Generator Voltage) → Brushes And Brush Rigging (c=78%)

(Generator And Auxiliaries) and (Generator Power Transformers) → Generator Power Transformers (c=64.8)

Output

(Antecedent) → (Consequent)

(Antecedent) → (Forced Outage)

(Brushes And Brush Rigging) and (Circuit Breakers - Generator Voltage) → Forced Outage (c=95%)
(Maintenance Outage) and (Bus Duct, Bus, Cable) → Forced Outage (c=89 %)

Output

(Antecedent) → (Consequent)

(Antecedent) → (Forced Outage)

(Brushes And Brush Rigging) and (Circuit Breakers - Generator Voltage) → Forced Outage (c=95%)

(Generator And Auxiliaries) and (Generator And Auxiliaries) → Forced Outage (c= 93.6%)

(Cooling Water Systems) and (Maintenance Outage) → Forced Outage (c=93.3 %)

(Maintenance Outage) and (Turbines) → Forced Outage (c=93.2 %)

(Governor System) and (Maintenance Outage) → Forced Outage (c= 92%)

(Generator Power Transformers) and (Forced Outage) → Forced Outage (c= 92%)

(Generator Power Transformers) and (Maintenance Outage) → Forced Outage (c= 88.9%)

(Maintenance Outage) and (Bus Duct, Bus, Cable) → Forced Outage (c=89 %)

Time Gap Problem

Are all sequence patterns interesting?

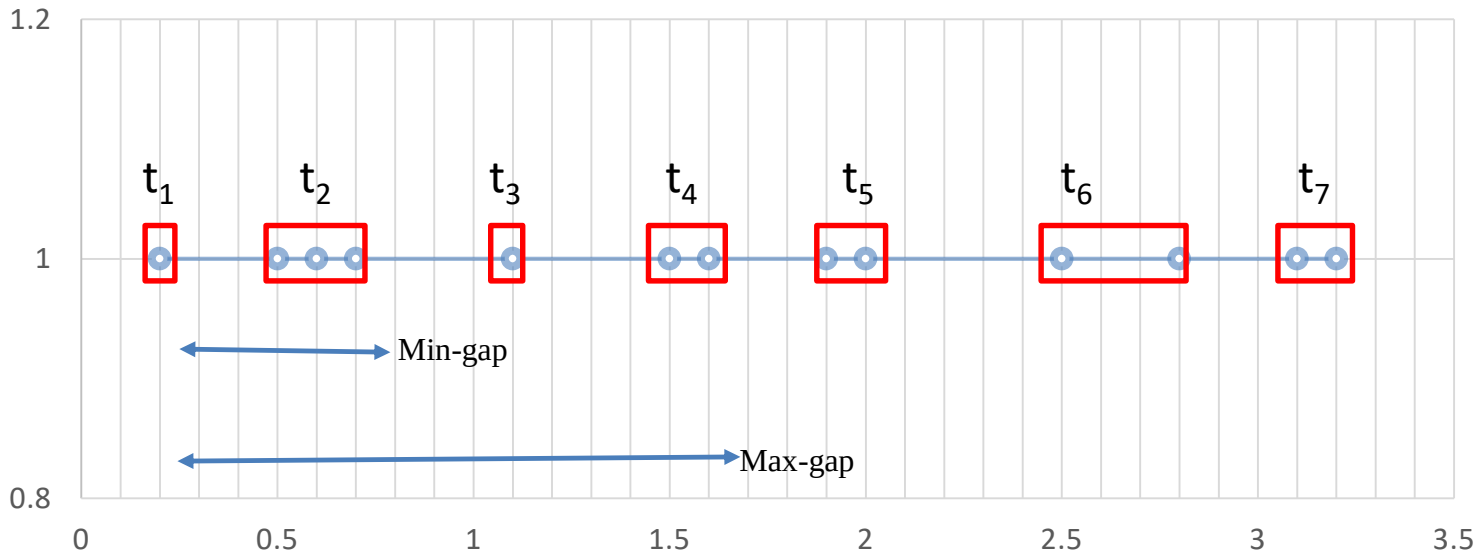
Time Gap Constraint in sequential pattern mining

- There are some challenges associated with this method, one of them is **Time Gap** constrain.

- In particular, when two events occur with a moderately long time gap, they cannot be considered as being part of the same sequence.

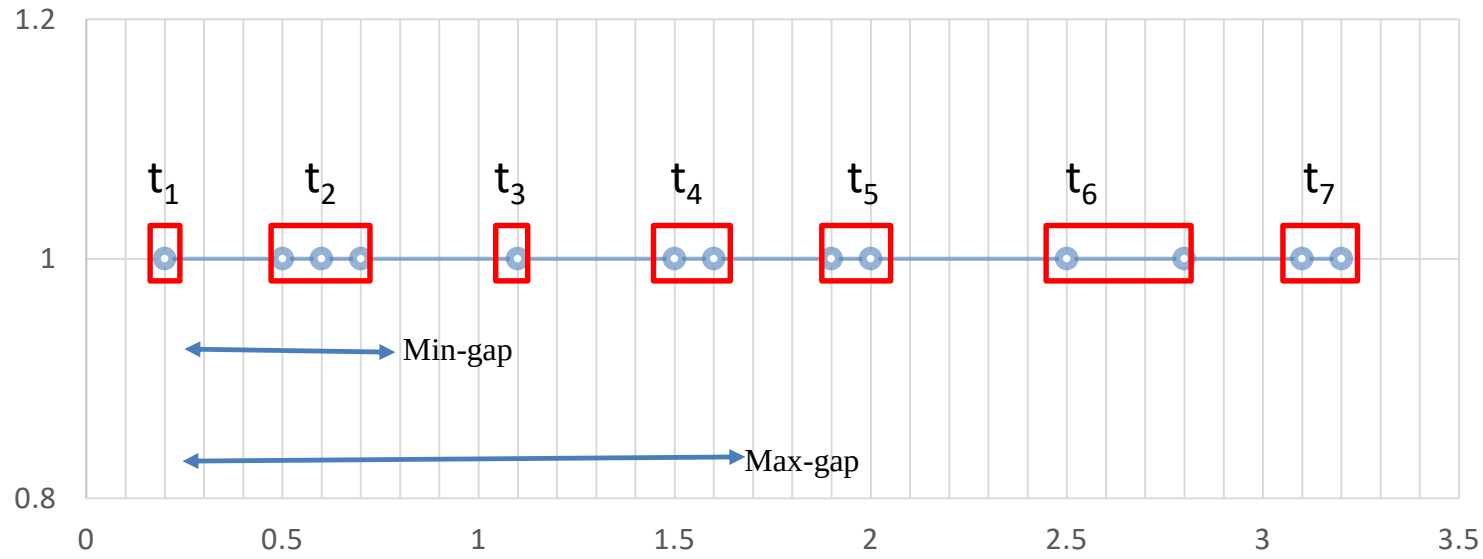
Time Gap Constraint in sequential pattern mining

- One of the important concern in Sequential pattern mining is the constraints of minimum and maximum gap between two consecutive events.



Constraints in sequential pattern mining

- A desired (time) gap between events in the discovered patterns is specified as a constraint:
- min gap ≤ gap ≤ max gap



Timing Constraints

max-gap = 2

Data sequence	Subsequence	Contained?
<{a,b} {c,d,e} {f,g} {h,i} {j}>	< {e} {i} >	Yes
<{a} {b} {c} {d} {e}>	< {a} {e} >	
<{a} {b,c} {d,e} {e,f}>	< {b} {d} {g} >	
<{a,b} {c} {d,e} {f,g} {h,i} {j,k}>	< {a,b} {k} >	

Timing Constraints

max-gap = 2

Data sequence	Subsequence	Contained?
<{a,b} {c,d,e} {f,g} {h,i} {j}>	< {e} {i} >	Yes
<{a} {b} {c} {d} {e}>	< {a} {e} >	No
<{a} {b,c} {d,e} {e,f}>	< {b} {d} {g} >	
<{a,b} {c} {d,e} {f,g} {h,i} {j,k}>	< {a,b} {k} >	

Timing Constraints

max-gap = 2

Data sequence	Subsequence	Contained?
<{a,b} {c,d,e} {f,g} {h,i} {j}>	< {e} {i} >	Yes
<{a} {b} {c} {d} {e}>	< {a} {e} >	No
<{a} {b,c} {d,e} {e,f}>	< {b} {d} {g} >	Yes
<{a,b} {c} {d,e} {f,g} {h,i} {j,k}>	< {a,b} {k} >	

Timing Constraints

max-gap = 2

Data sequence	Subsequence	Contained?
<{a,b} {c,d,e} {f,g} {h,i} {j}>	< {e} {i} >	Yes
<{a} {b} {c} {d} {e}>	< {a} {e} >	No
<{a} {b,c} {d,e} {e,f}>	< {b} {d} {g} >	Yes
<{a,b} {c} {d,e} {f,g} {h,i} {j,k}>	< {a,b} {k} >	No

Mining Sequential Patterns with Timing Constraints

- Approach 1:
  - Mine sequential patterns without timing constraints
  - Post-process the discovered patterns

- Approach 2:
  - Modify algorithm to directly prune candidates that violate timing constraints

Approaches and Data preparation

Consider only these state codes:

- Forced Outage (21) or Maintenance Outage (24) to a Planned Outage (25)

Delete some OCIDGE

- 99999
- 42100(Generator)

Max-GAP= 10 days
Min-Gap=0 day
Number of Rules:14
Number of Valid Transactions:476
Minimum Confidence:20.5
Maximum Confidence:100

output

(Antecedent) → (Consequent)

(Upstream Water Conditions) and (Upstream Water Conditions) → Trash Racks And Followers

Upstream Water Conditions , Trash Racks And Followers

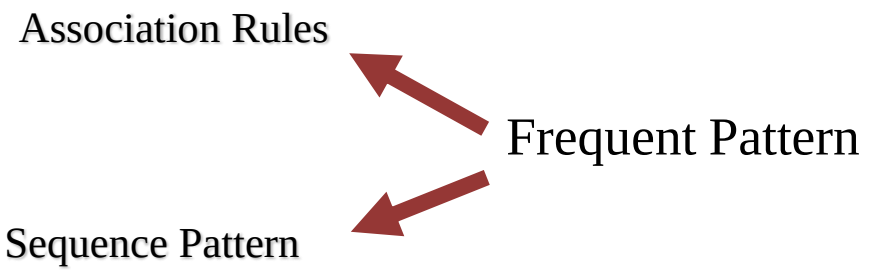
Exciter Transformer → Site Environment, Storms, Floods

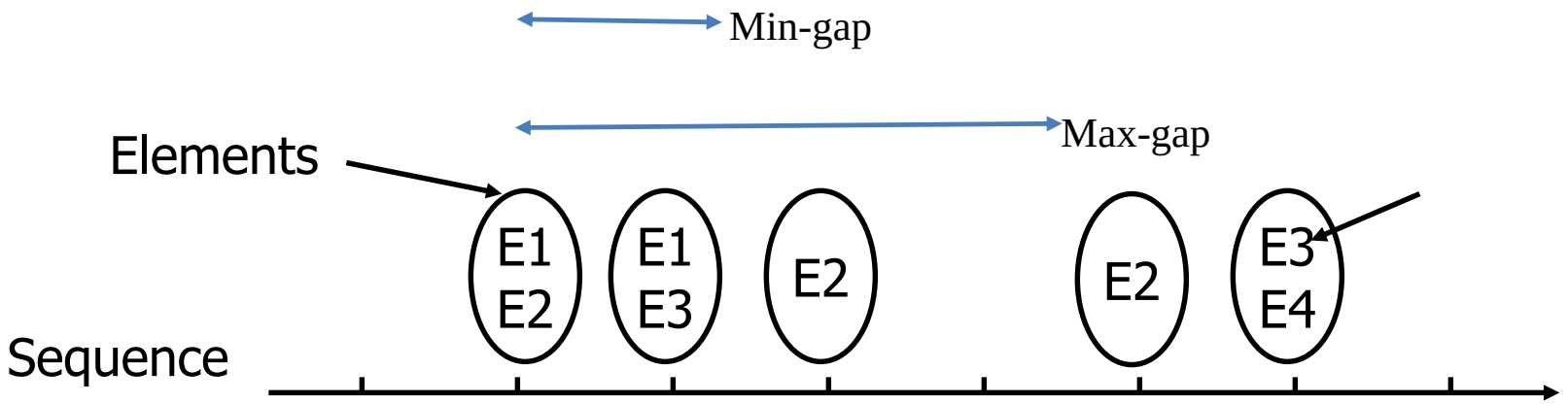
(Exciter Transformer) and (Site Environment, Storms, Floods) → Site Environment, Storms, Floods

(Channels & Tunnels) → (Channels & Tunnels)

Conclusion

UnitEvent	GDID	StateCode	ForcedOutage	Typ	Synchronous	Conden	CommonMode	OCIDes	Amplifica	Amplifica	Auxiliary	EodIndica	OutageTy	Comments	StartDateTime	FinishDateTime
15811437	HGU0719	21		1	NULL		1	G105200	NULL	NULL	NULL	NULL	NULL	load rejection	6/20/2017	6/20/2017
15811438	HGU0719	11	NUL		NULL	NULL		NULL	NULL	NULL	NULL	NULL	NULL		6/20/2017	6/22/2017
15811439	HGU0719	21		1	NULL		1	G105200	NULL	NULL	NULL	NULL	NULL	line trip cause	6/22/2017	6/24/2017
15811440	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		6/24/2017	6/24/2017
15811441	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		6/24/2017	7/9/2017
15811442	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		7/9/2017	7/9/2017
15811443	HGU0719	24	NULL		NULL	NULL	NULL	G129620	NULL	NULL	NULL	NULL	NULL	Penstock drair	7/9/2017	7/13/2017
15811444	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		7/13/2017	8/10/2017
15811445	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/10/2017	8/10/2017
15811446	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/10/2017	8/19/2017
15811447	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/19/2017	8/19/2017
15811448	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		8/19/2017	9/2/2017
15811449	HGU0719	24	NULL		NULL	NULL	NULL	G129620	NULL	NULL	NULL	NULL	NULL	Install Penstoc	9/2/2017	9/12/2017
15811450	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/12/2017	9/16/2017
15811451	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/16/2017	9/16/2017
15811452	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/16/2017	9/20/2017
15811453	HGU0719	24	NULL		NULL	NULL	NULL	G142260	NULL	NULL	NULL	NULL	NULL	AVR Inspectio	9/20/2017	9/20/2017
15811454	HGU0719	14	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/20/2017	9/20/2017
15811455	HGU0719	11	NULL		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL		9/20/2017	12/31/2017
15812121	HGU0693	21		1	NULL		NULL	G142171	NULL	NULL	NULL	NULL	F	Unit remains c	1/1/2017	12/31/2017
15812125	HGU0682	22	NULL		NULL		NULL	G141100	NULL	NULL	NULL	NULL	M	Closed By Yeai	1/1/2017	12/31/2017
15812126	HGU0689	21		1	NULL		NULL	G142171	NULL	NULL	NULL	NULL	NULL	Unit remains c	1/1/2017	12/31/2017





Thank you