CASPT2015

Conference on Advanced Systems in Public Transport
19-23 July 2015 | nhow hotel, Rotterdam, The Netherlands
Organized by Erasmus School of Economics and Rotterdam School of Management, Erasmus University

Mining the Cause of Delays in Urban Railways based on Association Rules

Hideyuki Yabuki *Tokyo Metro Co./ Chiba Inst. Tech*
Taku Ageishi *Chiba Inst. Tech.*
Norio Tomii *Chiba Inst. Tech.*

Tokyo Metro

CASPT2015

Conference on Advanced Systems in Public Transport
19-23 July 2015 | nhow hotel, Rotterdam, The Netherlands
Organized by Erasmus School of Economics and Rotterdam School of Management, Erasmus University

Tomii Lab. Chiba Institute of Technology.

Mining the Cause of Delays in Urban Railways based on Association Rules

We want to establish an approach to evaluate the effectiveness of the delay reduction countermeasures through data mining technique!

Tokyo Metro

Contents

- Background
- Research Objective
- Evaluation of delay reduction measures based on association rules
- Numerical Experiments
- Conclusions

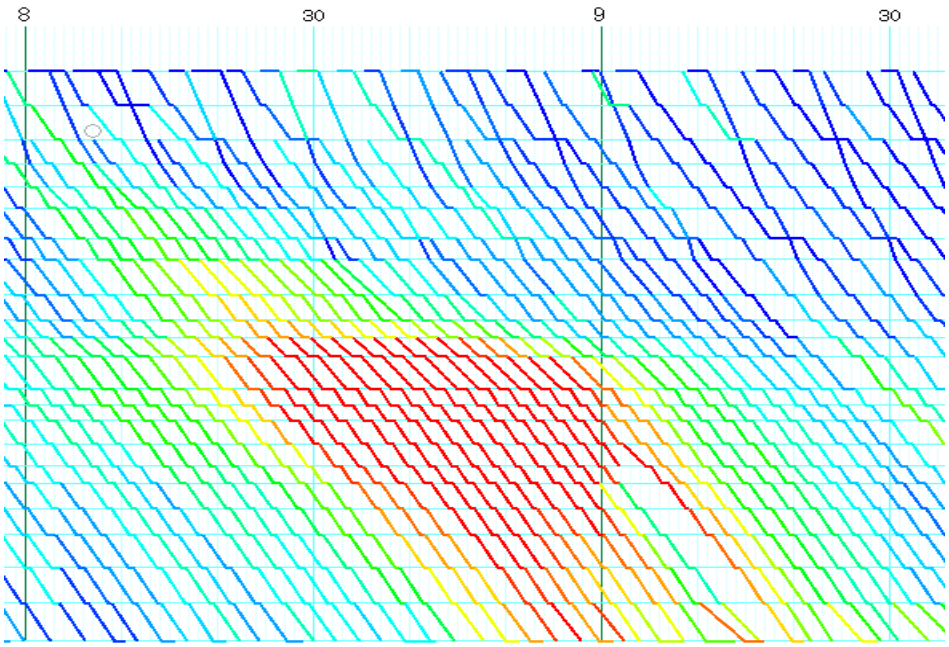
Background

Problems in urban railways in Japan

- Small delays (several minutes) often happen
- Delays propagate to other trains

Example

- < 1 minutes
- 1-3 minutes
- 3-5 minutes
- > 5 minutes



Background

Problems in urban railways in Japan

- Small delays (several minutes) often happen
- Delays propagate to other trains
- This is because
  - Trains are running densely

7	久 00		清 02		06	09	栗 12	栗 16	18	久 21	栗 25	半 28	31	清 33	栗 36	38	栗 41
	43	清 45	久 48	50	52	清 54	56	59									
8	清 01	栗 03	05	久 07	清 09	12	14	清 16	久 18	20	22	栗 25	清 27	29	31		
	久 33	清 35	38	40	久 42	清 44	46	48	51	清 53	栗 55	57	59				

28 Trains/hour

Background

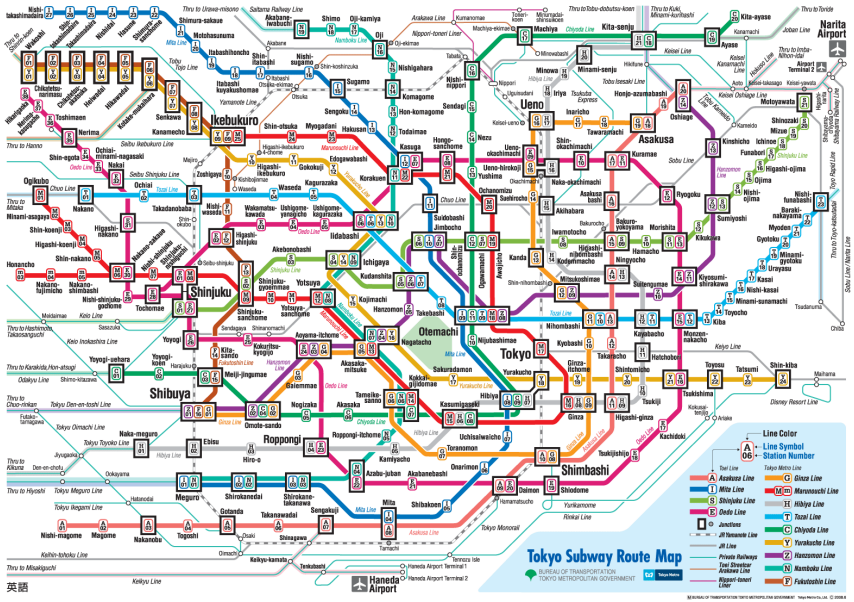
Problems in urban railways in Japan

- Small delays (several minutes) often happen
- Delays propagate to other trains
- This is because
  - Trains are running densely
  - Huge number of passengers

Background

Problems in urban railways in Japan

- Small delays (several minutes) often happen
- Delays propagate to other trains
- This is because
  - Trains are running densely
  - Huge number of passengers
  - Complex network and through trains from suburban area



Background

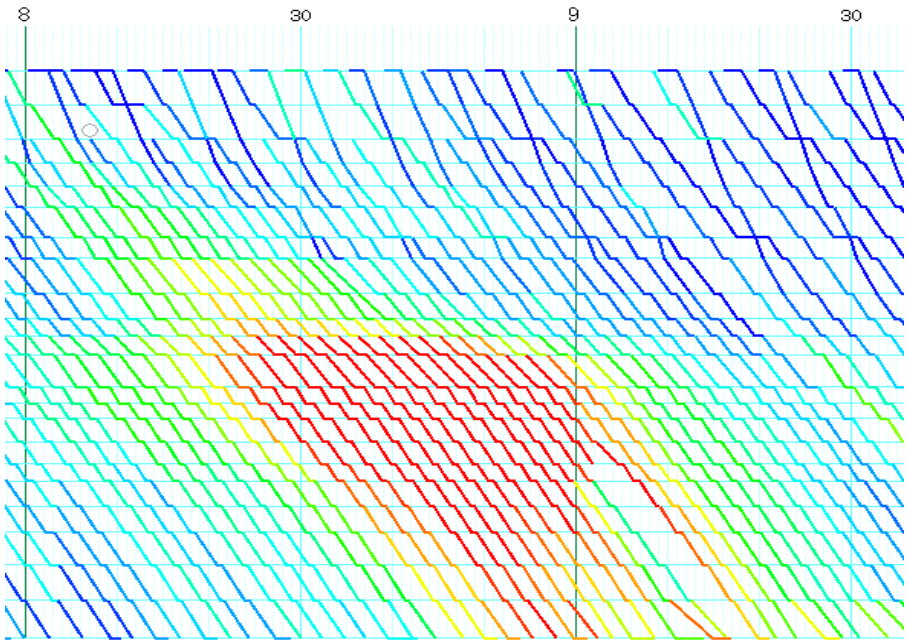
Causes of primary delays - Increase of dwell times:

1. Large number of passengers
2. Door jammed by bags, umbrellas, etc.
3. Sick passengers on train

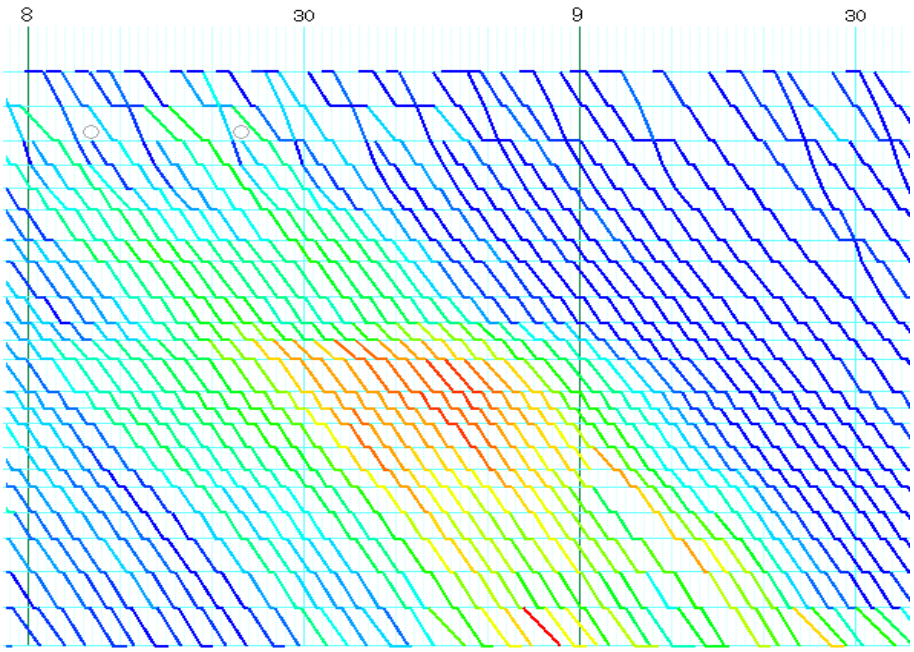
Delay reduction measures

- Revise timetables – adjust dwell times, intervals, etc.

before

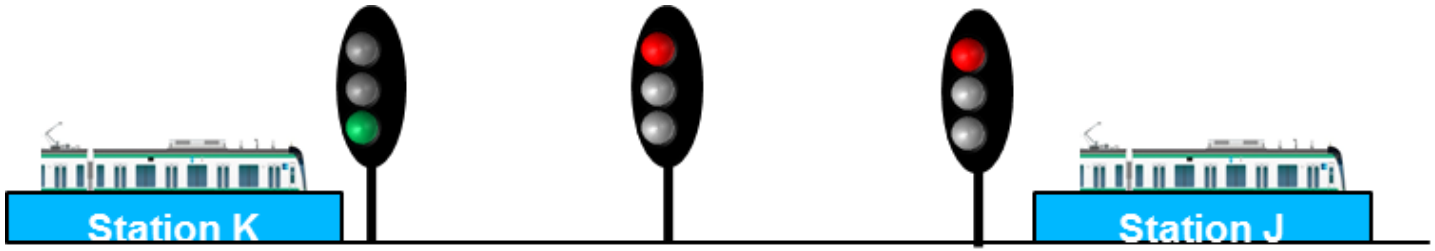


after



Delay reduction measures

- Revise timetables – adjust dwell times, intervals, etc.
- Improve signalling systems – shorten headways



Delay reduction measures

- Revise timetables – adjust dwell times, intervals, etc.
- Improve signalling systems – shorten headways
- Improve operations on platforms
  - Deploy additional staff to help passengers getting on and off
  - Change location where trains stop in a station

Background

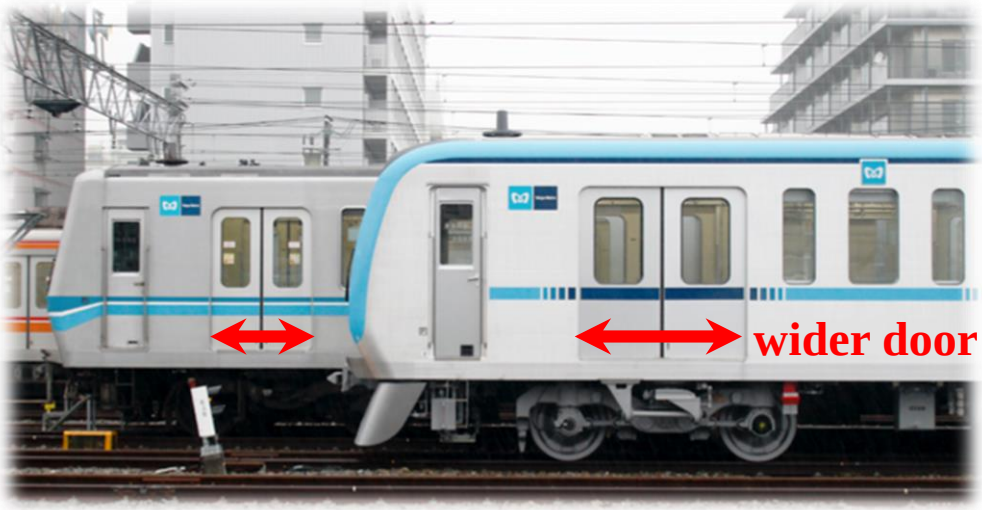
Delay reduction measures

- Revise timetables – adjust dwell times, intervals, etc.
- Improve signalling systems – shorten headways
- Improve operations on platforms
- Enlarge platforms

Background

Delay reduction measures

- Revise timetables – adjust dwell times, intervals, etc.
- Improve signalling systems – shorten headways
- Improve operations on platforms
- Enlarge platforms
- Introduce new type of rolling stock
  - More doors
  - Wider doors



Contents

The problem is:

Even if delays were reduced, it is difficult to evaluate:
- If these countermeasures were effective or not
- which countermeasures were effective?

Because

- Operations differ from day to day
- Multiple delay reduction measures are adopted together
- Timetables are highly complex

We need a quantitative approach!

Research Objective

Even if delays were reduced, it is difficult to evaluate:
- If these countermeasures were effective or not
- which countermeasures were effective?

We want to establish an approach to evaluate the effectiveness of the delay reduction countermeasures through data mining technique!

Contents

Research Concept

- Use of a data mining technique
  - Association rules
- Analyze daily train traffic records
- Identify strong relationships (rules)

Association rules

¶ Example

Beer

Diapers

Rule:
A lot of people buy Beer and Diapers together!

Association rules

Example

One day, a record of a purchase of person is 100.

Beer
40

Diapers
20

30

10

Purchase together

10

10

Rule	$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$	$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$	$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}$
Beer ⇒ Diapers  	$\frac{10}{100} = 0.1$	$\frac{10}{40} = 0.25$	$\frac{0.25}{0.1} = 2.5$

Association rules

Example

One day, a record of a purchase of person is 100.

Beer
40

Diapers
20

30 | 10 | 10 | 10

Purchase together

Rule	$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$	$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$	$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}$
Beer ⇒ Diapers  	$\frac{10}{100} = 0.1$	$\frac{10}{40} = 0.25$	$\frac{0.25}{0.1} = 2.5$

Association rules

Example

One day, a record of a purchase of person is 100.

Beer
40

Diapers
20

30

10

Purchase together

10

10

Rule	$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$	$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$	$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}$
Beer ⇒ Diapers  	$\frac{10}{100} = 0.1$	$\frac{10}{40} = 0.25$	$\frac{0.25}{0.1} = 2.5$

Association rules

Example

One day, a record of a purchase of person is 100.

Beer
40

Diapers
20

30

10

Purchase together

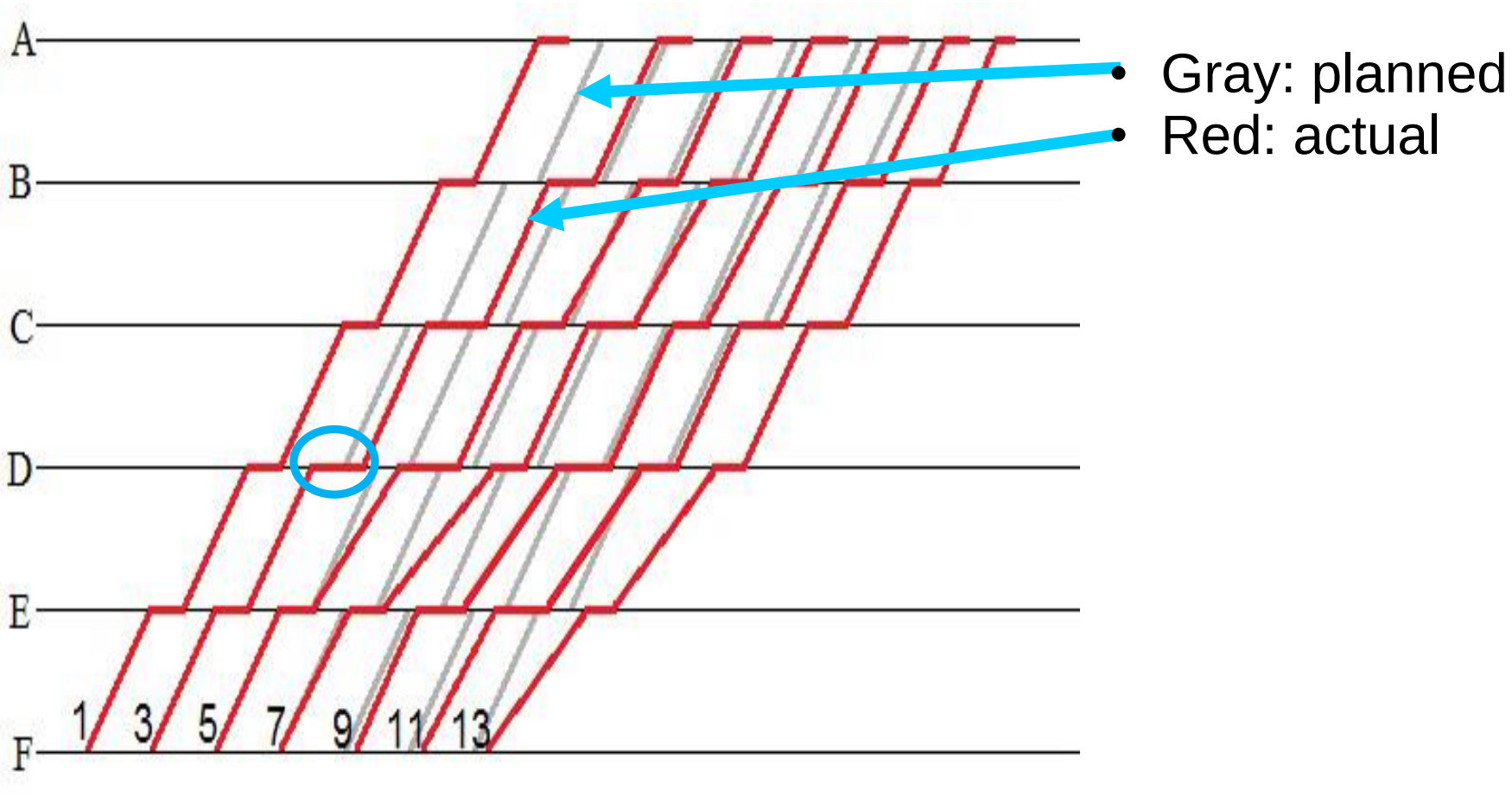
10

10

Rule	$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$	$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$	$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}$
Beer ⇒ Diapers  	$\frac{10}{100} = 0.1$	$\frac{10}{40} = 0.25$	$\frac{0.25}{0.1} = 2.5$

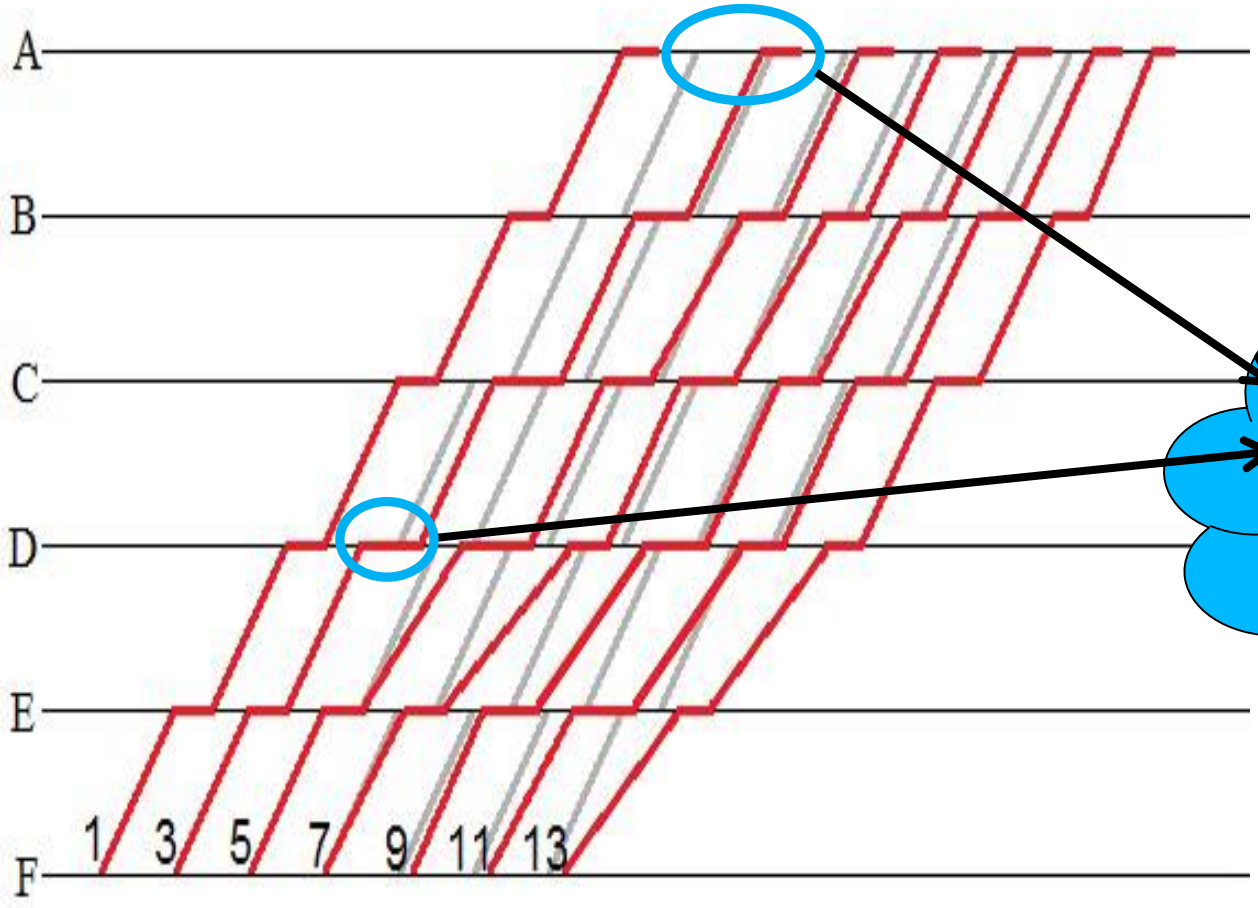
Basic idea:

- How we can apply association rules to detect the cause of delays?



Basic idea:

- How we can apply association rules to detect the cause of delays?

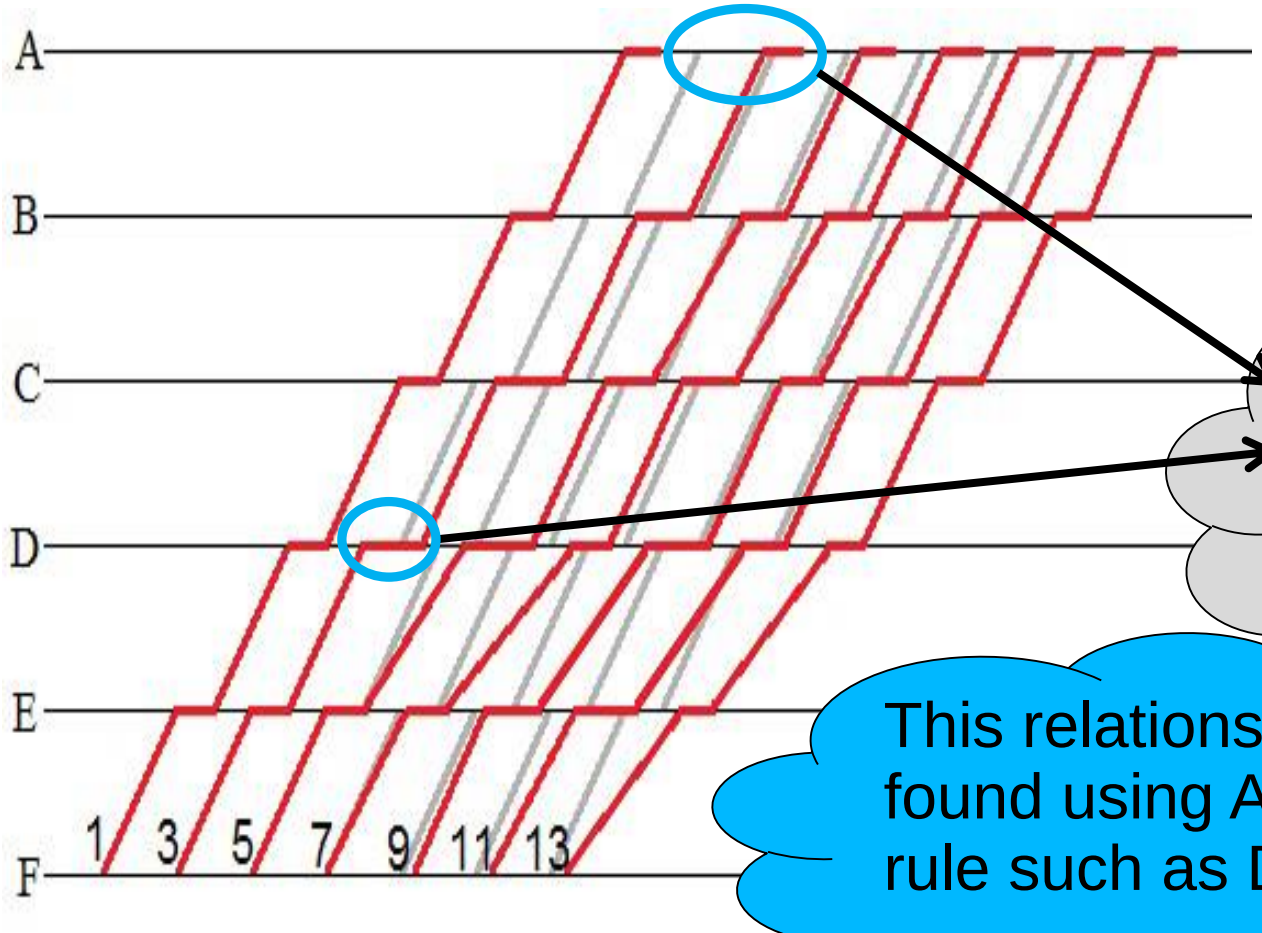


- Gray: planned
- Red: actual

If these two delays occur always at the same time, there exist a relationship!

Basic idea:

- How we can apply association rules to detect the cause of delays?



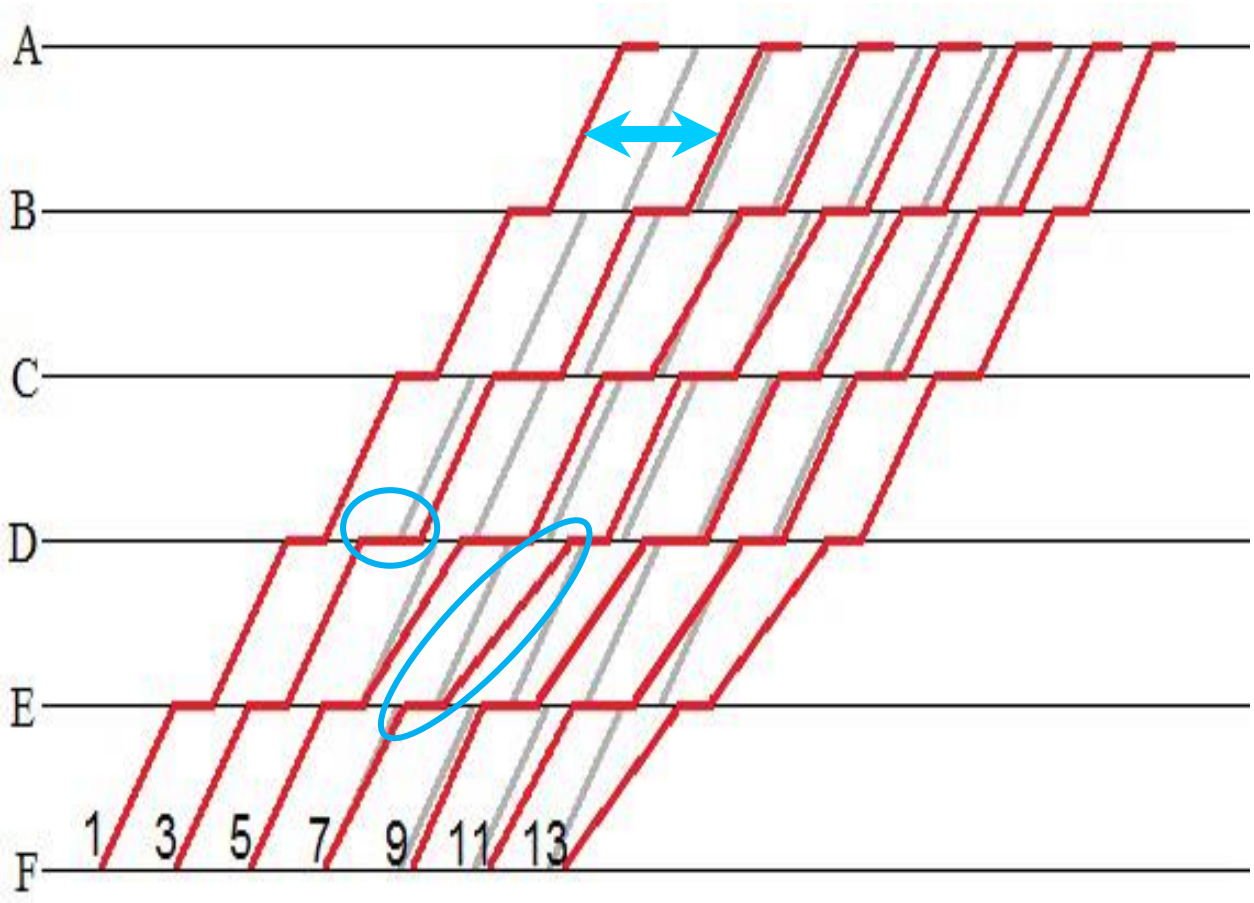
- Gray: planned
- Red: actual

If these two delays occur always at the same time, there exist a relationship!

This relationship can be found using Association rule such as D -> A.

Basic idea:

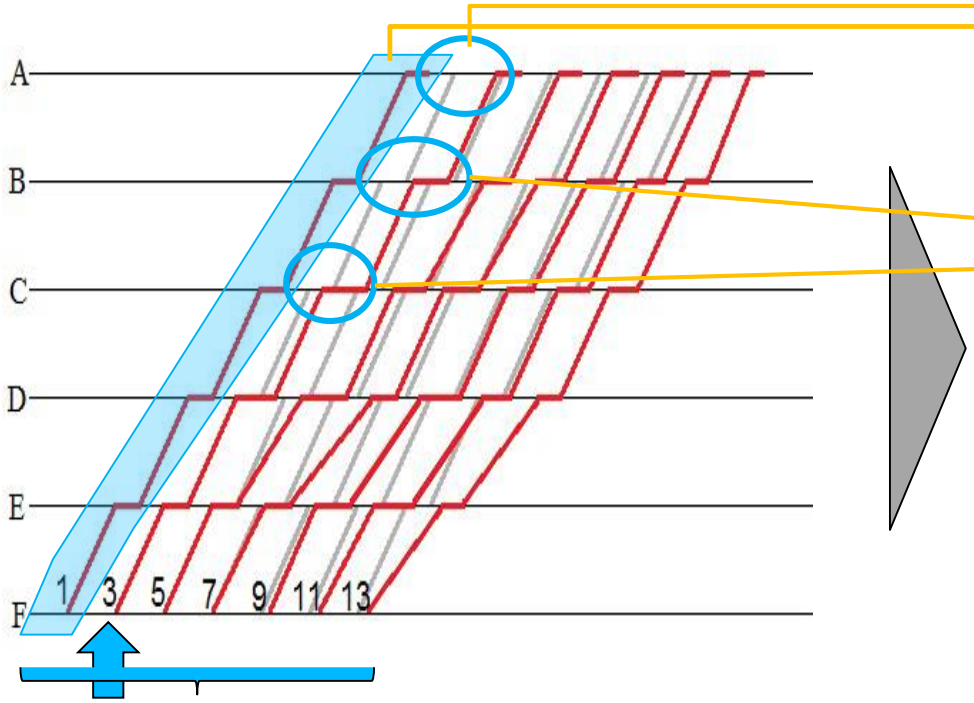
How we can apply association rules to detect the cause of delays?



- delay
- increase of dwell times
- increase of running times
- increase of intervals between trains

Association rules

How to make bit table



Station / Train	A	B	C	D	E	F
1	0	0	0	0	0	0
3	1	1	1	0	0	0
5	1	1	1	1	0	0
7	1	1	1	1	0	0
9	1	1	1	1	1	1
11	1	1	1	1	1	1
13	1	1	1	1	1	1

Transactions are 7 in this table.

Rule「A⇒D」

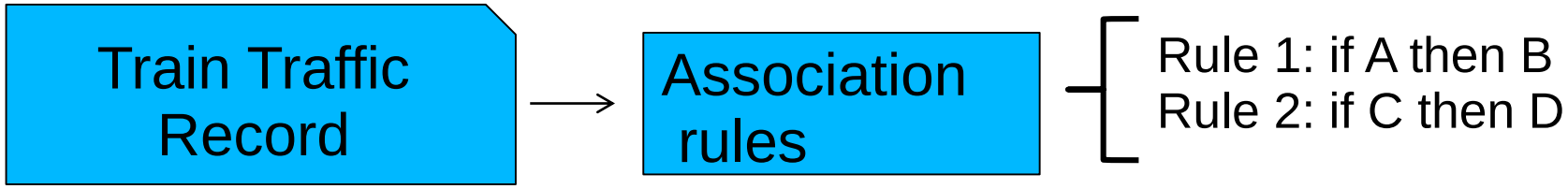
Confidence 5/6 ≒ 0.83

Support 5/7 ≒ 0.71

How we can evaluate delay reduction measures?

- We want to confirm if our approach works well or not

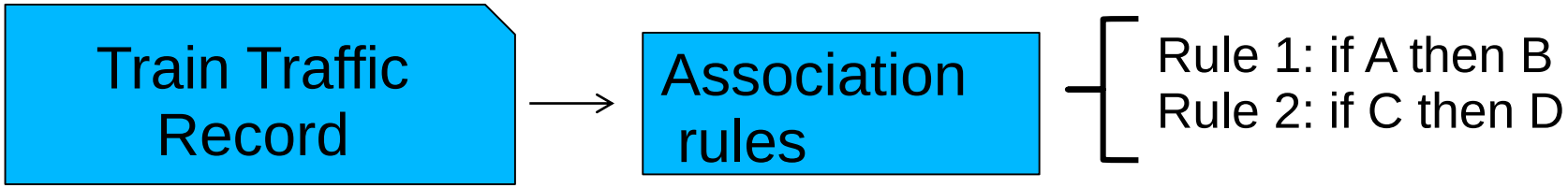
Before



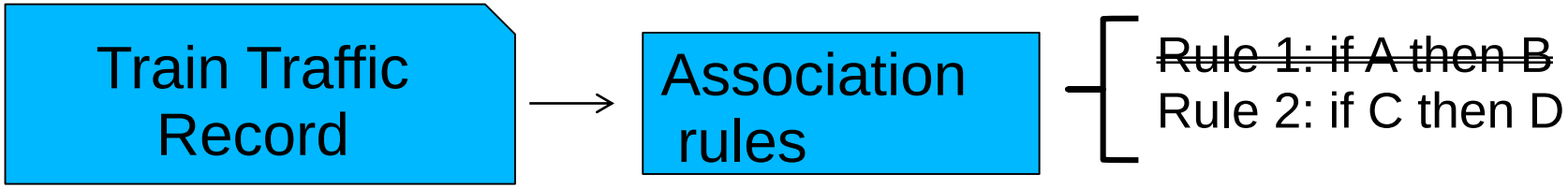
How we can evaluate delay reduction measures?

- We want to confirm if our approach works well or not

Before

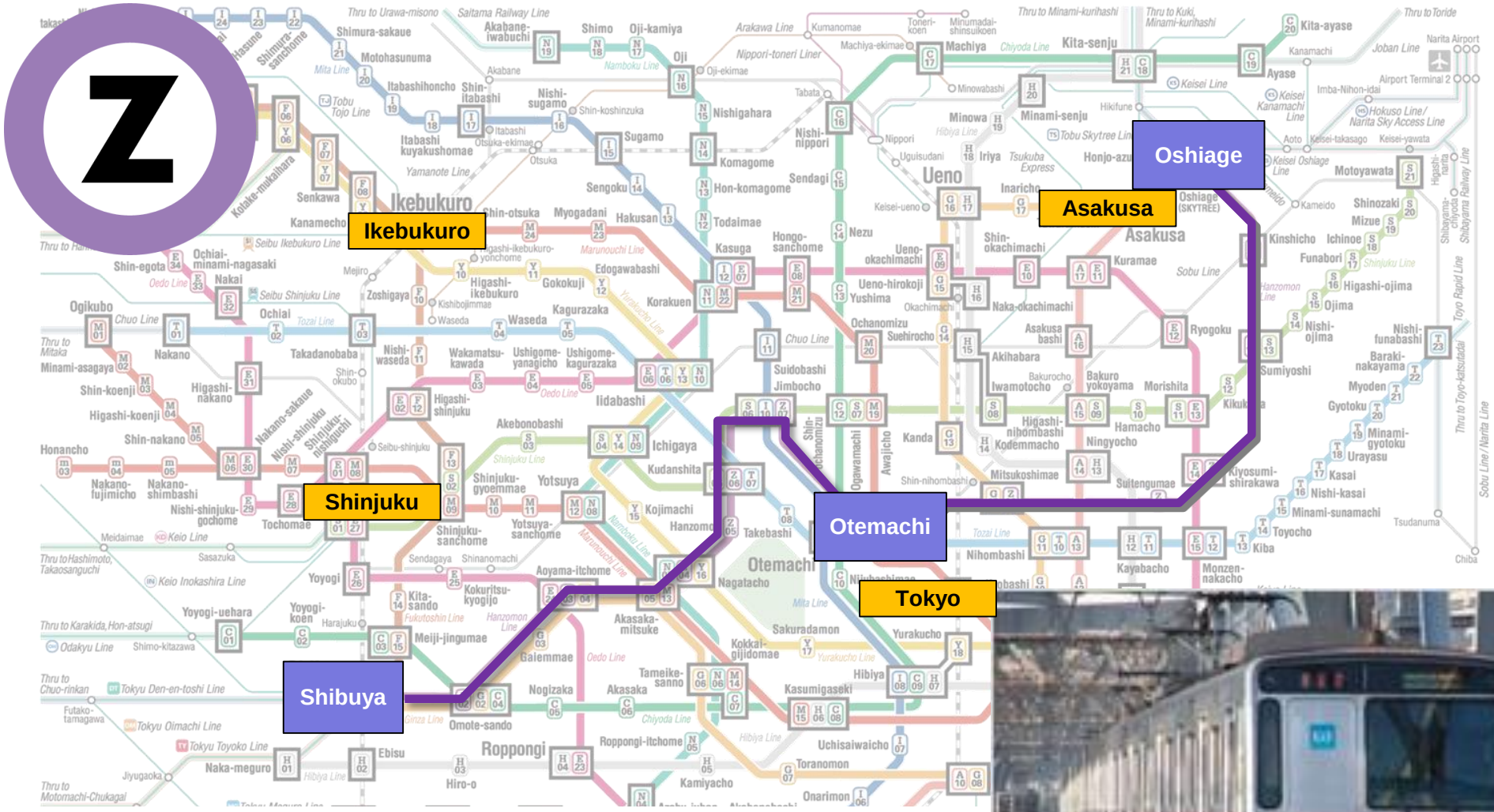


After delay reduction measure to Station A is introduced



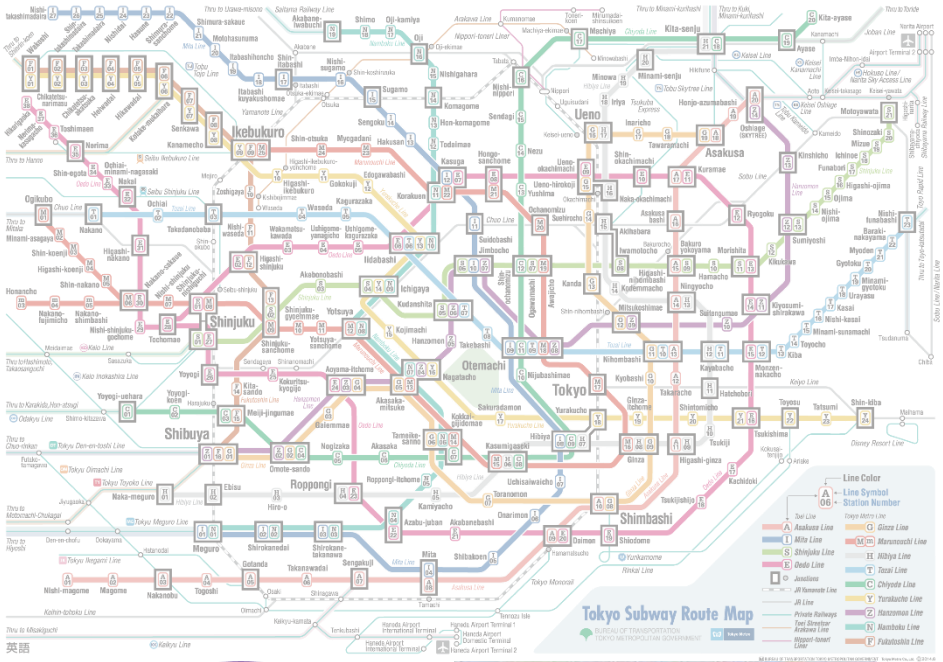
Contents

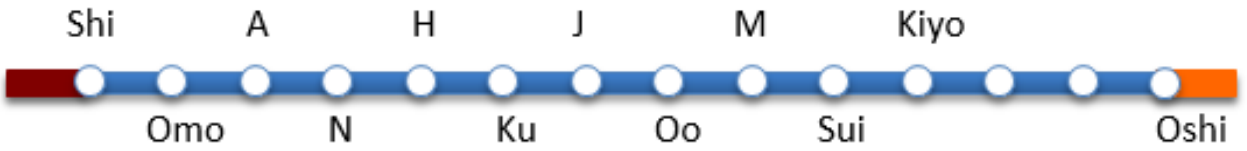
Hanzomon Line



Hanzomon Line

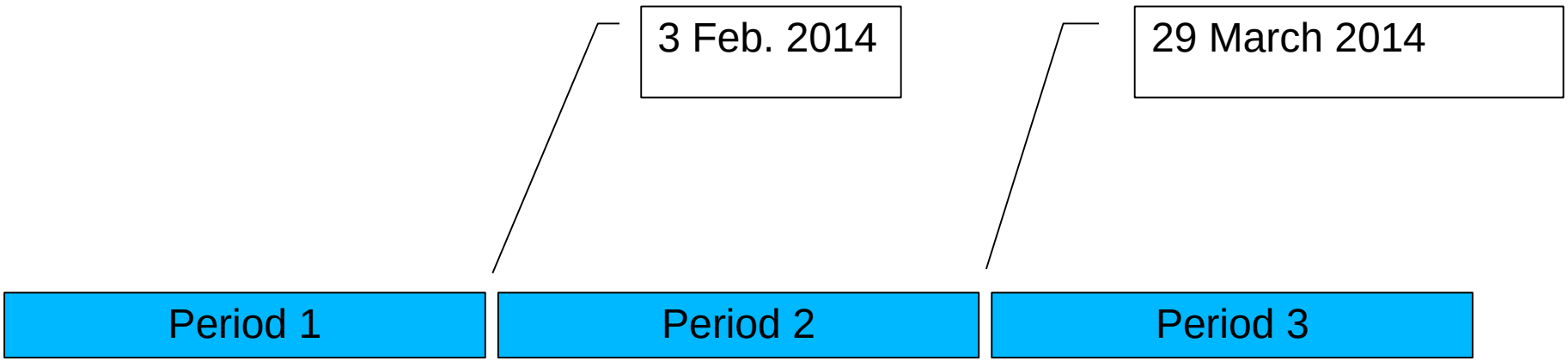
- located in the center of Tokyo
- Length: 16.8 km
- Number of stations: 14
- Rush hour frequency:
  28 trains/hour
- Through train services with:
  - Tokyu Denentoshi line
  - Tobu Skytree line





Numerical Experiments

- We applied our algorithm to the train traffic record data of the three periods, period 1, period 2 and period 3.



Experimental results

Period A

	Station	Confidence(Max)	Number of rules
1	Omo − A	1.0000000	16
2	Sui − Kiyo	1.0000000	13
3	H − Ku	1.0000000	4
4	N DT	0.8857143	6
5	A DT	0.8823529	7
6	Omo DT	0.8421053	2

Period 1

Rules concerning Station N and Ku were found!

Experimental results

Period A

	Station	Confidence(Max)	Number of rules
1	Omo－A	1.0000000	16
2	Sui－Kiyo	1.0000000	13
3	H－Ku	1.0000000	4
4	N DT	0.8857143	6
5	A DT	0.8823529	7
6	Omo DT	0.8421053	2

Period B

	Station	Confidence(Max)	Number of rules
1	J－Oo	1.0000000	5
2	Shi－Omo	1.0000000	12
3	H－Ku	1.0000000	11
4	Oshi	1.0000000	14
5	Sui－Kiyo	0.8750000	4
6	M DT	0.8750000	1
7	Omo DT	0.8461538	8

Period 1

Rules concerning Station N and Ku were found!

Period 2

Rule concerning Station Ku was found!
Rues concerning Station N was not found!

Experimental results

Period A

	Station	Confidence(Max)	Number of rules
1	Omo－A	1.0000000	16
2	Sui－Kiyo	1.0000000	13
3	H－Ku	1.0000000	4
4	N DT	0.8857143	6
5	A DT	0.8823529	7
6	Omo DT	0.8421053	2

Period B

	Station	Confidence(Max)	Number of rules
1	J－Oo	1.0000000	5
2	Shi－Omo	1.0000000	12
3	H－Ku	1.0000000	11
4	Oshi	1.0000000	14
5	Sui－Kiyo	0.8750000	4
6	M DT	0.8750000	1
7	Omo DT	0.8461538	8

Period C

	Station	Confidence(Max)	Number of rules
1	Omo－A	1.0000000	6
2	Sumi	1.0000000	9
3	Omo	0.8888889	4

Period 1

Rules concerning Station N and Ku were found!

Period 2

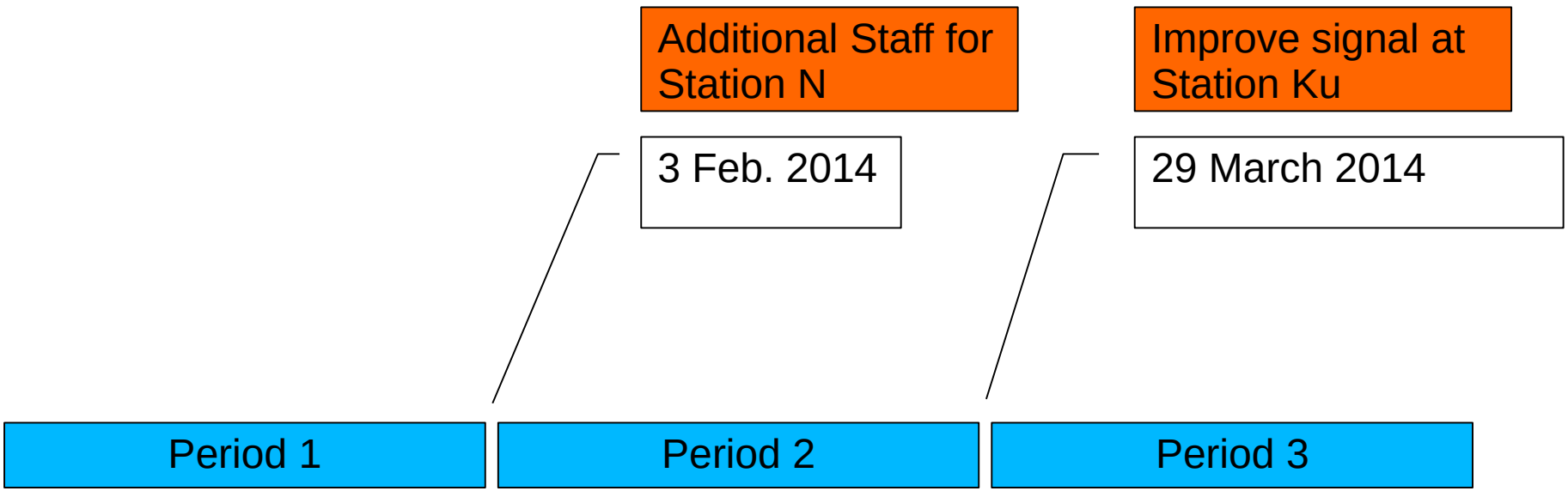
Rule concerning Station Ku was found!
Rule concerning Station N was not found!

Period 3

No rules about Station Ku and N were found!

Numerical Experiments

- We think these results are reasonable



Future work

- We have confirmed our algorithm is promising by numerical experiments
- We would like to
  - continue experiments for other data
  - further elaborate the algorithm
- In the next step, we would like to apply our algorithm to find out effective delay reduction measures

Conclusions

- We developed an algorithm based on association rules to evaluate the effectiveness of each delay reduction countermeasure.
- The algorithm takes daily traffic records as an input and discover strong rules.
- By comparing results before and after the adoption of the countermeasures, the developed method can successfully identified the strong relationships between countermeasures and delay.

Thank you very much for your attention