Poster ID: 18

Shota Suzuki, 4th Undergraduate Student
ARC-iTransportation Cluster, The University of Aizu

Poster Session at Graduate School Information Fair

Discovering Traffic Congestion Patterns in Big Transportation Networks

Importance

1. Japan recorded 2.61 thousand fatalities in 2022
2. 381 billion hours are lost annually in Japan.
3. 12 trillion yen lost annually.
4. Improving traffic safety is crucial to achieve SDGs.
5. Thus, it is important to discover traffic congestion patterns in large-scale transportation networks.

Table 1: Traffic incidents in Fukushima 2022 and 2023

	Total Incidents	Total Deaths (Including Elderly)			Total injuries	Total Property incidents
This Year	1,678 ↑	30 ↑	33 ↑	(17)	1,949 ↑	27,574 ↑
Last Year	1,559	25	25	(17)	1,790	27,393
Change of Quantity	119	5	8	(0)	159	181
Change of rate	7.6%	20.0	32.0%	(0.0%)	8.9%	0.7%

Challenge

1. How to model traffic congestion data?
2. What should be the mathematical model to define traffic congestion patterns?
3. What is the algorithm to find all traffic congestion patterns?

Our model and Algorithm

O ={O_1,O_2,…,O_n} be the set of objects

LD= $\cup_{o_j \in o}(oj, coordinatesoj)$

An Itemset, $G_i \subseteq$ O

A sequence, $S_j = \cup_{i=1}^{p} G_i$

A sequence database, SD = $\cup_{sid=1}^{k}(sid, Sj)$

Neighborhood itemset, $NG_i \subseteq$ O
$\forall O_i, Oj \in NGi, dist(Oi, Oj) \leq maxDist$

Neighborhood sequence, $NS_j = \cup_{i=1}^{p} NG_i$

If frequency of NS_j is greater than minimum support (minSup), we call it traffic congestion pattern

A depth-first search algorithm, GFSP, has been proposed

Table 2: Location database of road segments

Object	Coordinates
a	(0,0)
b	(0,1)
c	(3,0)
d	(1,1)
e	(3,2)
f	(3,3)
g	(4,4)

Table 3: Traffic congestion on road segments

Day	hour	*a*	*b*	*c*	*d*	*e*	*f*	*g*
1	1	1	1	0	0	0	0	0
	2	0	0	1	0	0	0	0
	3	0	0	0	1	1	1	0
2	1	1	0	0	1	0	0	0
	2	0	0	1	1	0	0	0
	3	1	0	0	0	0	1	0
3	1	0	1	1	0	0	0	0
	2	0	0	0	1	0	0	0
	3	1	1	0	1	1	0	0
	4	1	0	1	0	0	0	0
4	1	1	1	0	0	0	0	0
	2	0	0	1	0	0	0	0
	3	1	0	0	0	1	0	1
	4	0	0	0	1	0	1	0

convert →

Table 4: sequence database

sid	Sequences
1	$\langle\{a,b\},\{c\},\{d,e,f\}\rangle$
2	$\langle\{a,d\},\{c,d\},\{a,f\}\rangle$
3	$\langle\{b,c\},\{d\},\{a,b,d,e\},\{a,c\}\rangle$
4	$\langle\{a,b\},\{c\},\{a,e,g\},\{d,f\}\rangle$

↓

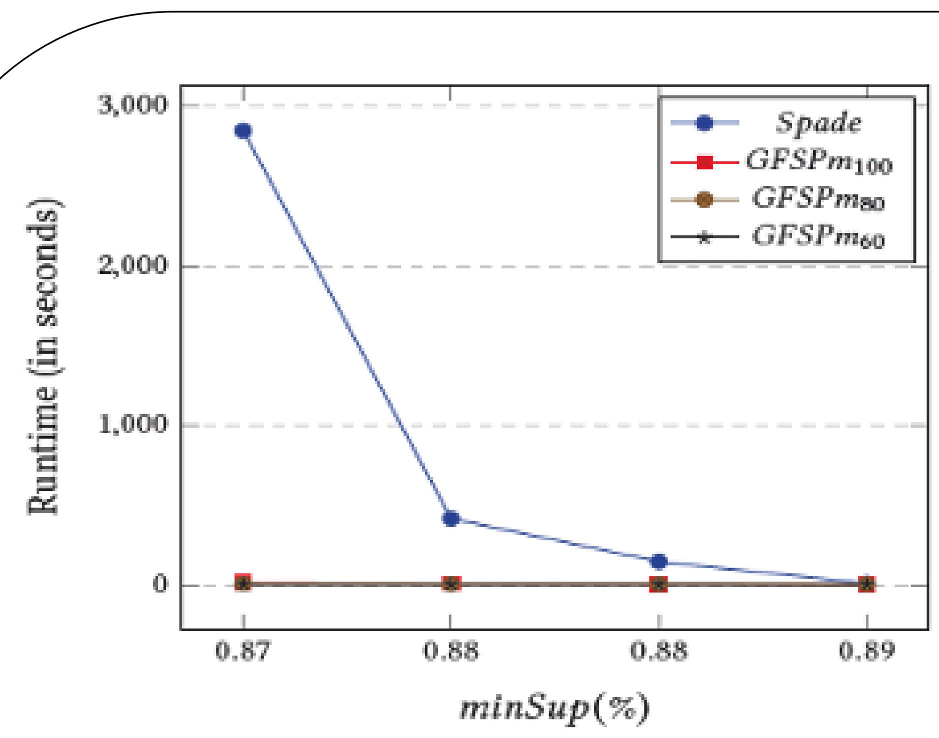
Mining algorithm

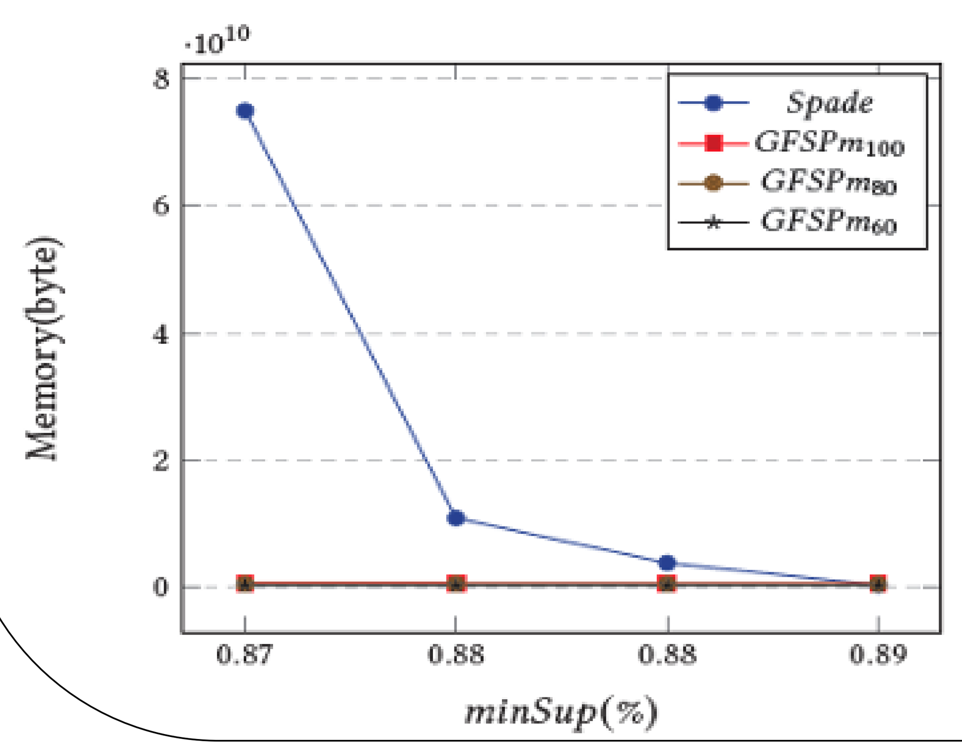
↓

The pattern ***<{ab} {d}>*** *appeared in two out of four sequences (50% of time)*

Meaning of above pattern:
1. Congestion was first observed on the roads a and b.
2. Later, congestion was observed on road d.

Experimental Results





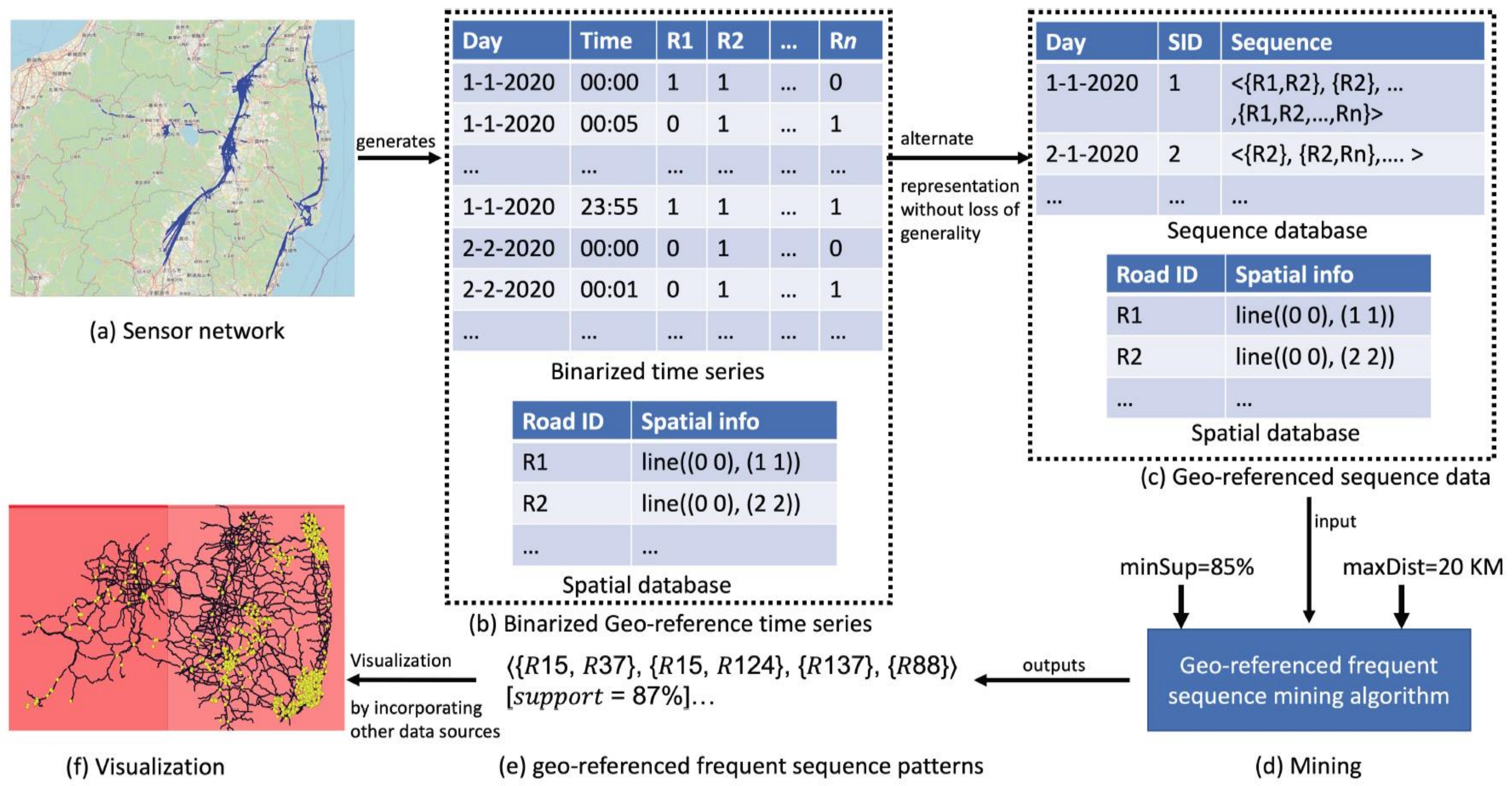


Figure 3: Case-study on Fukushima Traffic Congestion Data

Publication:

Suzuki Shota and Rage Uday Kiran: Towards Efficient Discovery of Spatially Interesting Patterns in Geo-referenced Sequential Databases. To be appeared in SSDBM 2023 (CORE RANK: A)