INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

R&D SHOWCASE 2021
Technology, Social Impact

Discovering Partial Periodic Spatial Patterns in Spatiotemporal Databases

Introduction

- Partial periodic pattern mining is an important data mining model with many real-world applications.
- A partial periodic pattern represents a set of items that repeats itself at regular intervals in the data. It is thus useful in characterizing the cyclic behavior
- The current partial periodic pattern models are not ideal for some applications because they disregard the crucial information about the spatial (or geometric) characteristics of an item.
- With this motivation, this paper proposes a more flexible model of partial periodic spatial pattern that exists in a spatiotemporal database.
- Three constraints, maximum inter-arrival time (maxIAT), minimum period-support (minPS) and maximum distance (maxDist), have been employed to determine the interestingness of a pattern in a spatiotemporal database.

Model of partial periodic spatial patterns

- Let $I = \{i_1, i_2, \cdots, i_n\}$, $n \geq 1$, be a set of spatial items. Let P_{ij} denote a set of coordinates for an item $i_j \in I$. The spatial database SD is a collection of items and their coordinates. That is, $SD = \{(i_1, P_{i1}),(i_2, P_{i2}), \cdots,(i_n, P_{in})\}$. Two items, $i_p, i_q \in I$, are said to be neighbors to each other if $Dist(i_p, i_q) \leq maxDist$
- A transaction $t_{tid} = (tid, ts, Y)$, where $tid \geq 1$ represents the transaction identifier, $ts \in R+$ represents the timestamp and $Y \subseteq I$ is a pattern. An (irregular) temporal database TDB is a collection of transactions. That is, $TDB = \{t_1, t_2, \cdots, t_m\}$, $1 \leq m \leq |TDB|$.
- Let $TS^X = \{ts^X_{tida}, ts^X_{tidb}, \cdots, ts^X_{tidc}\}$, $tid_a, tid_b, tid_c \in (1, |TDB|)$, denote the set of all timestamps in which the pattern X has appeared in the database
- Let ts^X_a and ts^X_b denote two consecutive timestamps at which X has appeared in TDB. Time difference between ts^X_a and ts^X_b is defined as an inter-arrival time of X, and denoted as iat^X_p
- **Period-Support** : Let $\hat{A} \subseteq IAT^s$ denote the set of all interesting inter-arrival times of s in S. That is, if there exists $iat^s_k \in IAT^s$: $iat^s_k \leq maxIAT$, then $iat^s_k \in \hat{A}$ .The period-support of s is the number of periodic occurrences of s.That is period-support(s) = $|\hat{A}|$.
- **Partial periodic spatial pattern:** A spatial pattern X is said to be a partial periodic spatial pattern if its period-support is no less than the user-specified minimum period-support (minPS). That is, X is a partial periodic spatial pattern if $PS(X) \geq minPS$ and $\max(Dist(i_p, i_q \mid \forall i_p, i_q \in X)) \leq maxDist$.

Proposed ST-ECLAT algorithm

Algorithm 1 ST-ECLAT (Set of items (I), spatial database (SD), temporal database (TDB), maximum distance ($maxDist$), maximum inter-arrival time ($maxIAT$), minimum period-support ($minPS$)

```
Scan the spatial database and find neighbors for each item using a distance function. Let N(i_j) denote the set of neighboring items for an item i_j.
Scan the temporal database and generate ts-list for each item. Simultaneously, calculate period-support for each item i_j by traversing its ts-list.
Find partial periodic spatial items by pruning uninteresting items that have period-support less minPS. Sort the partial periodic spatial items in descending order of their support values. Let S denote this sorted list of partial periodic spatial items. Let TS(S) denote the set of ts-lists for all partial periodic spatial items in S.
Initialize α = ∅ and call depthFirstSearch(α, TS(S), S).
```

TABLE I: Spatial database

Item	Coordinates
a	(0,0)
b	(3,3)
c	(0,7)
d	(5,3)
e	(5,6)
f	(7,7)
g	(7,1)

TABLE II: Neighbors

Item	Neighbors
a	bd
b	acdefg
c	bde
d	abefg
e	bcdfg
f	bdeg
g	bdef

TABLE III: Running example: temporal database

tid	ts	items	tid	ts	items	tid	ts	items
1	1	abcefg	5	6	acd	9	12	abef
2	2	abcf	6	7	bcdfg	10	13	abdef
3	3	ab	7	8	acde	11	14	eg
4	5	cdeg	8	11	acefg	12	15	acdfg

Algorithm 2 depthFirstSearch(α, $TS(N(\alpha))$, $N(\alpha)$)

```
while TS(N(α))! = ∅ do
    i_j, ts-list(i_j) = TS(N(α)).pop();{extract item and its ts-list from TS(N(α)) using pop function}
    output α ∪ i_j as a partial periodic spatial pattern;
    set N(α ∪ i_j) = N(α) ∩ N(i_j);
    set suffixItems = ∅.
    for i_k,ts-list(i_k) in TS(N(α)) do
        if i_k ∈ N(α ∪ i_j) then
            generate ts-list(α ∪ i_k) = ts-list(α) ∪ ts-list(i_j);
            PS(α∪i_k)=calculatePeriodSupport(ts-list(α∪i_k)).
            if PS(α ∪ i_k) ≥ minPS then
                suffixItems.append(i_k, ts-list(α ∪ i_k));
    depthFirstSearch(α ∪ i_j, suffixItems, N(α ∪ i_j));
```

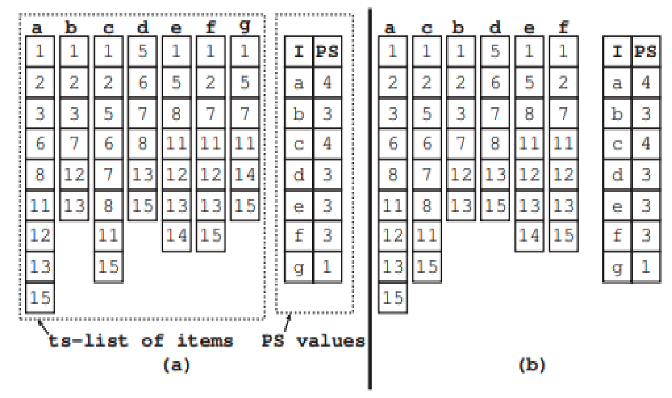


Fig. 1: Finding partial periodic spatial items. (a) After scanning the temporal database and (b) Sorted list of partial periodic spatial items

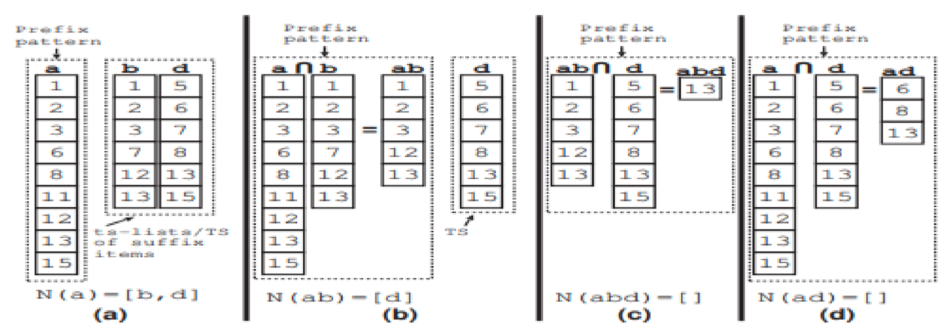
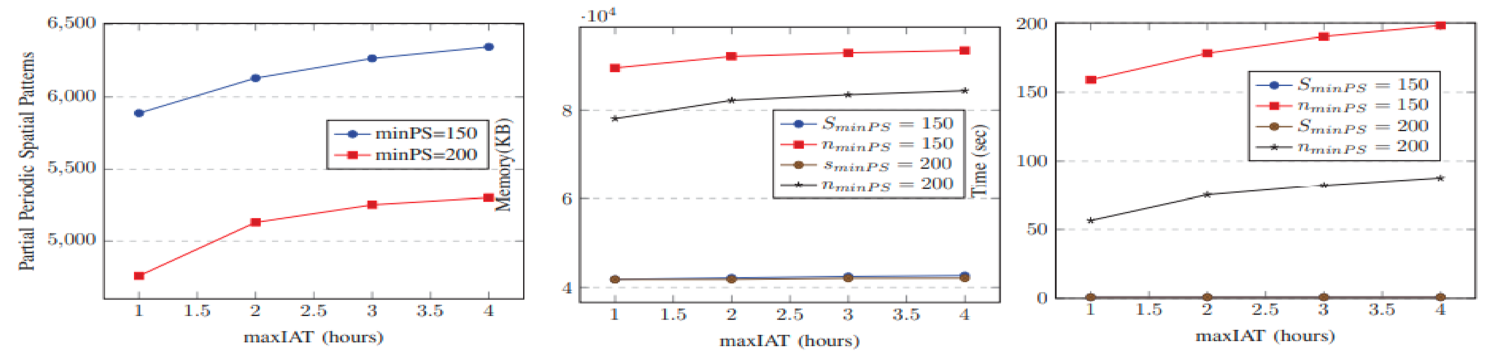


Fig. 2: Performing depth-first search on the itemset lattice using item a. (a) prefix pattern a, (b) prefix pattern ab, (c) prefix pattern abd and (d) prefix pattern ad

Performance Evaluation



(a) Number of patterns (b) Memory (c) Runtime

Fig. 3: Evaluation of n-ECLAT and ST-ECLAT algorithms

TABLE V: Few interesting patterns generated in Pollution database

S.No.	Patterns (or station ids)	color
1	{868, 872, 874, 876, 881}	Green
2	{4372, 4256, 4312, 4263, 4335,-4230, 4236, 4329, 4377}	Red
3	{1193, 1225, 1261, 1266}	Blue
4	{3946, 3954, 3503, 4032, 3097, 3939}	Pink

Fig. 4: Areas where people have been regularly subjected to unhealthy levels of PM2.5 concentrate

Publication

R. Uday Kiran, C. Saideep, K. Zettsu, M. Toyoda, M. Kitsuregawa and P. Krishna Reddy, "Discovering Partial Periodic Spatial Patterns in Spatiotemporal Databases," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 233-238.

Authors Name : R. Uday Kiran,C. SaiDeep, P. Krishna Reddy . Research Center: DSAC