[1]Swati Meshram

[2] Kishor P. Wagh

A Novel and Efficient Spatio-Temporal Co-location Pattern Mining Algorithm

Abstract: - Colocation pattern mining approaches aim at discovering neighboring relationships of distinct spatial features in geographic and temporal space.With big spatio-temporal dataset, there is large number of patterns often discovered. Then it is of importance to discover meaningful and patterns of interest which come as an aid in applications in use for humans and commercial use. To discover interesting patterns, we present in this article a co-location pattern mining algorithmic framework by considering neighbourhood, clustering, hashing, and mining methods. Neighbourhood relationship describes the closeness between the entities. The results of neighbourhood could be varied by varying the distance threshold. These objects exhibiting neighbouring entities are grouped using clustering technique. Clustering is a classical research approach that produces grouping of the entities. A clustering technique has been presented in the paper for spatio-temporal dimension that offers the advantage of faster grouping based on the neighbourhood relationship. Finally, a hash structure is utilized for faster access and retrieval of patterns. The proposed mining algorithm along with the distance and time threshold efficiently discovers the interesting spatio-temporal patterns and validates the patterns. The experiment conducted shows that the proposed co-location algorithm method yields effective and efficient outcome.

Keywords:*Co-location, pattern mining, spatio-temporal, neighbourhood, clustering.*

I. INTRODUCTION

In recent years, the field of spatiotemporal data mining has gained significant attention and has been recognized as an efficient and effective technique for discovering interesting patterns and capturing complex association. Researchers have devoted considerable effort to address key issues in spatiotemporal data mining, including the pre-treatment of spatiotemporal datasets and the development of effective methods for mining spatiotemporal information.

Spatiotemporal co-location pattern mining refers to the process of analyzing and discovering patterns of co-occurrence between entities of distinct types, in both space and time.Such patterns can provide valuable insights into the relationship and behavior of entities over time. Co-location pattern mining discovery has applications in various domains such as urban facility planning [1], criminology [2], to discover the diseases in neighbourhood [3]. Co-location patterns were studied to examine relationship between air pollutants emission in the neighbourhood and child cancer cases[4], hotspot detection in [5], road networks[6] and many more.

With increase in use of location-based services and global positioning system has led to voluminous spatio-temporal data generation. The data recorded could be point data representing objects as single geographical point, line data being collection of points or polygonal data representing collection of objects over an area.

In the past, the methods have been developed to mine spatial colocation patterns and very few works have been carried out on spatio-temporal co-location pattern mining. These methods are based on prevalent threshold to determine prevalent co-location patterns where the prevalent threshold is a user defined parameter. Spatial data space is continuous where the spatial instances cannot be represented as transactions ideally. Colocation pattern discovery is challenging for various reasons as: First, due to the absence of the transactions, association rules would

[1]* Research Scholar, Department of Computer Science and Engineering, Government College of Engineering, Amravati, Maharashtra, India, and

Assistant Professor, Department of Computer Science, SNDT Women's University, Mumbai, Maharashtra, India.

email: swati.meshram@computersc.sndt.ac.in, ORCID ID: 0000-0002-7569-7995

[2]Assistant Professor, Department of Information Technology, Government College of Engineering, Amravati, Maharashtra,

email: kpwagh2000@gmail.com, ORCID ID: 0000-0002-8189-3378

not be applicable directly. Second, complexity of the algorithms increases with increase in features and size of data. Third,determining the interestingness measure suitable for all types of patterns is very challenging.

In this paper we propose a novel technique for spatio-temporal colocation pattern mining.We have outlined core ideas of our framework and have also included experimented results to prove the applicability of the algorithm.

II. LITERATURE REVIEW

A. Types of Co-location patterns

The types of patterns are broadly categorized as shown in figure 1 are as follows.

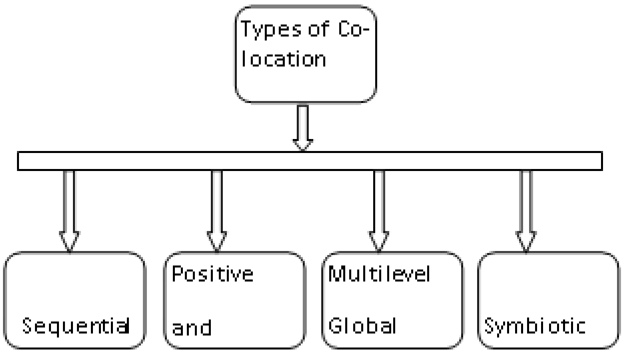

Figure. 1 Types of Co-location patterns.

1) Sequential co-location patterns

Is the type of patterns that exhibit a sequence of causal relationship and has been adopted to perform analysis of dengue disease spread. The case study relates a trend to when it rains and water is stagnant in the surrounding along with increase in temperature collectively leads to proliferation of mosquitoes and rise in dengue cases. This type of interconnection is a sequential co-location pattern, which are unveiled through the application of association rules on frequent item set to generate patterns [7].

2) Positive and negative co-location patterns

These types of patterns indicate a robust positive interaction and association among the features present in the dataset. Conversely, negative co-location patterns signify a prominent negative correlation and exhibit a mutually exclusive relationship or a contrast set [8]. In the context of [9], the algorithm is a join-based prevalent co-location pattern mining, which operates incrementally to identify negative co-location patterns. Initially, smaller-sized negative co-location patterns are identified and subsequently merged with other patterns to create larger-sized co-location patterns [10].

3) Multilevel – global and local patterns

As stated by Liu [11], the existence of co-location patterns at various levels is attributed to the spatial heterogeneity arising from unevenly distributed features. Global co-location patterns cover the entire study area, while local co-location patterns are confined to specific regions within the study area [12],[13]. The identification of such multilevel co-location patterns offers valuable insights into the interactions among diverse spatial phenomena. The determination of multi-level co-location patterns is facilitated by the participation index (PI); if the participation index of a candidate pattern surpasses a predefined threshold, it is classified as a global pattern. Conversely, if the participation index falls below the threshold, the candidate pattern is recognized as a local pattern. The natural neighborhood relationship, coupled with the participation index, serves as the basis for distinguishing between global and local patterns.

4) Symbiotic Co-location patterns

As indicated in the study [14], symbiotic mutualism patterns involve a reciprocal relationship where each feature within the pattern benefits from the association. The existence of certain features is conditional on the presence of others. For example, restaurants and ice-cream parlors tend to be in proximity, particularly in cities, towns, and communities. Another prevalent co-location pattern is {T. matsutake, pine trees}. Symbiotic patterns find significant application in ecological science and biology. The exploration of such patterns involves the utilization of multiple altered star neighborhoods, modified participation ratios, and symbiotic indices on the dataset.

Based on the given types of co-location pattern, research carried out is discussed further. Colocation pattern mining concept was introduced by Shekhar and Huang[15][16].The algorithm involved computation based on expensive spatial join operation. The work around has been developed on the basis of following structure. Based on the prevalence threshold which is a measure of interestingness of the pattern, the neighbourhood objects are searched for a relationship and is consistently observed over the entire study area to identify the support from various instances of this pattern. Compute its participation index, which is a measure of number of instances supporting this spatio-temporal dependency. Further, filter the patterns which have lower support to form strong co-location patterns.

To discover the patterns faster a partial join algorithm was introduced in [17]. These approaches were still considered costly when applied on large datasets and relied heavily on user defined parameters like prevalence threshold. Research based on association rules on spatial transactions were proposed in [18]. Others have adopted Apriori and FP-Growth techniques for mining association rules. These methods work on the property of downward closure for pruning of insignificant patterns. Big size co-location patterns discovery becomes inefficient in such techniques. Instance look-up, Star neighborhood-based relationship were identified and stored in a condensed way in a tree-based structure which provides join-less operation was proposed by Wang et al in [19]. Further, work was carried out which was independent of prevalence threshold through generation of maximal cliques and was undertaken in [20].

MapReduce based algorithms were designed for massive spatial data with parallel framework [21][22]. They require a large number of nodes in case of dense data. Parallel mining of co-location patterns is carried out on graphics processing unit (GPU's) in [23]. It requires similar distribution of all features to work efficiently.

A kernel-density-based model has been adopted to compute the prevalence index and is also based on the distance-decay property for co-location mining in [24]. Fraction score for quantifying the prevalence for the candidate pattern has been developed to reduce overcounting of data instances [25]. In [26], the algorithm is based on Delaunay triangulation and use of statistical information of the vertices in Delaunay triangulation to obtain k-order neighborhood instead of distance threshold. In [27], authors have adopted Voronoi diagrams to capture the neighborhood and normalized distance between instances. Whereas [28] adopts density peak clustering based technique for spatial neighborhood analysis which gives clusters exceeding the density threshold.

Further, to overcome the above stated issue where the dense data poses a huge challenge on processing. And requires users to provide a suitable distance and prevalence threshold. Moreover, no overcounting or ignorance of instances have to be dealt with. And where absolutely minimal work has been carried out on spatio-temporal co-location patterns. We propose a novel technique for spatio-temporal colocation pattern mining.

III. PROPOSED METHODOLOGY

The co-location pattern is a process to discover objects types that are often geometrically and temporally close to other objects types in the targeted study area. The object type is called as a feature that has multiple instances in the dataset. The object type could also be the events occurring at a specific time. Associated with every object instance is a location given in the geodetic coordinate system and the time of recording or occurrence of event. Thus, spatio-temporal dataset contains a collection of instances identified by id, time, location, and feature type.

Co-location pattern mining aims to discover a class of patterns based on the relationship of neighbourhood. The relationship with neighbors could be dependency for coexistence and survival. For an example, pharmacy stores are located near to the Hospitals, here the pharmacy stores receive benefit from the coexistence of Hospital due to large number of patients visiting for medical examination and are prescribed medicines. Another example is, many bank ATMs are found co-located near to the facilities visited by large numbers of people like railway stations, petrol pumps and offices.

Consider the dataset is a collection of Instances I, over a set of features F, time of occurrence given by T.

Definition 1 (Spatio-temporal dataset): Is a collection of instances or events I_i represented as <f_i,l_i,t_i>, consisting of set of non-spatial features f_i, spatial features as the location l_i in terms of longitude(Lon) and latitude(Lat) of the event. And the time of event happening or recording as t_i.

In co-location mining of patterns, we capture the neighborhood of objects to find the proximity between them and is most often bounded by threshold.

Definition 2(spatio-temporal neighborhood NR):The neighborhood relationship NR for any two Instances I_i and I_j of different features types is the geodesic and temporal distance between the two instances is close and is necessarily less than the distance and temporal threshold. As given in equation 1.

$$NR(I_i) = \{ I_j \mid I_j.F \neq I_i.F \wedge geodesicdistance(I_i, I_j) \leq distancethreshold \wedge (I_i.time, I_j.time) \leq timethreshold\} \quad (1)$$

where geodesic distance is given by haversine formula as in equation 2.

$$geodesicdistance(I_{i,}I_j) = 2 * EarthRadius * arcsin\sqrt{Sin^2(I_j.Lat - I_i.Lat) + Cos(I_{i.}.Lat).Cos(I_j.Lat) + Sin^2(I_j.Lon - I_i.Lon)} \quad (2)$$

with the Earth radius as 6371kms and distancethreshold is user specified parameter. The distance threshold is the maximum allowed distance to consider two locations as neighborhoods. We could regard the maximum distance of 500 meters to 1km as the distance threshold and the time interval is the time span between the two events.

Clustering in spatio-temporal domain is a classical problem of classification that assigns a single cluster label to all the neighbors data objects that are extremely close to each other that exhibit similar property [29]. In co-location mining we aim to group a collection of object instances of different types to form clusters.

For a feature selected as a core point and its set of neighbours form clusters for example (see fig. 2).The figure represents various data, grouped to form clusters. The candidate co-location patterns formed by clustering the features are cluster 1 includes University, Book store, Bus stop, Medical store and Shopping mall.Cluster 2 groups Railway station, Bus stop, Bank, medical store, shopping mall.Note that the features in a cluster could be redundant and are commonly included in a pattern as a single feature. As in the given example, the spatial bandwidth is the spatial distance threshold which provides the range to determine spatial neighbours. We assume spatial data far away does not exhibit a relationship or dependency. The temporal bandwidth estimates the higher correlation between data points that are closer in time than the ones that are far away. This reflects the time decay factor. The temporal factor aids in understanding the reaction or dynamics of entities involved. Our problem is to determine the co-location relationship on geographical data with respect to time.

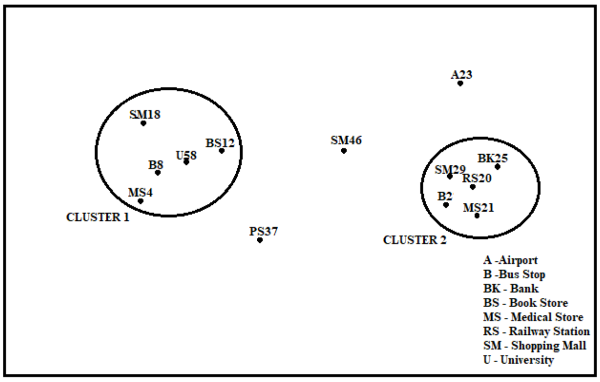

Figure 2. Co-location candidate patterns as circles representing clusters in a grid cell.

Definition 3 (Candidate co-location pattern CP): CP is a candidate co-location pattern which is a subset of non-spatial features over a specified region or a study area.

Definition 4(Participation ratio)The participation ratio is described as the proportion of feature instances within set CP relative to the total number of feature instances, presented as a percentage.

Definition 5(Participation Index)The Participation Index reflects the level of interaction or interest between the features of the instance.

Definition 6 (Spatio-Temporal Co-location Patterns STP) Is a collection of candidate co-location patterns which has high positive interaction between the objects and represented as $g_i \rightarrow g_j$ where g_i and g_j are set of non-spatio-temporal features and have no common features.

Algorithmic Description

Step 1:Form the grid cells based on maximum and minimum geodesic coordinates and divide them in equal distance. Read the data points and assign them to the grid.

Algorithm 1: Allocate data instances to grid cells.

1. Data←Read(Dataset)

2. <minLat, minLon> ← GetLatLon(Data)

3. <maxLat,maxLon> ← GetLatLon(Data)

4. Grids←formGrids (<minLat, minLon>,<maxLat,maxLon>,minDist)

Grid is essential as we want to calculate geodesic distance between two points. If the two data points grids are far or not the same nor nearby neighboring grid cells, they cannot be closer. Then there is no need to calculate their geodesic distance. As they may not belong to the same cluster.

Step 2: Understand the data points distribution. Compute the statistics about how many points are present for every feature.

For every feature find the data points. This helps to understand the population size of every feature.

Algorithm 2: To find the count of the features

Module GetFeature(I_i)

1. For each Instance I_i in Data do
2. F_k ← GetFeature(I_i)
3. FeatureBucket[F_k]. append(I_i)
4. End for
5. For each Feature F_k in Data do
6. WeightFeatureBucket[F_k] ←count(FeatureBucket[F_k])
7. End for
8. Return

Definition 6 (Density of grid cell denoted as η): Is the collection of instances in the grid cell gi and is given as.

$$\eta=\log(|\text{set of instances in grid cell gi}|) \quad (6)$$

Step 3: To capture the demography of the entire study area, the core points as the cluster centroids are chosen randomly depending upon the density of the grid cell. Apply single iteration of clustering to assign data instances to the closet cluster centroid to form clusters based on the distance and time threshold.

Algorithm 3: Form clusters for neighbours.Generate candidate colocation patterns.

1.For each grid cell $g_i \in$ Grid do

a. Choose η random corepoints from the cell to form cluster centroids.

b. for each instance I_i belonging to the grid g_i

b.1.Find the closet centroid and assign the instance to its cluster C_k if the geodesic distance between them is less than distance threshold and time threshold.

$CP_k = CP_k \cup F_i$ iff $NR(I_k)=I_i$

c. Form clusters to generate candidate patterns.

d. Call RemoveRedundantFeature(CandidatePattern).

e. Insert in HashStructure(CandidatePattern).

2.End For.

Algorithm 4:RemoveRedundantFeature

1.P← GetCandidatePattern(CP_k)

2.for each feature f_i in P

3. for each feature f_j in P where $i \neq j$

4. if $f_i == f_j$

5. P← P-{F_j}

6.return(P)

The above module accepts as an input a candidate pattern with or without redundant features and eliminates the redundant features in it. So that the current candidate pattern comparison with already existing patterns becomes easier and reduce the memory storage requirement by using condensed patterns.

Algorithm 5:HashStructure(Candidate Pattern)

/* Hash key generation module */

1.P_i← Candidate Pattern

2.initialize key←00000000(size of features in binary)

//form cluster and represent it as features binary candidate pattern

3.Convert P_i into binary bits and assign to key

//set key ←10101010 by OR with 00000000

set key ← binary candidate pattern

4.For each hashkey $h_k \in$ HashKeySet do

5. if (h_k binary AND key ==1)

6. set Bucket[key]:=S U{P_i} // S is the set of candidate pattern

7. set PatternCount[key]=PatternCount[key]+1

8. else

9. set Bucket[key]:={P_i}

10. set PatternCount[key]=1

11. end if

12.End For

Step 4: Pattern Mining

Algorithm 6: Spatiotemporal Pattern mining

1: Read the hash table

2: //For each pattern,

3:For each candidate as key $\in$ hash table

4: PR=count(hashtable[candidate]) //count the number of instances.

5: if(PR<min_count_threshold)

6: continue;

7: else

8: i=0

9: //Find the temporal difference

10: for each value ϵ hashtable[key]:

11: if(spatial_dist(hashtable[key][i],hashtable[key][i+1])<2)

12: diff=[abs(time_dist(hashtable[key][i],hashtable[key][i+1]))]

13: i++;

14: diff.sort(); // The maximum value will go to end of the list

15: $M(t_{ij}) = \frac{\sum_{i=0, j=i+1}^{n} diff_{ij}}{n}, diff_{ij} < \overline{diff_i}$

16: if (M(tij)>min_time_sup)

$STP \leftarrow STP \cup Key$

$MeanOccurenceRate \leftarrow M(t_{ij})$

$MeanDistance \leftarrow Spatial\_dist(hashtable[key][i], hashtable[key][i+1])$

17 end for

18: Display STP.

19: Stop

The workflow has been summarized in the figure 3.

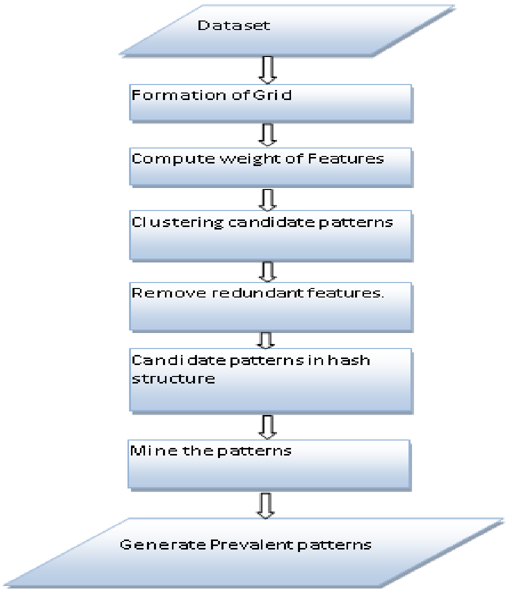

Figure 3. Workflow of proposed colocation mining algorithm

IV. EXPERIMENT RESULTS

In this section, we present experimental results on spatial and spatio-temporal data using the proposed approach. As python is an open source, high level programming language and is adopted by many researchers for data analysis and visualization. We have developed the program using Python programming in Colab environment, which is a Jupyter notebook environment running in google cloud which offers storage of data along with the TPU processing capability.

We have used the two dataset. First one is spatial synthetic dataset containing 999 records of features like education,residential, healthcare etc. Second is the swarm behaviour dataset.

Figure 4 and 5 shows the result of the pattern mining on the spatial dataset. On the x axis is the longitude and y axis is the latitude plotted. The plot has various colors for various features to form patterns. Like red color is residential, healthcare is shown in magenta, education in blue, hotel in cyan, insurance in green, retails in cornflower blue etc.

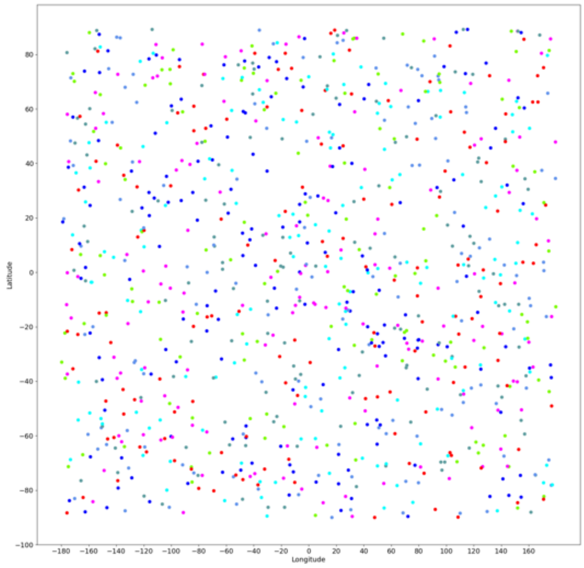

Figure 4: Visualization Result of proposed Colocation Pattern mining Algorithm on Spatial dataset.

The visualization result has been summarized below in the output figure 5. The result generated is colocation pattern healthcare is most often found with residential with support of 0.16. Insurance and residential pattern is supported with 0.12. Education and residential is a pattern with support of 0.11.Financial institution with residential is found co-located with support of 0.08.

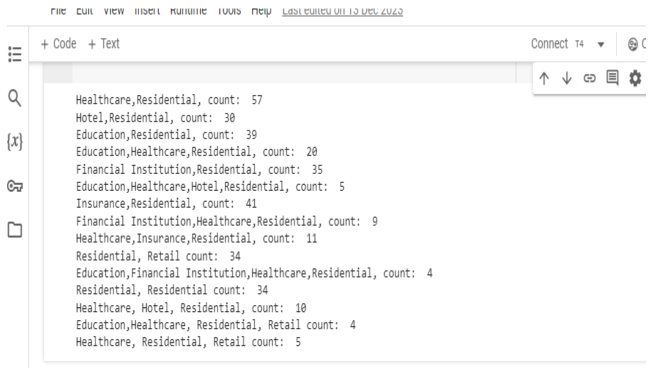

Figure 5: Result of proposed Colocation Pattern mining Algorithm on Spatial dataset.

The flock data, Swarm behavior has been collected from UCI repository which is a standard dataset and has been referred in [30]. We have used for our experiment the grouped dataset of birds called as boids. It is a spatiotemporal dataset of 200 birds with movement in their location. Each record is about the boids spatial location in current interval. We have considered, each interval as a new time interval and have used the spatial location of boids as the spatio-temporal dimensions. The data is available for download on the webaddress: https://archive.ics.uci.edu/dataset/524/swarm+behaviour

Preprocessing is performed on the boids data. Each boid is represented by x and y location that is the longitude and latitude of birds along with the velocity and other parameters. We have included only x and y attributes and have dropped the other attributes of the birds. So in all we have 400 columns for 200 birds and the experiment is run with 2500 records in each batch. The original dataset had 24017 records of intervals.

```
10 {116: [3, 26, 91, 158, 232, 291, 351, 496, 627, 815, 1007, 1058, 1113, 1165, 1227, 1282, 1368, 1481,
34 {258: [3, 26, 91, 232, 291, 1058, 1113, 1165, 1227, 1282, 1368, 1481, 1621, 1683, 1698, 1723, 1748,
188 {228: [3], 340: [176]}
310 {392: [3, 26, 91, 158, 232, 291, 351, 406, 496, 627, 815, 1007, 1368], 396: [1680]}
14 {192: [6, 54, 281, 346, 418, 544, 741, 981, 1047, 1110, 1156, 1208, 1247, 1290, 1382, 1529, 1654, 17
92 {200: [6, 18, 54, 72], 340: [573, 596, 621, 647], 232: [1379], 158: [1693, 1736], 312: [1943]}
136 {206: [6], 320: [226, 228], 390: [242]}
178 {334: [6], 210: [2478]}
202 {232: [6, 54, 125, 208, 281, 346, 418, 544, 741, 981, 2122, 2149, 2175, 2203, 2228, 2261, 2430, 246
130 {296: [11], 370: [2331, 2366, 2401, 2430, 2460]}
278 {378: [11], 340: [731, 814, 929]}
138 {146: [12], 338: [18, 72, 273, 281, 340, 346, 393, 418, 492, 628, 829, 1015, 1076, 1147, 1210, 1251
78 {354: [18, 72, 137, 209, 273, 340, 393, 492, 628, 829, 1015, 1076, 1210, 1251, 1297, 1394, 1541, 167
98 {182: [18, 72, 137, 209, 1210, 1251, 1297], 102: [1233], 398: [1870], 308: [1877, 1928, 1966, 2004]}
334 {378: [18, 72, 137, 209, 273, 340, 393, 492, 628, 829, 1015, 1980, 2085, 2173], 360: [105, 114], 38
38 {238: [20], 340: [146], 294: [393, 492], 256: [1721], 56: [2373]}
0 {208: [26, 91, 158, 232, 291, 351, 406, 1748, 1800, 1853, 1906, 1959, 2015, 2074, 2129, 2177, 2260, 2
124 {324: [61], 268: [1481, 1621, 1698, 1748, 1800, 1853], 306: [1726, 1775, 1816], 300: [2226], 258: [
140 {150: [61], 230: [1566], 194: [1721, 1778, 1830, 1885], 358: [1959, 2015, 2074]}
96 {262: [64], 208: [2015, 2074, 2129, 2177, 2222, 2260, 2293, 2372, 2417, 2461]}
228 {306: [64], 392: [105, 114], 388: [971]}
60 {130: [72, 137, 209], 74: [209, 273, 340, 393, 492, 628, 2311, 2349, 2389], 118: [1156, 1208, 1247,
```

Figure 6. The clustered co-location spatio-temporal patterns stored as hash map table along with the set of time intervals of boids found co-located.

```
10 {116: 28, 24: 1, 44: 1, 86: 1}
34 {258: 30, 350: 2, 234: 1}
188 {228: 1, 340: 1}
310 {392: 13, 396: 1}
14 {192: 18, 348: 5, 226: 1, 322: 1, 296: 10, 160: 2, 212: 1}
92 {200: 4, 340: 4, 232: 1, 158: 2, 312: 1}
136 {206: 1, 320: 2, 390: 1}
178 {334: 1, 210: 1}
202 {232: 19, 286: 1}
130 {296: 1, 370: 5}
278 {378: 1, 340: 3}
138 {146: 1, 338: 21, 158: 4}
78 {354: 35, 374: 5, 324: 16}
98 {182: 7, 102: 1, 398: 1, 308: 4}
334 {378: 14, 360: 2, 384: 2}
38 {238: 1, 340: 1, 294: 2, 256: 1, 56: 1}
0 {208: 21, 96: 22, 328: 1}
124 {324: 1, 268: 6, 306: 3, 300: 1, 258: 2}
```

Figure 7: A snapshot of clustered co-location spatio-temporal patterns with the count of times the two boids found co-located.

Figures 6 and 7 present the intermediate results of the proposed co-location algorithm. The clustering result groups the objects (i.e. boids) that are closer to each other with the distance radius of 2. Here value 2 is the difference between x and y coordinate is set to be less than 2 to find the proximity. It is interesting to note that by setting the radius as 20, 10, 5 results into capture of patterns where the objects are situated very far and results into large number of patterns generation where mining the patterns requires considerable time. But as the radius distance was set to 4, 2 and 1 discovered less pattern, as it places a tight requirement of proximity in terms of spatial and temporal distance.

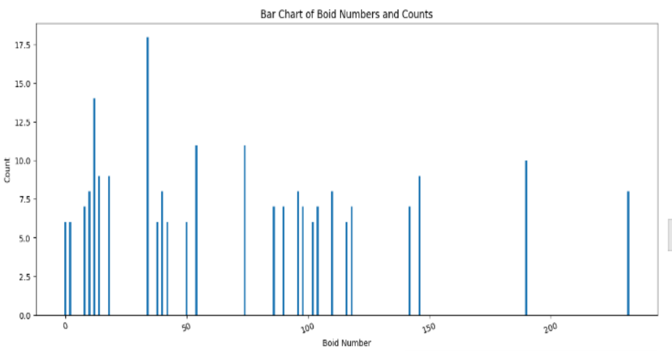

Figure 8: The boids with maximum number of co-location with other boids.

The figure 8 presents the result of size as the count which indicates the boids with maximum of number of co-location with other boids. The original dataset had 24017 records of intervals. We run the algorithm over a batch of 2500 time intervals and copy and merge the result in json file object.

V. CONCLUSION

The proposed co-location mining algorithm has been performed on spatial and spatio-temporal dataset.The algorithm is effective as even on a large dataset the algorithm has given better result.Mining is greatly affected by the radius of distance used to select the neighbours. If the distance is too large will result into many neighbours being selected. If the distance is set to low as 1 it results into very few neighbours getting selected. So if the distance is kept near to 2 has given better result.The result shows we have successfully applied the algorithm on spatio-temporal dataset.

Author contributions

Swati Meshram: Conceptualization, Methodology, programming, Writing-Original draft preparation, Validation.

Kishor P. Wagh: Supervision,Reviewing and Editing

Conflicts of interest

The authors declares that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Y. Chen, X. Chen, Z. Liu, and X. Li, “Understanding the spatial organization of urban functions based on co-location patterns mining: A comparative analysis for 25 Chinese cities,” Cities, vol. 97, p. 102563, Feb. 2020, doi: 10.1016/j.cities.2019.102563.

[2] Rogelio Santana-Arias, Tony Sam George, Aldanely Padrón-Salas, Xiomara Sarahí Sanjuan-Meza, Erick Alberto Landeros-Olvera, Patricia Elizabeth Cossío-Torres, The relationship between density of drug outlets, crime hot spots and family factors on the consumption of drugs and delinquent behaviour of male adolescent Mexican students”,Social Science & Medicine, Volume 279, 2021,113985,ISSN 0277-9536, https://doi.org/10.1016/j.socscimed.2021.113985.

[3] Calvo-Jurado, Carmen. "Modelling affected regions by the Iberian Quercus disease with proximity diagrams." Environmental Science and Pollution Research 28, no. 26 (2021): 34074-34086.

[4] J. Li, A. Adilmagambetov, M. S. Mohomed Jabbar, O. R. Zaïane, A. Osornio-Vargas, and O. Wine, “On discovering co-location patterns in datasets: a case study of pollutants and child cancers,” Geoinformatica, vol. 20, no. 4, pp. 651–692, Oct. 2016, doi: 10.1007/s10707-016-0254-1.

[5] S. Shekhar, M. R. Evans, J. M. Kang, and P. Mohan, “Identifying patterns in spatial information: A survey of methods,” WIREs Data Mining Knowl Discov, vol. 1, no. 3, pp. 193–214, May 2011, doi: 10.1002/widm.25.

[6] Xiaojing Yao, Xufeng Jiang,Dacheng Wang, Lina Yang, Ling Peng,Tianhe Chi,”Efficiently mining maximal co-locations in a spatial continuous field under directed road networks”, Information Sciences,Volume 542,2021,Pages 357-379, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2020.06.057.

[7] H. Alatrista Salas, S. Bringay, F. Flouvat, N. Selmaoui-Folcher, and M. Teisseire, “The Pattern Next Door: Towards Spatio-sequential Pattern Discovery,” in Advances in Knowledge Discovery and Data Mining, P.-N. Tan, S. Chawla, C. K. Ho, and J. Bailey, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 157–168. doi: 10.1007/978-3-642-30220-6_14

[8] X. Wu, C. Zhang, and S. Zhang, “Efficient mining of both positive and negative association rules,” ACM Trans. Inf. Syst., vol. 22, no. 3, pp. 381–405, Jul. 2004, doi: 10.1145/1010614.1010616

[9] G. Zhou, Z. Wang, and Q. Li, “Spatial Negative Co-Location Pattern Directional Mining Algorithm with Join-Based Prevalence,” Remote Sensing, vol. 14, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/rs14092103.

[10] Y. Jiang, L. Wang, Y. Lu, and H. Chen, “Discovering both positive and negative co-location rules from spatial data sets,” in The 2nd International Conference on Software Engineering and Data Mining, Jun. 2010, pp. 398–403.

[11] Qiliang Liu et.al,“An adaptive detection of multilevel co-location patterns based on natural neighborhoods,” International Journal of Geographical Information Science, vol. 35, no. 3, pp. 556–581, 2021, doi: 10.1080/13658816.2020.1775235.

[12] M. Celik, J. M. Kang, and S. Shekhar, “Zonal Co-location Pattern Discovery with Dynamic Parameters,” in Seventh IEEE International Conference on Data Mining (ICDM 2007), Oct. 2007, pp. 433–438. doi: 10.1109/ICDM.2007.102.

[13] W. Ding, C. F. Eick, X. Yuan, J. Wang, and J.-P. Nicot, “A framework for regional association rule mining and scoping in spatial datasets,” Geoinformatica, vol. 15, no. 1, pp. 1–28, Jan. 2011, doi: 10.1007/s10707-010-0111-6.

[14] J. Lu, L. Wang, Y. Fang, and J. Zhao, “Mining strong symbiotic patterns hidden in spatial prevalent co-location patterns,” Knowledge-Based Systems, vol. 146, pp. 190–202, Apr. 2018, doi: 10.1016/j.knosys.2018.02.006.

[15] S. Shekhar and Y. Huang, “Co-location rules mining: A summary of results,” in Proc. SSTD, 2001, pp. 236–256.

[16] Y. Huang, S. Shekhar, and H. Xiong, “Discovering colocation patterns from spatial data sets: A general approach,” IEEE Trans. Knowl. Data Eng., vol. 16, no. 12, pp. 1472–1485, Dec. 2004.

[17] J. S. Yoo and S. Shekhar, “A partial join approach for mining co-location patterns,” in Proc. ACM IWGIS, 2004, pp. 241–249.

[18] S. Rinzivillo and F. Turini, “Extracting spatial association rules from spatial transactions,” in Proc. Int. Workshop Geograph. Inf. Syst. (GIS), 2005, pp. 79–86.

[19] L. Wang et al., “A new join-less approach for co-location pattern mining,” in Proc. CIT, 2008, pp. 197–202.

[20] S. K. Kim, Y. Kim, and U. Kim, “Maximal cliques generating algorithm for spatial co-location pat-tern mining,” in Proc. FTRA Int. Conf. Secure Trust Comput., Data Manage., Appl., Jun. 2011, pp. 241–250.

[21] M. Sheshikala, D. Rao and R. Prakash, “Parallel approach for finding co-location pattern–a map reduce framework”, Proc. Comput. Sci. 89 (2016) 341–34

[22] Yoo, J.S., Boulware, D. & Kimmey, D. “Parallel co-location mining with MapReduce and NoSQL systems.”, Knowl Inf System,62(11),1433–1463(2020).https://doi.org/10.1007/s10115-019-01381-y

[23] W. Andrzejewski, P. Boinski,”Efficient spatial co-location pattern mining on multiple GPUs,”Expert Systems with Applications,Volume 93,2018,Pages 465-483, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2017.10.025.

[24] Xiaojing Yao, Liujia Chen, Ling Peng, Tianhe Chi,"A co-location pattern-mining algorithm with a density-weighted distance thresholding consideration",Information Sciences, Volume 396, 2017, Pages 144-161, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2017.02.040.

[25] H. K. -H. Chan, C. Long, D. Yan and R. C. -W. Wong, "Fraction-Score: A New Support Measure for Co-location Pattern Mining," 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 2019, pp. 1514-1525, doi: 10.1109/ICDE.2019.00136.

[26] Tran, Vanha, Lizhen Wang, and Hongmei Chen. "A spatial co-location pattern mining algorithm without distance thresholds." In 2019 IEEE International Conference on Big Knowledge (ICBK), pp. 242-249. IEEE, 2019.

[27] Xiaojing Yao, Liujia Chen, Congcong Wen, Ling Peng, Liang Yang, Tianhe Chi, Xiaomeng Wang and Wenhao Yu ,”A spatial co-location mining algorithm that includes adaptive proximity improvements and distant instance references”, International Journal of Geographical Information Science, 32:5, 980-1005,2018, DOI: 10.1080/13658816.2018.1431839

[28] Fang, Y., Wang, L., Hu, T. “Spatial Co-location Pattern Mining Based on Density Peaks Clustering and Fuzzy Theory.” In: Cai, Y., Ishikawa, Y., Xu, J. (eds) Web and Big Data. APWeb-WAIM 2018. Lecture Notes in Computer Science, vol 10988. Springer, Cham.,2018. https://doi.org/10.1007/978-3-319-96893-3_22

[29] Meshram, S., Wagh, K. P. Mining Intelligent Spatial Clustering Patterns: A Comparative Analysis of Different Approaches. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) March 2021,(pp. 325-330). IEEE

[30] Shadi Abpeiker, Kathryn Kasmarik,Motion behaviour recognition dataset collected from human perception of collective motion behaviour,Data in Brief,Volume 47,2023,108976,ISSN2352-3409, https://doi.org/10.1016/j.dib.2023.108976