Graph Representation and analysis of spatial data set

M.Vasavi [1] Dr.A.Murugan [2]

K Venkatesh sharma 3

1 Research Scholar, SRM Institute of Science & Technology, Kattankulanthur, India

2 Professor, SRM Institute of Science Technology, Kattankulanthur, India

2 Professor, SRM Institute of Science Technology, Kattankulanthur, India

Abstract: -

Cascading spatial temporal pattern mining is the process of getting a partial order set of getting space and time in one order pair. The order pairs are disjoint and unique with location constraint. In this article taken the crime data set and represented the ordered pairs as nodes. The event occurred next is taken as edge from one node to another node. Graph terminology as homogeneous and heterogeneous with kinds of problems are solves. Representation of Graph gives us the crime data analysis with location wise and helping us to predict the next occurrence instance. An alternate way of modeling the objects in data sets is to represent those using graphs. Frequent pattern discovery is by sub graphs from entire data sets. Experiment evaluation of the performance of a pattern using data sets.

Key words: Partial order set, Cascading spatiotemporal pattern, crime data analysis.

Introduction: -

Spatiotemporal data are used in social applications including climate safety and health. Data represented as location and time slot. These data may also contain and related type disease or crime instance data. Sequential and non-sequential patterns for large data sets have mining terminologies taken [1],[2],[3],[4],[5],[6]. General purpose patterns mining algorithms for spatial, geometric and location wise nature of data sets gives characterize each domain.

In this approach each entity of the event as edge corresponds to the relations of each edge with an event. Frequent pattern discovery as sub graph over the data set. The advantages or graph representation is given by various researchers [7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17], for representation of edges in events and nodes as relationships occurrence. Techniques for sub structuring of a graph are given in [18],[19],[20],[21], and with classification is framed in the article as instances [22]. Framing of the algorithm for a frequent pattern with computational terms is the main objective of this article. Finding frequent patterns with sub graph expansion and applying apriori [1].

Main objective implementation is graph representation to minimize the storage representation, minimize the storage and computation, generating candidate pattern effectively, applying optimization for each candidate key generation enabling scaling of large data sets and labeling the sub graphs.

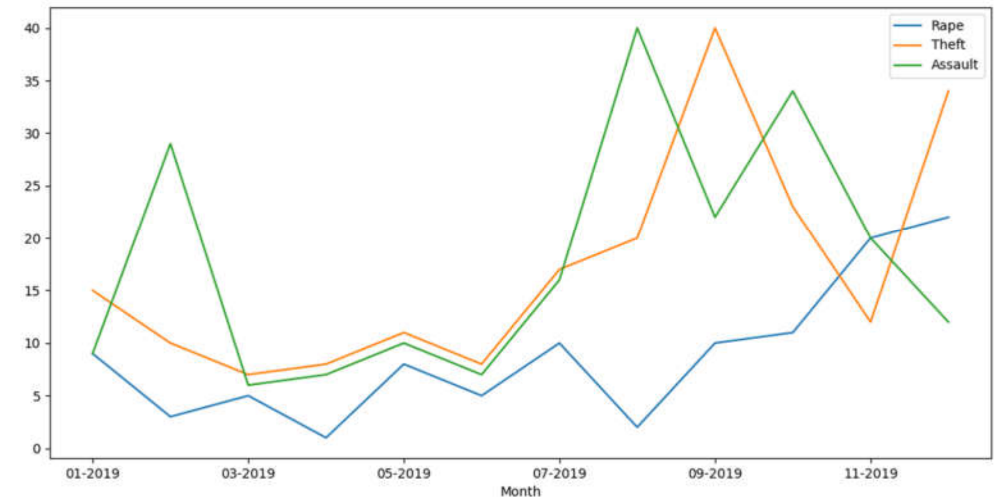

Figure1-ST dataset crime event result from location-

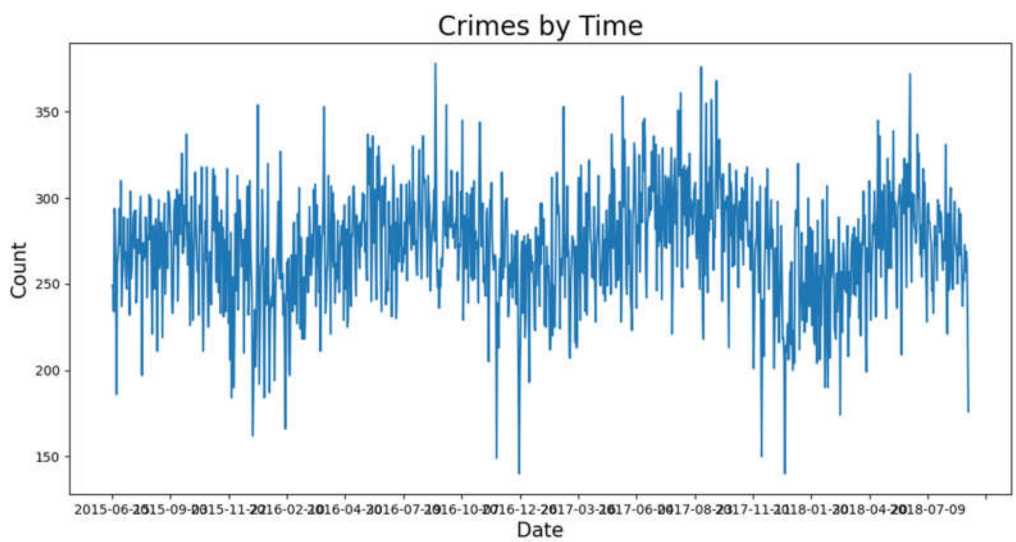

figure 1.1 (a) shows the time slot for each eventSpatial feature include each event type and interest of time slot.

Handling ST data is often handling time and space semantics as ST autocorrelation, partial order and heterogeneity. Implicit relationships between the events and frequent pattern mining techniques [23,24]applied to spatial-temporal order of data input and columns. For any event to the relation of each instance the individual event. The figure shows an aggregate of the dataset and the partitions. The representation is relationship between the time and location information in an occurrence of data instances. The relationship of each event may be viewed as partial order set both totally ordered and unordered. Event instances are taken as the data semantics in implicit partial order and occurrence and each instance of o-occurrence.

This article had shown the directed acyclic graph terminologies of each instances and representation of each partial order set of events in each location with filtering approaches to get the prediction of next occurrence of crime data.

Basic Applications Terminology: -

Cascading spatial-temporal data set may provide the information as geographic changes , public safety [25,26,27,28,29,30], public health [31], ecology and climate science with effect of each [32,33,34]. Crime generators and crime attractors [35,36,37]. A well knows theoretical area of study in public safety is environmental criminology [38]. Crime patterns may help for law enforcement and mitigation of crime. Figure 1.3(a) crime generators in the location.

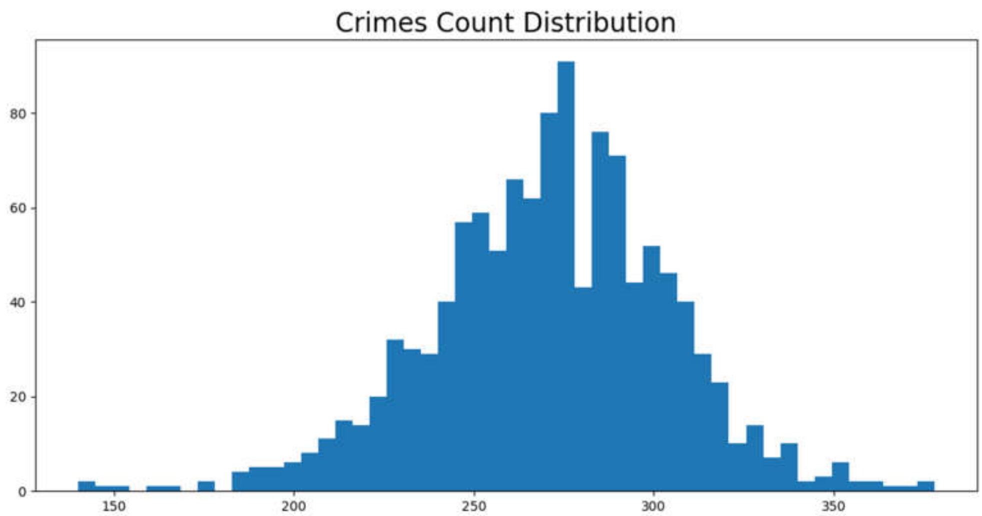

figure 1.2(b) shows the total crimes at different locations in particular year.

Challenges: -

Data are presented for connected data set and generate of connected data sets such as Social networks, biological networks and transportation systems. This article gives the graph representation of the connected events of partial order set of crime data instances [39].
Basic representation of connected data is possible by graph data structure to build predictive models of each connected data sets. A graph data structure has nodes and edges. Nodes represent the event tim e and location. The edge represents next occurrence of an event or leading event.

Node Terminology with types: -

Graph with one node and one edge type is homogeneous. Graph with two or more types of nodes and types of edges called heterogeneous. An online crime event with next partial order set of occurrences of an event is given as nodes and edges are given with location wise. Another classifies terminology is static and dynamic[39]. Static graphs do not change with time and dynamic graph changes its structure overtime. When data nodes are added, the new instances of each crime event are helping us to predict the next occurrence of crime event and the generators of crime. Removing the node and edge will change the event occurrence instances. When we apply the filters to each data set with time and space values in iteration of the filter, then the graph with the node and edge connectivity will be dynamically updated as space is taken into consideration. Basic terminology is having some problems with the node and edge representation for partial order set of events. To solve the problems, in this article, we see the various categories of connected data representations for each data instance as partial order and disjoint data sets.
Solution 1: - Crime data instance as partial order data pairs with node classification, edge prediction, crime detection and graph identification in homogenous graph terminology.
Solution 2: - crime data instance as partial order data pair with node classification, edge prediction, crime detection and graph identification in heterogeneous graph terminology.

Discussion of solutions1 and solution2: -

Solution1: - [Homogenous]

Given a connected events instance with location data having the edge representation taken as partial order data pairs with node classification, edge prediction, crime detection and graph identifications in homogenous graph terminology.
Step 1: - Node classification for homogeneous data set with partial order set of data instances

Step 2: - Edge prediction with data instances not varying and location also constant.
Step 3: - crime detection with case 1 and case 2 analyses and getting the set of instances represented as nodes and edges.
Step 4: - over all representation of resultant crime data with the predicted pattern shown as each node with space and time and location details.

Step1: - Node classification for homogeneous data set with partial order set of data instances

Node attribute instances are the inferring missing or incomplete attributes of each node, if the event is given in the collection of nodes set of data instances. Inference gives the ability to bring the neighborhood analysis information and predicting the crime at each location.
Figure 2: - example of node classification and crime analysis report
Homogenous graph has single type of node and edge. Node data as space and time in one location example: - theft at location ‘a’ and time ‘t’ taken in area slot ‘at’. No change in term a and t with at also not changeable. These instances and inferences may not help us for crime prediction.

Usability: - Occurrence of same event in one location with one event may help , if occurrence is repeated for some instance of time slots. Law enforcement will be done based on the iterative time slots of same event type in same location.
Usability 1: - Node classification can be used in consideration of location and time as constant and graph nodes and edges are static.
Figure-3: - Constant time and location graph
Usability 2: - Node classification can be used in consideration of location and time as constant and graph nodes and edges are dynamic.

Time slot at area ‘a’ with time t1, t2,…..tn
Location if at time slot with space id as s1,s2,….sn
Area data frame with time and space in multi graph representation as t1[s2],t2[s2]…..tn[sn] can be computed for digraph notation with each time set instances collected for each space slot as sum of(time slots[si])+sum of(space slots)+area computing with coordinates. tn[sn] will be computed as per the spanning sequence from each set of node-to-node traversal to get event instances occurrence in sequence.
Step 2: - Edge prediction with data instances not varying and location also constant.
Collection of data instances with the edge taken into consideration in one location. Link prediction for finding the collection of data instances for predicting the crime occurrence with recommendation of the node as crime generators and provide crime creator terminology as event data set.

Step 3: - crime detection with case 1 and case2 analysis and getting the set of instances represented as nodes and edges.
Data sequence of instances for each time slot for
location space taken each set of iteration of space and time. Iteration gives the occurrence as pair of same set of then we take a count for each iteration for getting the most occurrences as predicted event in crime. This may help for getting crime event occurred in one location as time slot same and frame a law to stop crime.

Step 4: - over all representation of resultant crime data with the predicted pattern shown as each node with space and time and location details.

Solution 2: -[Heterogeneous]
Crime data instance as partial order data pair with node
classification, edge prediction, crime detection and graph identification
in heterogeneous graph terminology. Given a connected events instance with location data having the edge representation taken as partial order data pairs with node classification, edge prediction, crime detection and graph identifications in heterogeneous graph terminology.
Step 1: - Node classification for heterogeneous data nodes with partial order set of data instances
Step 2: - Edge prediction with data instances not varying and location varying.
Step 3: - crime detection with case 1 and case2 analysis and getting the set of instances represented as nodes and edges.
Step 4: - over all representation of resultant crime data with the predicted pattern shown as each node with space and time and location details.

Homogenous graph has single type of node and edge. Node data as space and time in different locations example:- theft at location 'a' and time 't' taken in area slot 'at'. Changes for term a and t with at also changeable. These instances and inferences may help us for crime prediction.
Usability:-
Occurrence of same event in varying location with varying event may help , if occurrence is repeate d for some instance of time slots. Law enforcement will be done based on the iterative time slots of same event type or different event type in same location or different locations.
Usability 1:- Node classification can be used in consideration of location and time as varying and graph nodes and edges are static at some time and space.
Usability 2:- Node classification can be used in consideration of location and time as varying and graph nodes and edges are dynamic.

Step 2:- Edge prediction with data instances varying and location varying.
Data instances have space and time collection with data taken at location with varying sequence.
Step 3:- crime detection with case 1 and case2 analysis and getting the set of instances represented

as nodes and edges.

Step 4:- over all representation of resultant crime data with the predicted pattern shown as each node with space and time and location details.

Conclusion:-

Data set in graph with each node represent the event with space and time and edge represents the occurrence of event sequence. Two events with the edge correspond to the directed sequence of next event. Directed representation in both static and dynamic are demonstrated with crime data set. Graph categorization is shown with basic steps in homogenous and heterogeneous as collection of data. Demonstration of graph is applied for spatial location of each event, dynamic graph with time change Structural node representation as node events and relationship as edges, event representation with crime terminology**,** Same events at different location**,** Representation of crime data as node and time of a particular event**,** example of node classification and crime analysis report, different events in same location(the graph should be multiple.), time varying and space constant graph(Static), event by considering time and location varying, sequence of event with time**,** homogeneous and heterogeneous graphs representation (two graphs in one figure), resultant graph with space and time shown with sequence of event occurrence. Filtering techniques are applied on each data set instances to get patterns of each event. References:-

[1] R.Agarwal and R.Srikant, “Fast algorithms for mining association rules,” Proc. 20^{th} int’l Conf. Very Large Data Bases (VLDB), pp.487-499,sept.1994.

[2]R.Srikant and R.Agarwal, “ Mining sequential patterns: Generalizations and performance improvements,” Proc, Fifth int’l Conf. Extending Database Technology (EDBT),pp. 3-17, 1996.

[3] R.C.Agarwal, C.C.Agarawal, and V.V.V. Prasad, “ A Tree projection Algorithm for generation of frequent item sets, j, parallel and distributed computing, vol. 61, no. 3, pp. 350-371, 2001.

[4]M.J.Zaki and K. Gouda, “Fast Vertical Mining using Diffsets,” Technical Report 01-1, Dept. of Computer science, Rensselaer polytechnic inst., 2001.

[5]J.Han, J.pei, and Y.Yin, “Mining Frequent Patters without candidate generation, “ Proc. ACM SIGMOD Int’l Conf. Management of Data, May 2000.

[6] J.Pei, J. Han, B. Mortazavi-Asl, H.Pinto, Q.Chen, U. Dayal, and M.C.Hsu, “PrefixSpan: Mining Sequential Patterns Efficiently by prefix-projected pattern growth, “Proc. 2001 Int’l Conf. Data Eng. (ICDE ’02), july 2002.

[7]H.Kalvianene and E.Oja, “comparisions of Attributed Graph matching Algorithms for Computer vision”, Proc. STEP-90, Finnish Artificial Intelligence Symp., pp. 354-368, June 1990.

[8]L.B. Holder, D.J. Cook, and S. Djoko, “Substructure Discovery in the SUBDUE System,” Proc. AAAI Workshop Knowledge Discovery in Databases, pp. 169-180, 1994.

[9]T.K. Leung, M.C. Burl, and P. Perona, “Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching,” Proc. Fifth IEEE Int’l Conf. Computer Vision, June 1995.

[10]K. Yoshida and H. Motoda, “CLIP: Concept Learning from Inference Patterns,” Artificial Intelligence, vol. 75, no. 1, pp. 63-92, 1995.

[11] Y. Amit and A. Kong, “Graphical Templates for Model Registration,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 3, pp. 225-236, 1996.

[12]E.G.M. Petrakis and C. Faloutsos, “Similarity Searching in Medical Image Databases,” Knowledge and Data Eng., vol. 9, no. 3, pp. 435-447, 1997.

[13]A. Srinivasan, R.D. King, S.H. Muggleton, and M. Sternberg, “The Predictive Toxicology Evaluation Challenge,” Proc. 15th Int’l Joint Conf. Artificial Intelligence (IJCAI), pp. 1-6, 1997.

[14]C.-W.K. Chen and D.Y.Y. Yun, “Unifying Graph-Matching Problem with a Practical Solution,” Proc. Int’l Conf. Systems, Signals, Control, Computers, Sept. 1998.

[15]L.Dehaspe, H. Toivonen, and R.D. King, “Finding Frequent Substructures in Chemical Compounds,” Proc. Fourth ACMSIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD98), pp. 30-36, 1998.

[16] D. Dupplaw and P.H. Lewis, “Content-Based Image Retrieval with Scale-Spaced Object Trees,” Proc. SPIE: Storage and Retrieval for Media Databases, vol. 3972, pp. 253-261, 2000.

[17] J. Gonzalez, L.B. Holder, and D.J. Cook, “Application of GraphBased Concept Learning to the Predictive Toxicology Domain,” Proc. Predictive Toxicology Challenge Workshop, 2001.

[18] R.D. King, S.H. Muggleton, A. Srinivasan, and M.J.E. Sternberg, “Structure-Activity Relationships Derived by Machine Learning: The Use of Atoms and Their Bond Connectivities to Predict Mutagenicity by Inductive Logic Programming,” Proc. Nat’l Academy of Sciences, pp. 438-442, 1996.

[19] A. Srinivasan and R.D. King, “Feature Construction with Inductive Logic Programming: A Study of Quantitative Predictions of Biological Activity Aided by Structural Attributes,” Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 37-57, 1999.

[20] C. Borgelt and M.R. Berthold, “Mining Molecular Fragments: Finding Relevant Substructures of Molecules,” Proc. 2002 IEEE Int’l Conf. Data Mining (ICDM), 2002.

[21]M. Deshpande, M. Kuramochi, and G. Karypis, "Automated Approaches for Classifying Structures," Proc. Second Workshop Data Mining in Bioinformatics (BIOKDD '02), 2002.

[22] C. Hansch, P.P. Maolney, T. Fujita, and R.M. Muir, "Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients," Nature,vol. 194, pp. 178-180, 1962.

[23] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining (First Edition). Number 0321321367. Addison-Wesley Publishing Company, Boston, MA, USA, 2005.

[24]Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: current status and future directions. Data Mining and Knowledge Discovery, 15(1):55–86, 2007.

[25]L.Anselin, J.Cohen, D.Cook, W.Gorr, and G.Tita. Spatial analyses of crime. Criminal justice, 4:213-262,2000.

[26] P.Brantingham and P. Brantingham. Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3):1–26, 1995.

[27]P.L. Brantingham and P.J. Brantingham. Environment, routine and situation: toward a pattern theory of crime. Routine activity and rational choice, 5(2):259–94, 1993.

[28] J.E. Eck, S. Chainey, J.G. Cameron, M. Leitner, and R.E. Wilson. Mapping crime: Understanding hot spots. 2005.

[29] J.E. Eck and D. Weisburd. Crime places in crime theory. Crime and place, crime prevention studies, 4:1–33, 1995.

[30] P.J. Brantingham and P.L. Brantingham. Environmental criminology. Sage Publications Beverly Hills, CA, 1981.

[31] David M. Morens, Gregory K. Folkers, and Anthony S. Fauci. The challenge of emerging and re-emerging infectious diseases. Nature, 430:242–249, July 2004.

[32]Committee on Strategic Advice on the U.S. Climate Change Science Program; National Research Council: Restructuring Federal Climate Research to Meet the Challenges of Climate Change. The National Academies Press, Washington D.C., 2009.

[33] Strategic plan for the climate change science program. http://www.climatescience.gov/Librar/stratplan2003/final/ccsptratplan2003-chap9.html,2003.

[34] L. Hein, M.J. Metzger, and R. Leemans. The local impacts of climate change in the ferlo, western sahel. Climatic change, 93(3):465–483, 2009.

[35] P. Brantingham and P. Brantingham. Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3):1–26, 1995.

[36] J.E. Eck and D. Weisburd. Crime places in crime theory. Crime and place, crime prevention studies, 4:1–33, 1995.

[37] P.J.Brantingham and P.L. Brantingham. Environmental criminology. Sage Publications Beverly Hills, CA, 1981.

[38]P.L.Brantingham and P.J. Brantingham. Environment, routine and situation: toward a pattern theory of crime. Routine activity and rational choice, 5(2):259–94, 1993.

[39] https://medium.com/stellargraph/knowing-your-neighbours-machine-learning-on-graphs-9b7c3d0d5896.