Analysis of Spatial Data Using Spatial Clustering In Consideration Of Disaster Scenarios

Bhagyashree Jadhav[1], Sakshi Sanap[2], Aishwarya Pangarikar[3]

[1] *BE, (Information Technology), SITRC, Nashik,Maharashtra, India.*
[2] *BE, (Information Technology), SITRC, Nashik,Maharashtra, India.*
[3] *BE, (Information Technology), SITRC, Nashik,Maharashtra, India.*

ABSTRACT

Spatial grouping of the occasions dispersed over a topographical district has numerous significant applications, including the appraisal of necessities of the individuals influenced by a debacle. In this paper we think about spatial grouping of online life information (e.g., tweets) produced by advanced cells in the catastrophe area. Our objective in this setting is to discover high thickness territories inside the influenced territory with plenitude of messages concerning explicit needs that we call just as "circumstances". Tragically, a direct spatial bunching isn't just precarious or inconsistent within the sight of portability or changing conditions yet in addition neglects to perceive the actuality that the "circumstance" communicated by a tweet stays substantial for some time past the hour of its discharge. We address this by partner a rot work with every data content and characterize a gradual spatial bunching calculation (ISCA) based on the rot model. We study the presentation of steady grouping as a component of rot rate to give bits of knowledge into how it tends to be picked fittingly for various circumstances.

Keyword: - *Spatial Big Data Analytics, Crowd Big Data, etc….*

1. INTRODUCTION

Informal communities, especially Twitter, have gotten a well-known stage for imparting data significant for salvage furthermore, recuperation during debacles [1, 2]. For instance, the accompanying data was utilized during and after the Great East Japan Quake [3, 4]: (1) The tag '# j help me' was utilized on Twitter following the quake and wave as a route for crisis faculty to quickly distinguish individuals needing salvage; (2) Google tweeted a connection on Twitter to its Google Individual Finder apparatus, which empowers individuals to scan for missing relatives. A mechanized investigation of the twitter information to remove and comprehend the "circumstance" and the requirements of the influenced individuals can take into account fast and centered assistance all together to amplify advantage to the influenced individuals [5]. Our concentration in this paper is to distinguish these "circumstances" in light of the examination of the tweets in a manner that perceives that the circumstance communicated by the tweets stays substantial over some undefined time frame as opposed to being of momentary worth as it were. Spatial huge information investigation is the study of finding covered up designs in geospatial information.

It has a great deal of utilizations particularly in crisis circumstances to help salvage and recuperation exercises during or following a debacle by comprehension the "circumstances, for example, wounds to the individuals, requirement for therapeutic care, lack nourishment or water, and so on. Spatial grouping is the procedure of collection a lot of spatial articles into groups so that the articles inside a group have high similitude than those across bunches. By applying spatial grouping, it becomes conceivable to find problem area regions, which we characterize as high thickness bunches with a given circumstance (e.g., regions where numerous individuals are harmed or need some kind of help), and track the movement of such territories. Fig. 1(a) shows the conveyance of quake related tweets (with catchphrases 'seismic tremor', and which implies catastrophe in Japanese) in the Kumamoto Earthquake that struck at Kumamoto City of Kumamoto Prefecture in Kyushu Region, Japan in 2016. Fig. 1(b) shows the shake map saw to the east of Kumamoto City [6]. This figure plainly appears that the dispersion of the debacle related tweets can give some helpful data regarding the degree of harm, the genuine needs of local people, and so on.

In this paper our principle objective is to comprehend circumstances in a hazardous situation through spatial grouping of social huge information. Such examination can be helpful to find problem area regions (or the districts of interests), where crisis supplies should be sent as quickly as time permits.

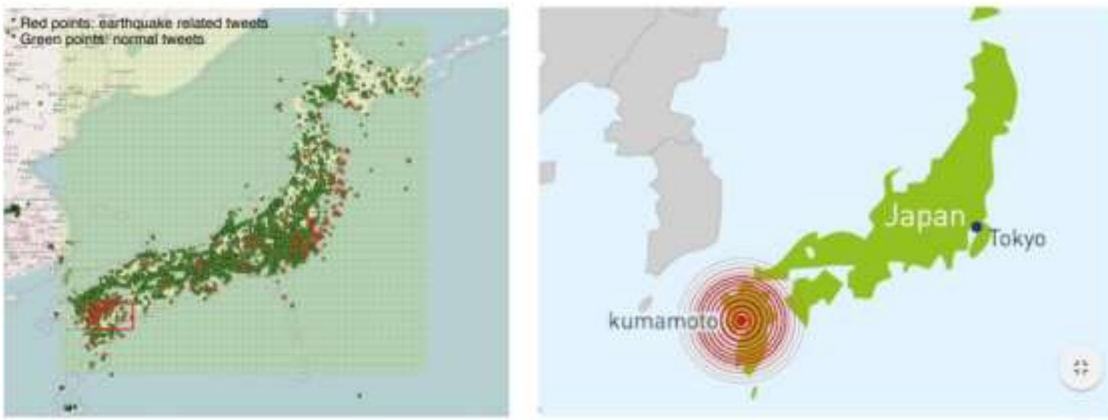

Fig -1: Distribution of Kumamoto Earthquake (April-May, 2016). (a) Hotspot of Earthquake related Tweets after Kumamoto Earthquake and (b) the region of the epicenter as obtained the ground reality

1.1 Existing System

Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work Introduction related your research work.

1.2 Objectives

We address this by associating a decay function with each information content and define an incremental spatial clustering algorithm (ISCA) based on the decay model. We study the performance of incremental clustering as a function of decay rate to provide insights into how it can be chosen appropriately for different situations.

1.3 Contribution

It's been observed repeatedly that much of the data has a popularity pattern: Very hot when the data is generated, and then the popularity wanes. The data may become hot again. For example, think of the files associated with the paper you are working on. It's popular for some days or weeks, and then there is no activity until you come back to it again. For measurement data coming from the cyber physical infrastructure or physics experiments, the new data is hot only for some time. It may become hot again, but perhaps with decreasing peak popularity, i.e., short term decay functions superimposed on a long term decay function. In storage and other systems, the popularity is often captured using caching mechanisms such as LRU, but LRU is useful only at short time scales and small access granularities (blocks or objects). At longer time scales and large blobs of data, the stickiness may provide more insights.

2. LITERATURE SURVEY

1.1 A spatial data mining method for mineral resources potential assessment

On the basis of multi-source geology spatial database and traditional spatial data mining, a spatial data mining method for mineral resources potential assessment was proposed in this paper, which the spatial characteristics and uncertainty of geology data were reasonable to consider. The method mainly include continuous geological spatial data discretization, spatial relationship abstracting and attribute transforming, mining metallogenic association rules and quality assessment, comprehensive evaluation of metallogenic association rules and potential assessment. Finally, the experiment of mineral potential assessment for iron deposits was performed in Eastern Kunlun, Qinghai province, China, using spatial data mining method and weights-of-evidence model, respectively. The results indicate that the prediction accuracy of spatial data mining was obvious higher than weights-of-evidence model's, the method is suitable for mineral resources potential assessment and its effectiveness is good.

1.2 GeoKSGrid: A geographical knowledge grid with functions of spatial data mining and spatial decision

Motivated by the lack of a geographical problem solving environment that is adequate to provide end users with reliable, open, distributed, and long lasting spatial data analyzing and knowledge discovery services, a novel geographical knowledge service platform - GeoKSGrid with functions of spatial decision support and distributed parallel data mining is described in this paper. The basic concepts and state-of-the-art knowledge in grid research is discussed first. Then, the design of system architecture and the implementation of several most important modules of GeoKSGrid are illustrated. Finally, some demonstrative applications of the geographical knowledge services in real industry contexts is examined, combining with a brief interpretation of the processing results which confirm the practical value of the services and knowledge grid platform.

1.3 DD-Rtree: A dynamic distributed data structure for efficient data distribution among cluster nodes for spatial data mining algorithms

Parallelizing data mining algorithms has become a necessity as we try to mine ever increasing volumes of data. Spatial data mining algorithms like Dbscan, Optics, Slink, etc. have been parallelized to exploit a cluster infrastructure. The efficiency achieved by existing algorithms can be attributed to spatial locality preservation using spatial indexing structures like k-d-tree, quad-tree, grid files, etc. for distributing data among cluster nodes. However, these indexing structures are static in nature, i.e., they need to scan the entire dataset to determine the partitioning coordinates.

This results in high data distribution cost when the data size is large. In this paper, we propose a dynamic distributed data structure, DD-Rtree, which preserves spatial locality while distributing data across compute nodes in a shared nothing environment. Moreover, DD-Rtree is dynamic, i.e., it can be constructed incrementally making it useful for handling big data. We compare the quality of data distribution achieved by DD-Rtree with one of the recent distributed indexing structure, SD-Rtree. We also compare the efficiency of queries supported by these indexing structures along with the overall efficiency of DBSCAN algorithm. Our experimental results show that DD-Rtree achieves better data distribution and thereby resulting in improved overall efficiency.

3. PROPOSED SYSTEM

We discuss the detailed research problem and our proposed approach using Notations are listed in Table I. In the following a topical situation refers to the main situational information that is expressed or discussed in a twitter message. A spatial or point object is the location or place that the topic is associated with, here it refers to the geotag in a twitter message. The spatial objects in rectangular shape denote the point data that are not associated with a targeted situation, whereas the ones in circular shape are the points associated with the targeted situation. In a disaster scenario, the targeted situation indicates the damage level after the disaster, so that the rescue services can be provided as soon as possible. As the tweet messages are generated continuously by the users, various

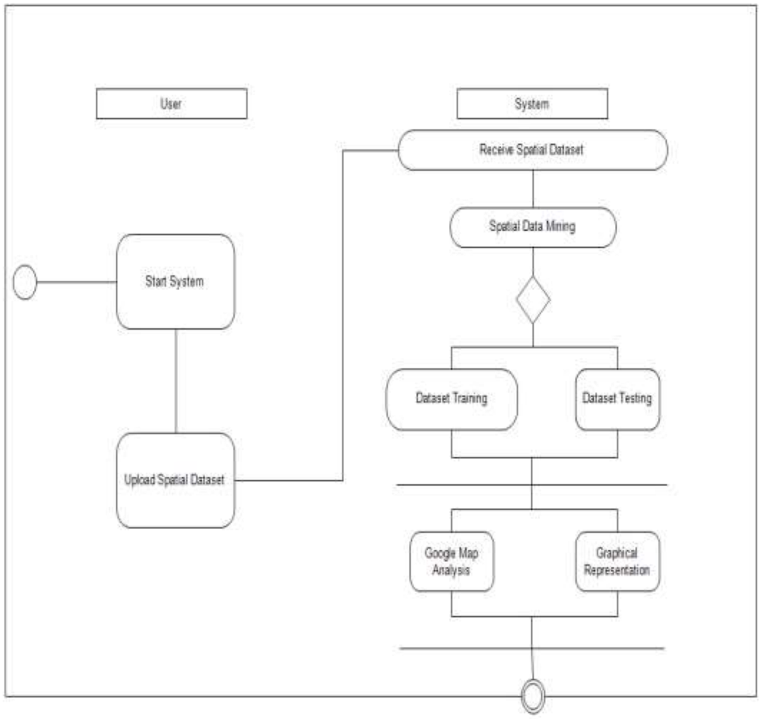

Fig -2 Proposed System

Opinions/messages exist for the same situation, which also evolves with time. Although a spatial clustering implemented at dierent time instants can nd high density areas, such an approach will lead to unstable or unreliable identication of the hot-spot areas, as shown . An unstable clustering is of little use to rescue operations which may take a signicant amount of time to plan and execute. We address this issue by associating the messages with a decay function that assigns less weights to the older messages and propose an in cremental clustering scheme so that the hot-spots regions can evolve steadily and smoothly over time Information Energy of the Tweets It's been observed repeatedly that much of the data has a popularity pattern: Very hot when the data is generated, and then the popularity wanes. The data may become hot again. For example, think of the les associated with the paper you are working on. It's popular for some days or weeks, and then there is no activity until you come back to it again. For measurement data coming from the cyberphysical infrastructure or physics experiments, the new data is hot only for some time. It may become hot again, but perhaps with decreasing peak popularity, i.e., short term decay functions superimposed on a long term decay function. In storage and other systems, the popularity is often captured using caching mechanisms such as LRU, but LRU is useful only at short time scales and small access granularities (blocks or objects).

4. SYSTEM ANANLYSIS

Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work Conclusion related your research work

5. RESULT & DISCUSSION

This section presents evaluation results of ISCA that are obtained from a synthetic dataset as well as from a real disaster-related social data. A. Simulation on Synthetic Data Set We first simulated ISCA using a synthetic database obtained from [22]. The database is composed of several datasets that model the temporal evolution of the information contents in a two dimensional space. The datasets were generated by Gaussian distributions whose mean and/or variance change over time. We use the "3C2D2400Spiral" dataset, which presents a helix alike movement of 3 clusters. These three clusters could be considered as three groups of population with dynamic ratios of the situation over the time series, as shown in Fig. The point data distribution of all three groups and the hotspot over the time series are shown in Fig. and 7(a). We can observe that the hot-spot corresponding to these two data sets is mainly concentrated on Group 2, which has the highest percentage of situation labels among the three groups. Metrics Used: We believe that the information contained in the crowd data, e.g., a tweet, about the situation remains valid for some time. For how long it remains valid is a property of the situation, and this property is modeled as energy decay in Section III-A.

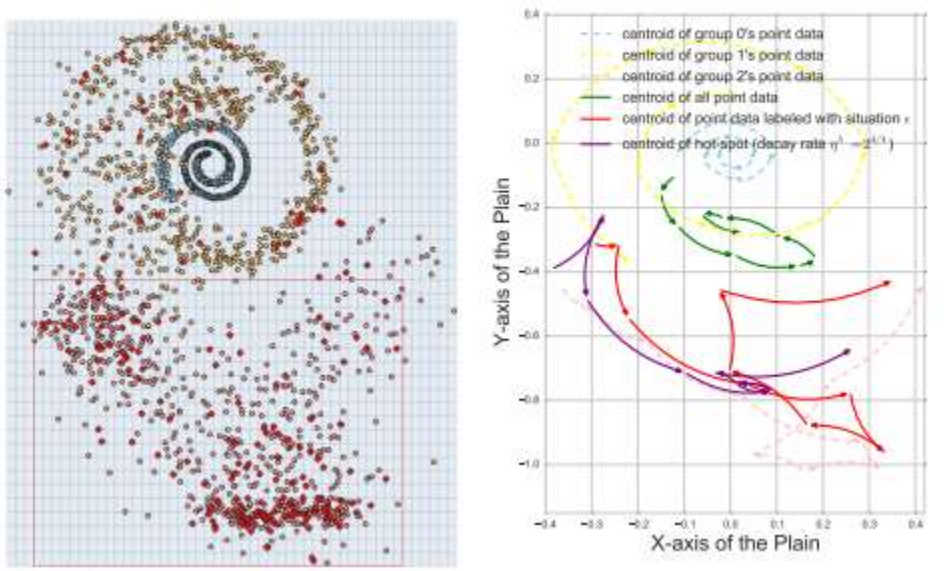

Fig -2 Overall Hot-spot & Centroid Movement.

With the exponential decay base η = 2, different values of λ (the exponential decay exponent) represent different values of 'half-life' for the decay, i.e. the time when the weight of the tweet becomes one-half. In order to estimate the effect for the decay ratio λ on the spatial clustering result under different 'half-life' periods, we introduce a sensitivity metric. Assume that the centroid of the point objects at two time instances t and t + 1 are μt and μt+1, and that of the hotspots are Ct and Ct+1 respectively.

6. CONCLUSIONS

We proposed an incremental spatial clustering algorithm based on information weight decay, in order to achieve a stable and reliable decision-making based on dynamic big crowd data. We have evaluated the proposed method by using a disaster related social data set as well as a synthetic data-set. The incremental spatial clustering using the decay model provides an adaptive observation of dynamic changes to the crowd situation. We introduce the metric of sensitivity to assist the user of this system to select proper decay rate to observe the spatial clustering result. Our Application help in analysis of disaster situation.

7. ACKNOWLEDGEMENT

We are profoundly grateful to Prof. Shital R. Bedase, our Project Guide for her expert guidance and continuous encouragement all the time since the projects commencement to its completion. We express deepest appreciation towards Prof.Prashant Gawande, our Project Coordinator, for continuously letting us know about the upcoming project competitions, improvement and additions of the modules in the project. And we must express sincere heartfelt gratitude towards Dr.Pawan R. Bhaldhare, Head of Department of Information Technology and Dr.S.T. Gandhe, Principal, Sandip Institute of Technology and Research Center, and to all the staff members of Information Technology Department who saw our growth and helped us in every way possible.

8. REFERENCES

[1]. Gao et al., “Harnessing the crowdsourcing power of social media for disaster relief,” IEEE Intelligent Systems, vol. 26, pp. 10–14, 2011.

[2] Yin et al., “Using social media to enhance emergency situation awareness,” IEEE Intelligent Systems, vol. 27, pp. 52–59, 2012.

[3] Slater et al., “Social media, information, and political activism in japan’s 3.11 crisis,” The Asia-Pacific Journal: Japan Focus, 2012.

[4] Wang et al., “Big data analytics for emergency communication networks: A survey,” IEEE Communications Surveys Tutorials, vol. 18, pp. 1758–1778, 2016.

[5] Wu et al., “Disaster network evolution using dynamic clustering of twitter data,” in IEEE ICDCS Workshops, 2017, pp. 348–353.

[6] Asian Disaster Reduction Center, “2016 Kumamoto Earthquake Survey Report (Preliminary),” http://www.adrc.asia/publications/201604 KumamotoEQ/ADRC 2016KumamotoEQ Report 1.pdf, accessed: March, 2017.

[7] Kulldorff, “Spatial scan statistics: models, calculations, and applications,” in Scan statistics and applications. Springer, 1999, pp. 303–322.

[8] Auchincloss et al., “A review of spatial methods in epidemiology, 2000– 2010,” Annual review of public health, vol. 33, pp. 107–122, 2012.

[9] Singh et al., “K-means v/s k-medoids: A comparative study,” in National Conference on Recent Trends in Engineering Technology, 2011.

[10] Xu et al., “A distribution-based clustering algorithm for mining in large spatial databases,” in IEEE ICDE, 1998, pp. 324–331.

[11] Lin et al., “A deected grid-based algorithm for clustering analysis,” WSEAS Transactions on Computers, vol. 7, pp. 125–132, 2008. 12] Bezerra et al., “A new evolving clustering algorithm for online data streams,” in IEEE EAIS, 2016, pp. 162–168.