International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Survey on Spatio-Temporal Clustering

Vaishali V. Dhundale[1], M. S. Takalikar[2]

[1]Department of Computer Engineering, PICT, Savitribai Phule Pune University

[2]Professor, Department of Computer Engineering, PICT, Savitribai Phule Pune University

Abstract: ***Clustering is one of the major data mining methods in large databases for knowledge discovery. The importance of spatial and spatio-temporal data mining is growing with the increasing attention to social media and vast amount of spatio-temporal data generated by mobile devices, GPS, weather forecasting. Thus, clustering plays important role in broad range of applications. Density based clustering DBSCAN is commonly used for spatial data mining. Various density based clustering algorithms and its modifications are used for spatio-temporal clustering. Literature survey, challenges, applications of spatio-temporal data are introduced in further sections.***

Keywords: Spatio-temporal data, density based clustering, spatio-temporal data analysis, data mining

1. Introduction

Data mining is the extraction of hidden predictive information from large databases. Data mining tools predict future trends and behaviors to make proactive, knowledge-driven decisions also aims at modeling relation-ships and discovering hidden patterns in large databases. Data mining is broadly classified into two categories supervised and unsupervised.

Clustering is unsupervised classification process of grouping large datasets according to their similarity into meaningful clusters. i.e. patterns in the same cluster are similar in some sense and patterns in different clusters are dissimilar in the same sense. Now days, clustering of spatio-temporal data is challenging task due to increased data collected from weather forecasting, medical imaging and geographic information systems also the increase in general availability of large volume of data through internet, remote sensing devices. Large numbers of applications include location-based services, climatology, land-use classification and global change using satellite imagery, finding crime hot spots, traffic system, migration of birds, fishing control, patients cancer analysis(medical imaging) and so on. Thus, new methods are needed to analyze spatial and spatio-temporal data to extract interesting, useful, and non-trivial patterns.

Spatio-temporal clustering is a process of grouping objects based on their spatial and temporal similarity. It is relatively new subfield of data mining which gained high popularity due to various applications that record position, time and environmental properties of an object in real-time system like location based service applications. Spatio-temporal data refers data of an object having spatial and temporal properties i.e. a position of an object in space over a period of time or state of an object.

The spatio-temporal object contains spatial, temporal and non-spatial attributes. Spatio-temporal object can be defined as an object that has at least one spatial and one temporal property. The spatial properties are location and geometry of the object. The temporal property is timestamp or time interval for which the object is valid. For ex. a moving object. An event that may happen at a certain time t and location x and this describes a spatial and temporal phenomenon of an object.

Spatio-temporal clustering is based on two dimensions.

1) Spatial Dimension: Considering objects associated to fixed location or move i.e. dynamic location changing with time.
2) Temporal Dimension: Full history of object is kept. Each object can change status so considering recent value (updated snapshot) is known.

Spatio-temporal data mining applications range from transportation, remote sensing, satellite telemetry, monitoring environmental resources and geographic information systems (GIS).

As discussed in [14] the significance of spatio-temporal data analysis and mining is growing with the increasing availability and awareness of huge amount of geographic and spatio-temporal datasets in many important application domains like

- Meteorology: All kinds of weather data, moving storms, changes in freezing level, droughts.
- Biology: Animal movements, mating behaviour, species relocation and extinction.
- Crop sciences: Harvesting, soil quality changes, land usage management, seasonal grasshopper infestation.
- Forestry: Forest growth, forest fires, hydrology patterns, canopy development, planning tree cutting, planning tree planting.
- Medicine: Patient's cancer developments, supervising developments in embryology.
- Geophysics: Earthquake histories, volcanic activities and prediction.
- Ecology: causal relationships in environmental changes, tracking down pollution incidents.
- Transportation: Traffic monitoring, control, tracking vehicle movement, traffic planning, vehicle navigation, fuel efficient routes.

2. Challenges in Spatio-temporal Clustering

Spatio-temporal data is complex in nature due to various reasons, so while clustering of spatio-temporal data has many issues and challenges.

- Continuous and discrete changes of spatial and non spatial properties of the spatio-temporal objects.
- Influence of neighboring spatio-temporal objects on one another. For example spread of fire is influenced by rain and changing wind speed and direction.

As spatiotemporal data carries multi-dimensional information such as time, location, geometry and non-spatial attributes of spatiotemporal objects, so the processing, analysis and mining is done for multiple levels of granularity.

Spatio-temporal data consists both spatial as well as temporal property so while clustering, controlling influence of spatial and temporal component on each other.

The attributes of neighboring patterns may have significant influence on a nearby pattern and it should be considered. For example, spatio-temporal event like hurricane will have influence on traffic jam pattern.

Scale effect in space and time is a challenging issue in spatio-temporal data analysis and mining. Scale in terms of spatial resolution or temporal granularity can have a direct impact on the kind and strength of spatio-temporal relationships [16] that can be discovered in datasets.

The unique characteristic of spatio-temporal datasets requires significant modification of data mining techniques so that they can exploit the rich spatial and temporal relationships and patterns embedded in the datasets.

3. Spatio-temporal Analysis

Spatio-temporal analysis [14] is categorized in spatial data analysis, temporal data analysis, dynamic spatio-temporal data analysis and static spatio-temporal data analysis.

3.1 Spatial Data Analysis

Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful patterns from large spatial datasets. The spatial data analysis analyzes how thematic attributes data changing with respect to a distance from a spatial reference at a specified time. For example study of change in temperature and humidity values when moving away from sea coast at a given time. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation.

3.2 Temporal Data Analysis

The temporal data analysis fixes the spatial dimension and analyzes how thematic attributes data change with time. Analysis of rainfall, temperature and humidity of a given region over a period of time is an example of this kind. Temporal data stored in a temporal database is different from the data stored in non-temporal database in that a time period attached to the data expresses when it was valid or stored in the database. Conventional databases consider the data stored in it to be valid at time instant now, they do not keep track of past or future database states. By attaching a time period to the data, it becomes possible to store different database states.

3.3 Dynamic Spatio-temporal Data Analysis

The dynamic spatio-temporal data analysis fixes thematic attributes dimension and analyzes how spatial properties change with time. Analysis of moving car data, spread of fire are examples of this category.

3.4 Static Spatio-temporal Data Analysis

The static spatio-temporal data analysis fixes the temporal and thematic attribute dimensions and studies the spatial dimension. An example of this is finding locations having same rainfall at same time. Analysis of large volume of spatio-temporal data without fixing any dimension is very difficult and complex. However the data mining can be used to uncover unknown patterns and trends within the data.

4. Literature Survey on Spatio-temporal Clustering

Clustering is one of the major data mining methods in large databases for knowledge discovery. It is the process of grouping large data sets according to their similarity. Spatial data records information regarding the location, shape and its effect on features (e.g. geographical features). When such data is time variant, it is called spatio-temporal data. Spatio-temporal data mining is an upcoming research topic which focuses on studying and implementing novel computational techniques for large scale spatio-temporal data [15]. Spatio-temporal clustering algorithms [2, 13] have to consider the spatial and temporal neighbors of objects while extracting the clusters.

In real-world applications, we experience various kinds of spatio-temporal data. Kisilevich et al. [4] divided spatio-temporal data into five categories including spatio-temporal events, georeferenced variables, georeferenced time series, moving objects, and trajectories. In spatio-temporal event data, there is a set of events, each occurred in a spatial location and coming with its timestamp. Clustering this type of data aims to find a set of events that are close to each other in both space and time. One of the commonly used methods for clustering these types of data is scan statistics [5], [6]. In this method, one moves a cylindrical window of variable size and shape, across a geographical region to detect clusters of events with the highest likelihood ratios.

In [7], an extended version of FCM has been proposed to find circular clusters of hotspots in spatio-temporal geographical information system data. For each timestamp, the events are clustered based on their spatial location, and then, a comparison between occurred clusters in consecutive time stamps has been performed to conclude some interpretations about events.

Wang et al. [8] proposed two spatio-temporal clustering methods, which are called ST-GRID and ST-DBSCAN, to detect seismic events in China and neighboring countries. The ST-GRID method used a multidimensional grid that covers the entire spatio-temporal feature space. Then, by merging the dense neighbor cells, spatio-temporal clusters were formed. ST-DBSCAN extended DBSCAN by redefining density reachability using spatial and temporal radius. Both methods exploited an ordered k-dist graph [9] to determine their parameters. Georeferenced time series are composed of a set of fixed geographical coordinates, each corresponding to one or more time series. Georeferenced variables data form a special case of georeferenced time series where only the most recent point of time series is available. Clustering this type of data aims to group objects based on their spatial closeness and temporal similarities.

In [10], FCM has been used to cluster weather time series. The Pearson correlation coefficient was employed as the similarity measure expressing closeness of two time series and a method to determine the number of clusters has been proposed. However, the method does not involve the spatial part of data in the clustering process.

Deng et al. [11] proposed a density-based spatio-temporal clustering. In this method, a spatial proximate network has been constructed using Delaunay triangulation and a spatio-temporal autocorrelation analysis was employed to define the spatio-temporal neighborhood.

In [3], an extended version of FCM was proposed for image segmentation by considering the spatial location of pixels. This method has been considered by Coppi et al. [12] for clustering spatio-temporal data. In this approach, a spatial penalty term that was calculated using a spatial contiguity matrix has been added to the objective function to guarantee an approximate spatial homogeneity of the clusters.

Derya Birant et al. [13] proposed ST-DBSCAN modification of DBSCAN with concept of density factor which is degree of density of cluster. It discovers clusters according to non-spatial, spatial and temporal values of the objects.

Keiichi Tamura et al. proposed (ε, ʈ) density based spatial clustering algorithm which extracts spatial as well as temporal clusters associated with local topics and events.

5. Analytical Survey

The classic clustering techniques performs poor for spatial data mining compared to density based clustering algorithms. Density based algorithms typically regards clusters as dense regions of objects in the data space which are separated by regions of low density (representing noise). Density-based methods can be used to filter out noise (outliers) and discover clusters of arbitrary shape.

Density-Based Spatial Clustering and Application with Noise (DBSCAN) is a clustering algorithm based on density. Clustering is done through growing high density area and it finds arbitrary shaped clusters. There are two important objects clusters and noise, for DBSCAN algorithm. All points in data set are divided into points of clusters and noise. The key idea of DBSCAN is that for each point of a cluster the neighbourhood of a given radius has to contain at least a minimum number of points, i.e. the density in the neighbourhood has to exceed some threshold. The theoretical complexity of DBSCAN in low dimensional spatial data is O(Nlog(N)).

The following table 1 shows analytical survey regarding spatio-temporal clustering.

Table 1: Analytical Survey

Sr. No	*Paper Title*	*Algorithm*	*Dataset*	*Description/Method*
1	ST-DBSCAN : An algorithm for clustering spatio-temporal data	ST-DBSCAN	Remotely sensed data(historical extent of marine areas)	Concept of density factor introduced
2	Mining spatial-temporal clusters from geo-databases	ST-GRID, ST-DBSCAN	Integrating seismic catalog in China and neighbor countries	*k*-dist graph used
3	Fuzzy cluster analysis of spatio-temporal data	FCM	Weather data	The Pearson correlation coefficient used
4	A fuzzy clustering model for multivariate spatial time series	Extended FCM	Spatiotemporal Time series	Spatial Penalty coefficient
5	P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos	P-DBSCAN	Geotagged photos	Used adaptive density for optimizing search
6	DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method	DB-SMoT	Trajectories of fishing boats	Direction based stops and moves
7	Density-based Spatiotemporal clustering algorithm for extracting bursty areas from georeferenced documents	Extended DBSCAN with (ε, ʈ)	Geotagged tweets from twitter	DBSCAN with (ε, ʈ) gives spatial as well as temporal clusters

6. Conclusion

Spatio-temporal data mining is becoming very important field of research as it focuses the data for not only static view point but also on time and space. Thus it is useful to locate future statistics based on time and space. Density based clustering algorithm DBSCAN can detect arbitrary shaped clusters, so used in spatial data mining, spatio-temporal data mining on large scale. Classification of spatio-temporal data, survey, challenges, applications also spatio-temporal data analysis is presented in this paper.

References

[1] Ester, M., Kriegel, H.-P., Sander, J. and Xu, X. (1996) "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" In Proc. Second

International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, E. Simoudis, J. Han and U. Fayyad, Eds., AAAI Press.

[2] Manso, J. A. ; Times, V. C. ; Oliveira, G. ; Alvares, Luis Otavio ; Bogorny, V. "DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method" In: IEEE International Conference on Intelligent Systems (IS), 2010, Londres. Proceedings of the IEEE International Conference on Intelligent Systems.

[3] D. L. Pham, "Spatial models for fuzzy clustering," Comput. Vis. Image Understand., vol. 84, no. 2, pp. 285–297, 2001.

[4] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatiotemporal clustering," in Data mining and Knowledge Discovery Handbook. New York: Springer, 2010, pp. 855–874.

[5] M. Kulldorff, "Prospective time periodic geographical disease surveillance using a scan statistic," J. Roy. Statist. Soc. A, vol. 164, no. 1, pp. 61–72, 2001.

[6] H. Izakian and W. Pedrycz, "A new PSO-optimized geometry of spatial and spatio-temporal scan statistics for disease outbreak detection," Swarm Evol. Comput., vol. 4, pp. 1–11, Jun. 2012.

[7] F. Di Martino and S. Sessa, "The extended fuzzy C-means algorithm for hotspots in spatio-temporal GIS," Expert Syst. Appl., vol. 38, no. 9, pp. 11829–11836, Sep. 2011.

[8] M. Wang, A. Wang, and A. Li, "Mining spatial-temporal clusters from geo-databases," in Proc. 2nd Int. Conf. Adv. Data Mining Appl., 2006, pp. 63–270.

[9] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," Data Mining Knowl. Discovery, pp. 226–231, 1996.

[10] Z. Liu and R. George, "Fuzzy cluster analysis of spatio-temporal data," in Proc. 18th Int. Symp. Comput. Inf. Sci., Antalya, Turkey, 2003, pp. 984–991.

[11] M. Deng, Q. Liu, J. Wang, and Y. Shi, "A general method of spatiotemporal clustering analysis," Sci. Chin. Inf. Sci., pp. 1–14, 2011.

[12] R. Coppi, P. D'Urso, and P. Giordani, "A fuzzy clustering model for multivariate spatial time series," J. Classif., vol. 27, no. 1, pp. 54–88, Mar. 2010.

[13] Derya Birant & Alp Kut, (2007) "ST-DBSCAN: An algorithm for clustering spatio-temporal data", Data & Knowledge Engineering, Volume 60, Issue 1, January, Pages 208-221.

[14] Venkateswara Rao, Govardhan A. and Chalapati Rao "Spatiotemporal data mining: issues, tasks and applications", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.1, February 2012

[15] John F. Roddick and Myra Spiliopoulou. A survey of temporal knowledge discovery paradigms and methods. IEEE Trans. on Knowl. and Data Eng., 14(4):750–767, ISSN 1041- 4347, July 2002.

[16] K.Venkateswara Rao, Dr. A.Govardhan & Dr.K.V.Chalapati Rao, "Discovering Spatiotemporal Topological Relationships", The second international workshop on Database Management Systems, DMS-2011, July, Chennai, India, Springer Proceedings LNCS-CCIS 198.

Author Profile

Vaishali V. Dhundale, persuing ME in Computer Engineering from PICT, Savitribai Phule Pune University, Pune.