International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 7 Issue 9, September 2020

ISSN: 2394-4404

Spatiotemporal Data Mining In Context

Sylvanus Chime

MSc. Information Technology, University Of East London, United Kingdom

***Abstract:** Temporal, spatial and spatiotemporal data provide us with quality information about climate science and climate change respectively. The mining of Spatiotemporal data has become increasingly important and it has created the popularity of mobile phones, Internet-based map services, GPS devices, digital Earth, weather services, satellite, RFID, sensor, wireless, video technologies and far-reaching implications*

Some of the examples of spatiotemporal data mining include the discovering the evolution of the history of land and cities, determining earthquakes, discovering weather patterns and global warming trends.

Moreover, the increased availability of geographical data and the recent growth in observations and model outputs indicate new opportunities for data miners. This document maps climate requirements to solutions available in temporal, spatial and spatiotemporal data mining.

It will also make a case for the development of novel algorithms to address these issues, discusses the recent literature, and proposes new directions.

A new version of this document will provide an illustrative case study to prove that relatively simple data mining approaches can provide new scientific insights with high societal impacts.

I. OVERVIEW

Spatiotemporal data is data that relates to both time and space. Spatiotemporal data mining could be defined as the process of uncovering patterns and vital knowledge from spatiotemporal data. Some of the examples of spatiotemporal data mining include the discovering the evolution of the history of land and cities, determining earthquakes, discovering weather patterns and global warming trends.

The mining of Spatiotemporal data has become increasingly important and it has created the popularity of mobile phones, Internet-based map services, GPS devices, digital Earth, weather services, satellite, RFID, sensor, wireless, video technologies and far-reaching implications *(Jiawei Han, ... Jian Pei, in Data Mining (Third Edition), 2012).*

A spatiotemporal database is used in storing, querying, retrieving information that is related to space and time.

Some examples include:

- ✓ Historic process that involve tracking of tectonic activities
- ✓ Identification of objects in motion, which categorically can take up only a single position at a point in time.
- ✓ Any database that contains information of wireless communication networks, which may be active only for a short timespan within a geographic region.
- ✓ Any index of different species in a given region, where in the space of time additional species maybe added or existing species may migrate or die out.

An extension of spatial databases and temporal databases is also known as Spatiotemporal databases and it embodies spatial, temporal, and spatiotemporal database concepts. Thus, capturing spatial and temporal aspects of data and deals with;

- ✓ Geometry of change over time
- ✓ Location of objects in motion over invariant geometry
- ✓ (*Ralf Hartmut Güting; Markus Schneider (2005). Moving Objects Databases*).

II. IMPLEMENTATIONS

For practical reasons, spatiotemporal databases are not based on the relational model, chiefly among them that the data is multi-dimensional, capturing complex structures and behaviors. Although there exist numerous relational databases with spatial extensions (Ralf Hartmut Güting; Markus

Schneider (2005). *Moving Objects Databases*. Academic Press. ISBN 978-0-12-088799-6).

Before year 2008, there was no Relational Database Management System (RDBMS) products with spatiotemporal extensions. Although they exits some products such as the open-source TerraLib, that uses a middleware approach for data storage in a relational database. Apart from specified spatial domain, there are however no official or de facto standards for spatiotemporal data models and their querying and management. Moreover, there is still an argument that the theory of this area is incomplete.

There is an alternative approach which is the constraint database system such as MLPQ (Management of Linear Programming Queries). GeoMesa is an open-source distributed spatiotemporal index built on top of Bigtable-style databases using an implementation of the Z-order_curve to create a multi-dimensional index combining space and time *(Brent Hall; Michael G. Leahy (2008). Open Source Approaches in Spatial Data Handling).*

In diverse domains, considerable volume of spatiotemporal data is frequently collected and examined. Theses include, climate science, social sciences, neuroscience, epidemiology, transportation, mobile health, and Earth sciences. It is necessary to always keep in mind that spatiotemporal data is different from relational data for which computational approaches are developed in the data mining community for multiple decades.

The presence of these attributes introduces additional challenges that needs to be dealt with. Approaches for mining spatiotemporal data have been studied for over a decade in the data mining community.

In this article, we present a broad survey of this relatively young field of spatiotemporal data mining. We discuss different types of spatiotemporal data and the relevant data mining questions that arise in the context of analyzing each of these datasets.

Considering the characteristics of the data mining challenges examined, we classify the literature on spatiotemporal data mining into six major categories: clustering, predictive learning, change detection, frequent pattern mining, anomaly detection, and relationship mining.

We discuss the various forms of spatio-temporal data mining problems in each of these categories (Gowtham Atluri, Anuj Karpatne, Vipin Kumar)

- ✓ *CLUSTERING:* refers to the grouping of instances in a data set that share similar feature values. Novel challenges arise due to the spatial and temporal aspects of different types of ST instances.
- ✓ *PREDICTIVE LEARNING:* The basic objective of predictive learning methods is to learn a mapping from the input features to the output variables using a representative training set. Both the input and output variables can belong to different types of ST data instances, thus resulting in a variety of predictive learning problem formulations.
- ✓ *FREQUENT PATTERN MINING:* is the process of discovering patterns in a data set that occur frequently over multiple instances in a data set, e.g., frequently bought groups of items in market-basket transactions.
- ✓ *ANOMALY DETECTION:* Anomalies are traditionally defined as instances that are remarkably different from the majority of instances in the data set.
- ✓ *CHANGE DETECTION:* this approach has also been explored for identifying changes in time-series. For example, variational approaches for switching state space models have been used to discover. The problem of change detection involves identifying the time point when the behavior of a system undergoes a significant deviation from its past behavior.
- ✓ *RELATIONSHIP MINING:* relationships among pairs of time series can be discovered using any of the similarities in instances.

III. SOURCES OF ST DATA & MOTIVATION FOR ANALYZING THEM IN DIFFERENT APPLICATION DOMAINS

It was mentioned earlier that in diverse domains, considerable volume of spatiotemporal data is frequently collected and examined such as social media, healthcare, agriculture, transportation, and climate science (ACM Computing Surveys, Vol. 1, No. 1, Article. Publication date: November 2017).

Within this chapter, we try to describe the different sources of spaciotemporal data and the motivation for analysing them in different application domains.

CLIMATE SCIENCE: in this application domain, one can study data pertaining to current and past atmospheric and oceanic conditions (e.g., temperature, pressure, wind-flow, and humidity) is collected and studied in climate science [Karpatne et al. 2013].

The purpose in studying this data is to discover relationships and patterns in climate science that advance our understanding of the Earth's system and help us better prepare for future adverse conditions by informing adaptation and mitigation actions in a timely manner.

NEUROSCIENCE: Continuous neural activity captured using a variety of technologies such as Functional Magnetic Resonance Imaging (fMRI), Electroencephalogram (EEG), and Magnetoencephalography (MEG) is studied in neuroscience [Atluri et al. 2016].

The purpose in studying this data is to understand the governing principles of the brain and thereby determine the disruptions to normal conditions that arise in the case of mental disorders.

ENVIRONMENTAL SCIENCE: Studying the data pertaining to the quality of air, water, and environment is one of the objectives of environmental science [Thompson et al. 2014].

Studying these environmental data sets is helpful in detecting changes in levels of pollution, identify the causal factors that contribute to pollution, and to design effective policies to reduce the different types of pollution

PRECISION AGRICULTURE: data collected and studied in precision agriculture helps in detecting plant disease and it also helps in understanding the effect of multiple factors such as compaction during planting, misapplication of fertilizer, as well as their inter-relationships [Mahlein 2016].

Knowledge gotten from this would help in laying down proper procedures in future crop cycles to mitigate the risks due to the factors that adversely affect the crop yield.

EPIDEMIOLOGY/ HEALTH CARE: Electronic health record data that is widely stored in hospitals provide demographic information pertaining to patients as well diagnosis made on patients at different time points [Matsubara et al. 2014].

This data is studied to discover spatiotemporal patterns in different diseases and to study the spread of an epidemic.

SOCIAL MEDIA: Users of social media portals such as Twitter and Facebook post their experience at a given place and time. Each social media post captures the experience of a user at a given place and time [Tang et al. 2014]. Using this data one can study collective user experience at a given place for a given time period.

IV. PROPERTIES OF ST DATA

The presence of space and time introduces a rich diversity of spaciotemporal data types and representations, which forms several means of formulating spaciotemporal database management (STDM) problems and methods.

We have already described some of the generic properties of spaciotemporal data, and then described its basic types of data available in different applications. Building on these descriptions, we mention some of the common ways of defining and representing its instances, and generic methods for computing similarity among different types of its instances.

Spatiotemporal data introduces challenges as well as opportunities for classical data mining algorithms. The two main generic properties of spatiotemporal data are

AUTO-CORRELATION: In domains involving spatiotemporal data, the observations made at nearby locations and time stamps are usually not independent. However, they are correlated with each other. This auto-correlation in spatiotemporal data sets results in a coherence of spatial observations.

HETEROGENEITY: implies that every instance belongs to the same population and is thus identically distributed. Hence, observations made in winter are differently distributed than the observations made in summer.

V. CONCLUSION & FUTURE WORK

The most used methods of data mining are created on the believe that data instances identically distributed and independent. Anyway, this believe is violated when dealing with spaciotemporal data because data instances are structurally related to one another in the context of space and time and show varying properties in different spatial regions and time periods. Any negligence on these dependencies during data analysis can lead to inaccurate and interpretability of data (ACM Computing Surveys, Vol. 1, No. 1, Article. Publication date: November2017).

A review document on issues arising from STDM and its methods attempts to builds a foundation of spaciotemporal data types and properties that can help in identifying the relevant problems and methods for any class of spaciotemporal data encountered in real-world activities.

It has also tried to present a survey of STDM approaches for several studies in connection with data mining problems such as clustering, predictive learning, frequent pattern mining, anomaly detection, change detection, and relationship mining.

This article attempted to provide a good understanding of the field of spatiotemporal data mining.

However, considering the vast literature on this topic and time limitations, I was only able to cover a small fraction of work in this fast-growing area of research.

Anyways, I hope this article provides a good foundation for consolidating the rich and diverse literature on STDM research, which has been explored in several different application contexts for different types of spaciotemporal data.

REFERENCES

[1] Data Mining (Third Edition) The Morgan Kaufmann Series in Data Management Systems 2012, Pages 543-584

[2] Data Handling in Science and Technology Volume 32, 2020, Pages 305-331Ralf Hartmut Güting; Markus Schneider (2005). Moving Objects Databases

[3] Jiawei Han, ... Jian Pei, in Data Mining (Third Edition), 2012

[4] Brent Hall; Michael G. Leahy (2008). Open Source Approaches in Spatial Data Handling

[5] M. Chen, T. Kanade, in Medical Image Recognition, Segmentation and Parsing, 2016

[6] Gowtham Atluri, Anuj Karpatne, Vipin Kumar

[7] Ralf Hartmut Güting; Markus Schneider (2005). Moving Objects Databases

[8] C.C.Aggarwal.2015. Mining Spatial Data. In Data Mining. Springer,531–555. C.C.Aggarwal.2017.

[9] SpatialOutlierDetection. InOutlier Analysis.Springer,345–368. R.Agrawaletal.1995.

[10] Discoveringclustersinmotiontime-seriesdata. In CVPR, Vol.1. IEEE, 375–381. H.Altetal.1995.

[11] International Conference on Data Mining. G.Atlurietal. 2015.

[12] ACMComputing Surveys, Vol. 1, No. 1, Article. Publicationdate: November2017

[13] GowthamAtluri, Anuj Karpatne, and Vipin Kumar. 2017. Spatio-Temporal Data Mining: A Survey of Problems and Methods. ACM Comput. Surv. 1,1(November 2017), 37 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn