International Journal of Computer Science and Network (IJCSN)
Volume 1, Issue 4, August 2012 www.ijcsn.org ISSN 2277-5420

A Survey on Spatio-Temporal Data Mining

[1]DIPIKA KALYANI, [2]Prof. Setu Kumar Chaturvedi

[1]Computer Technology and Application

Technocrats Institute of Technology,

Bhopal (M.P), India

[2]Computer Science and Engineering

Technocrats Institute of Technology,

Bhopal (M.P), India

Abstract

Data mining is the process of searching valuable information by analyzing large volumes of data through automatic or semi-automatic means to discover meaningful patterns and rules. The field of spatio-temporal data mining is concerned with such analysis in the case of spatial and temporal interdependencies. Many interesting techniques of spatio-temporal data mining are proposed and shown to be useful in many applications. Spatio-temporal data mining brings together techniques from different fields such as machine learning, statistics and databases. Here, we present an overview of spatio-temporal data mining and discuss its various tasks and techniques in detail. We have also listed a few research issues of spatio-temporal data mining.

Keywords. Spatio-Temporal Data Mining; Spatio-Temporal Data; Tasks and Technique

1. INTRODUCTION

Spatio-temporal data mining refers to the extraction of implicit knowledge, spatial and temporal relationships or other patterns from spatio-temporal data. *Data mining*, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help one to focus on the most important information in their huge data warehouses [2]. Spatio-temporal data mining involves extracting and analyzing useful information stored in large spatio-temporal databases. This new discipline today finds application in a wide and diverse range of business, scientific and engineering scenarios. For example, several terabytes of remote-sensing image data are gathered from satellites around the globe.

1.1 Spatial Concepts

The field of *spatial data mining* is where the spatial aspect of the data defines a relationship between every data point (close-to, within, north-of, etc). *Spatial data mining techniques* extract useful patterns from spatial data sets. Important attribute of SDM is location. *Spatial data* in GIS is defined as elements that can be stored in a map, images, graph and tabular forms. Spatial data mining techniques extract useful patterns from spatial datasets. *Spatial Association Rule* occurs when a predicate in either the antecedent or the consequent contains a spatial relationship.

1.2 Temporal Concepts

The field of *temporal data mining* is where the temporal aspect of the data defines a relationship between every data point (before, during, etc). *Temporal data mining techniques* extract the relationships or patterns from historical data sets by placing greater emphasis on the temporal element of data. Important attribute of TDM is Time [1]. A *Temporal Database* is real world database that maintains past, present, and future data.

1.3 Spatio-Temporal Concepts

Spatio-temporal concepts integrate both spatial and temporal concepts. It deals with both spatial and temporal relationships. Two important attributes of STDM are Location and Time. The field of *spatio-temporal data mining* is where this relationship is both defined by the spatial and temporal aspects of the data and is extremely challenging due to the increased search space for knowledge [5]. The *importance* of spatio-temporal data mining is growing with the increasing importance of large datasets such as maps, virtual globes, and repositories of remote-sensing images.

1.4 Spatio-Temporal Data

Spatio-temporal data usually records the states over time of an object, an event or a position in space. Such data can be found in several application fields, such as traffic management, environment monitoring, weather forecast, etc. ST Data is a set of spatio-temporal sequences, S. Each element of the sequence is represented by its spatial and temporal attributes ($x_1, x_2 \dots x_n$, t), where x_i, $1 \leq I \leq n$, is a spatial attribute and t a temporal attribute. Spatio-temporal data are stored in 3-d format (2-d space info + time).

Management of spatio-temporal data has gained much interest during these past few years mainly due to rapid advancements in telecommunications, which facilitates collection of large datasets. The recent advances and price reduction of technologies like Satellite Images, Sensor Networks, and GPS devices have facilitated the collection of spatio-temporal data.

Analysis of spatio-temporal data is inherently challenging. Spatio-temporal data require complex data pre-processing, transformation, data mining, and post-processing techniques to extract novel and understandable patterns [4]. Some of the new methods include those for discovering interactions, detecting spatial outliers and location prediction on spatio-temporal pattern mining.

Vast amount of ST data is obtained from various fields of which few are given below:

- Meteorology: weather data, moving storms, tornados, droughts, etc.
- Biology: animal movements, species relocation, extinction, etc.
- Crop sciences: grasshopper infestation, harvesting, soil quality changes, etc.
- Forestry: forest growth, forest fires, tree cutting, planning tree planting, etc.
- Geophysics: earthquake histories, volcanic activities, prediction, etc.

2.SPATIO-TEMPORAL DATA MODELS

In the past, research in spatial and temporal data models and database systems has mostly been done independently. *Spatial data models* have focused on modeling and querying geometries associated with objects while *temporal data models* have focused on modeling and querying temporally evolving data. *Spatio-temporal data models* are built combining temporal and spatial data models. Aim is to develop spatio-temporal models and evaluate both the accuracy and the complexity of such models. Spatio-temporal data model can describe both continuous and discrete change.

There are two ways to accommodate temporal and spatial data models;

- The embedding of a temporal awareness in spatial data models and,
- The accommodation of space into temporal data models.

There are different modeling techniques that can be explored depending on the how the data is collected.

2.1 Geographical Spatio-Temporal Data Mining

The technology of geographical spatio-temporal data mining is still in its infancy of research. In *geographical spatio-temporal data mining*, uncertainty is involved at each step, from data pre-processing through data conceptualization until association rules extraction [3]. A large number of techniques, such as geographical spatio-temporal transaction, data conceptualization, and storage methods assigned with spatio-temporal semantics, remain unsolved so far. Discovery of geographical association rules is computation of multivariate spatio-temporal correlations or multivariate spatio-temporal variability. Different geographical association rules are mined out of spatio-temporal transactions. The difference between general data mining and geographical data mining is the computation of geographical spatio-temporal correlations. These correlations are evaluated with the methods of geo-statistics interpolation, wavelet data decomposition, fuzzy c-means clustering, etc.

2.2 Applications of Spatio-Temporal Data Mining

Spatio-temporal data mining has its application in many fields. Few real-world applications include:

- Public Health (e.g. spread of disease)
- Public Safety (e.g. crime hot spots)
- Mobile-commerce industry
- Local instability in traffic
- Migration of birds
- Autonomous navigation
- Fleet tracking
- Fishing control
- Pedestrian behavior analysis

3. MODELS AND PATTERNS

Outputs of data mining algorithms can be categorized in the structures of these algorithms which are classified as models and patterns. These structures may be used to achieve data mining objectives. A *model* is a global, high-level and abstract representation for the data. Recently proposed space-time data models integrate time and space as the primary dimensions of data while other attributes are subordinate. Models can be classified as predictive or descriptive. *Predictive models* are used in forecast and classification applications while *descriptive models* are useful for data summarization. For example, spatio-temporal traffic models are useful for traffic incident detection. On the other hand, clustering is a good example of descriptive modeling techniques.

A *Pattern* is a local structure that makes a specific statement about a few variables or data points. "*Pattern mining*" is a data mining method that involves finding existing patterns in data. Spatio-temporal movement patterns are useful in domains such as traffic management, traffic flows and animal tracking. The traffic incident detection problem can be viewed as recognizing incident patterns from observed data. Recently, spatio-temporal pattern mining has attracted many research efforts [6]. Since data collected from detectors have temporal and spatial components that must be taken into consideration in the mining process. Pattern mining can be used as a tool to identify terrorist activity. Pattern mining looks for patterns that might be associated with terrorist activity — these patterns might be regarded as small signals in a large ocean of noise. Mining spatio-temporal movement patterns is receiving increasing interest from the data mining community.

While this distinction between models and patterns is useful from the point of view of categorizing data mining algorithms, there are cases when such a distinction becomes blurred.

4. SPATIO-TEMPORAL DATA MINING TASKS

Both the "spatial" and the "temporal" prefixes have added substantial complexity to data mining tasks. Nevertheless, to investigate both "spatial" and "temporal" relations at the same time complicates the data mining tasks even more. Spatio-temporal data mining tasks can be classified as: (i) Segmentation, (ii) Dependency analysis, (iii) Deviation and outlier analysis, (iv) Trend Discovery and (v) Generalization and characterization.

In this section, we provide a brief overview of spatio-temporal data mining techniques as relevant to above specified tasks.

4.1 Segmentation

Spatio-temporal Data Mining Task	Techniques	
	Static Spatial Data	Spatio-Temporal Data
Segmentation	Cluster analysis, Bayesian classification, Decision tree	Temporal extension to clustering and classification

4.1.1 Clustering

Clustering can be used to identify locations with similar incidents of ecosystem disturbance. Clustering is the task of partitioning data into groups of similar objects. Clustering spatio-temporal data can also help in social network analysis, which is used in tasks like targeted advertising and personalization of contents. *Spatio-Temporal clustering* is equivalent to detecting and tracking moving objects.

Cluster analysis provides the capability to investigate the spatio-temporal variation of data. In cluster analysis the optimal number of clusters could vary with the temporal scale of input data. Cluster analysis of spatio-temporal data, so-called *regionalization*, is to analyze the spatial variability of one or more physical variables and to decompose a large complex area into smaller homogeneous regions.

Clustering can be used for data reduction purposes. Two potential *applications* of clustering: (1) to assist users in categorizing different types of ecosystem disturbance events and (2) to facilitate real-time exploration and analysis of high-resolution eco-climatic data.

4.1.2 Classification

In classification, each object presented to the system is assumed to belong to one of finitely many classes and the goal is to automatically determine the corresponding category for the given input sequence. There are many examples of sequence classification applications, like speech recognition, gesture recognition, handwritten word recognition, etc.

In order to build models for classification that fully exploits the spatio-temporal nature of these data it has led to the investigation of *Bayesian Networks* C*lassifiers*. Bayesian networks are probabilistic models that facilitate the discovery of complex relationships in spatio-temporal datasets. BN are transparent in the way that they model spatio-temporal data. A BN consists of a *directed acyclic graph* consisting of links between nodes that represent variables in the domain. The links are directed from a *parent* node to a *child* node, and with each node there is an associated set of *conditional probability distributions*.

The *Dynamic Bayesian Network* is an extension of the BN that can model time series. Links in a DBN can be between nodes in the same time slice or from nodes in previous time slices. A *Spatial Bayesian Network* to be a BN that represents data of a spatial nature and a Spatial Dynamic Bayesian Network to be a BN that represents spatio-temporal data. *Temporal Bayesian network*, also known as dynamic bayesian Network is a BN where the N nodes represent variables at differing time slices. A S*patio-temporal Bayesian network* is a special TBN that assumes dependencies between variables based on some spatial neighborhood. A set of operators are required for learning network structures that exploit the spatial nature of the dataset. An extension of these models which is designed to classify data is called Spatio-Temporal Bayesian Network Classifier.

4.2 Dependency Analysis

Spatio-Temporal Data Mining Task	Techniques	
	Static Spatial Data	Spatio-Temporal Data
Dependency Analysis	Association Rules	Prediction And Temporal extension to Association Rules

Mining for spatial dependency involves finding patterns in the form of rules to predict the value of some attribute based on the value of other attributes, taking into account that the values of attributes of nearby spatial objects tend to systematically affect each other. On the other hand, mining for temporal dependency involves finding meaningful time-related rules such as the valid time periods during which association rules hold. Traditional analysis tools are inadequate for handling the complexity of mining spatio-temporal patterns.

4.2.1 Prediction

Prediction deals with forecasting of future values of some attribute based on the value of other attributes over time. Spatio–temporal data are associated with time and space. To extract knowledge from spatio-temporal data one needs to build a predictive model. For example investigations on *earthquake predictions* are based on the assumption that all of the regional factors can be filtered out and general information about the earthquake patterns can be extracted. The prediction of the earthquakes is a very difficult and challenging task; we cannot operate on only one level of resolution. Various computational methods and tools are used for Earthquake prediction.

4.2.2 Association Rules

An A*ssociation Rule* takes the form *A→B* where *A* and *B* are sets of predicates. There are many applications of association rule mining in spatio-temporal domain. A *spatio-temporal association rule* occurs when there is a spatio-temporal relationship in the antecedent or consequent of the rule.

Association rule mining seeks to discover associations among transactions encoded within a database. Association rule mining uses the concepts of support and confidence to identify interesting rules. The *support* is the probability of a record in the database satisfying the set of predicates contained in both the antecedent and consequent. The *confidence* is the probability that a record that contains the antecedent also contains the consequent. Most attempts to apply Association Rule Mining technique to spatial-temporal domains maps the data to transactions, thus losing the spatio-temporal characteristics. *Spatio-temporal association rules (STARs)* describe how objects move between regions over time [13].

Multiple level association rule mining is supported by mining rules at varying levels of the concept hierarchy to find the hierarchy resolution that best captures the rule. The development of concept hierarchies through data classification demonstrates a methodology to support multiple levels spatio-temporal association rule mining. Association rule mining is a promising analytical tool for spatio-temporal data analysis. There has been work on spatial association rules and temporal association rules but very little work has addressed both spatial and temporal dimensions.

4.3 Deviation and Outlier Analysis

Spatio-Temporal Data Mining Task	Techniques	
	Static Spatial Data	Spatio-Temporal Data
Deviation and Outlier Analysis	Outlier Detection	Temporal extension to Outlier Detection

Outliers can be defined as observations which appear to be inconsistent with the remainder of the dataset. *Outlier detection* is a data mining technique like classification, clustering, and association rules. A *Spatial Outlier* is an object whose non-spatial attribute value is significantly different from the values of its spatial neighbors. A *Temporal Outlier* is an object whose non-spatial attribute value is significantly different from those of other objects in its temporal neighborhood. *Spatio-Temporal Outlier* combines S-Outlier and T-Outlier definitions. A STO is a spatio-temporal object represented by a set of instances (oid, si, ti), where the spacestamp si, is the location of object oid at timestamp ti. The thematic attributes of STO are significantly different from those of other objects in its spatial and temporal neighborhoods.

An *outlier detection algorithm* first identifies S-Outliers and then T-Outliers. However, the identification of first T-Outliers and then S-Outliers yields the same result. So ST-Outliers and TS-Outliers are identical. Many approaches are proposed to identify the spatio-temporal outliers.

For *spatio-temporal outlier detection* there exists a three-step approach to detect spatio-temporal outliers in large databases. These steps are clustering, checking spatial neighbors, and checking temporal neighbors. Clustering is a basic method for outlier detection. It checks the spatial and temporal neighbors of the potential STOs identified in the clusters. If the semantic value of such an STO does not have significant differences with its spatial neighbor, it may not be a STO; then check its temporal neighbors. If the difference with the temporal neighbors is not large, this checking is not a STO. Otherwise, it is confirmed as a STO.

4.4 Trend Discovery

Spatio-Temporal Data Mining Task	Techniques	
	Static Spatial Data	Spatio-Temporal Data
Trend Discovery	Discovery of common trends and Regression	Sequence Mining

Sequence mining is mining frequent sequences satisfying a given regular expression. A *sequence* is an ordered list of discrete items, such as a sequence of letters or a gene sequence.

There are many application domains where data are represented as sequences. In the *medical domain*, symptoms exhibited by a patient can be ordered according to their occurrence in time, and some patterns can be found that relate a certain subsequence of symptoms with a particular disease. Also, genetic analysis must take into account the sequential nature of DNA. In the *financial domain*, the daily price of a stock during can be naturally represented as a sequence of values. In the *market analysis domain*, finding patterns representing the buying behavior of the person help in predicting future sales of products. In the *WEB domain*, finding patterns in the sequence of web pages that person visits, helps in predicting which pages the person will visit next.

Extracting frequent subsequences from a database of sequences is an important data mining task. Frequent sequence mining approaches are often based on the use of an Apriori-like candidate generation strategy, which typically requires numerous scans of a potentially huge sequence database. A more efficient strategy for discovering frequent patterns in sequence databases requires only two scans of the database. The first scan obtains support counts for subsequences of length two. The second scan extracts potentially frequent sequences of any length and represents them as a compressed frequent sequences tree structure. Frequent sequence patterns are then mined from the FS-tree. Incremental and interactive mining functionalities are also facilitated by the FS-tree.

4.5 Generalization and Characterization

Spatio-Temporal Data Mining Task	Techniques	
	Static Spatial Data	Spatio-Temporal Data
Generalization and Characterization	Attribute Oriented Induction	Temporal extension to Attribute Oriented Induction

A*ttribute-Oriented Induction* method is based on generalization hierarchy and summarizing the general relationships between attributes at higher concept levels. A generalization hierarchy can explicitly be specified by a domain expert. Several authors have investigated attribute-oriented induction methods for extracting generalization hierarchies for spatio-temporal data [7].

Common data mining tasks include deriving the general characteristics of data, classifying them into different groups, extracting association rules between attributes or objects, finding relationships between individual data items, and detecting trends and deviations. Attribute oriented induction technique is use to mine characteristic and discriminant rules from given spatio-temporal data. The attribute oriented induction technique is a form of generalization utilizing conceptual hierarchies as background knowledge in the discovery process. The induction technique provides a general characterization of a selected group of items based on the commonness of their attribute values.

5. RESEARCH ISSUES FOR STDM

Presently, several open issues can be identified in this research field ranging from the definition of suitable mining techniques able to deal with spatio-temporal information to the development of effective methods to analyze the produced results. Few of the research issues are discussed below.

1. To reveal spatial and temporal relationships among spatial entities at various scales, as scale effect in space and time is a challenging research issue in geographic analysis.
2. To develop spatio-temporal models and evaluate both the accuracy and the complexity of such models.
3. To modify the data mining techniques so that they can identify efficiently the spatial and temporal features embedded in the datasets of a given application domain.
4. To develop mechanisms to test and validate spatio-temporal data mining results, particularly that test the validity of spatial and temporal relations, and to reconcile discrepancies in data.
5. To develop efficient and general methods that can support complex spatio-temporal data types structures as well as the scalability issue as the amount of currently collected data increases at exponential rates.
6. Exploration of efficient methods due to the large amount of spatio-temporal data and the complexity of spatio-temporal data types, data representation, and spatial data structure.
7. Distributed data mining has become necessary for large and multi-scenario datasets requiring resources, which are heterogeneous and distributed. This constitutes an additional research aspect of spatio-temporal data mining.

8. To find the impact of using different classification schemes on the results of spatio-temporal association rule mining.
9. To investigate how more sophisticated interestingness measures and meta-mining approaches may be used to improve the utility and efficiency of applying association rule mining to spatio-temporal data.
10. To integrate data from different data sources at different levels that extends from spatio-temporal association rule mining to many types of spatio-temporal statistical analyses.

6. CONCLUSION

Spatio-temporal data mining is a promising research area dedicated to the development and application of computational techniques for the analysis of spatio-temporal data. In this paper, we have provided an overview of spatio-temporal data models. We have discussed many tasks and techniques of spatio-temporal data mining in detail. Due to the increasing computerization in many fields, these days vast amounts of data are routinely collected. Also in all data mining applications, the primary constraint is the large volume of data. New methods are needed to analyze spatio-temporal data to extract interesting and useful patterns. The field of spatio-temporal data mining is relatively young. Research accommodating both spatial and temporal data mining is sparse and wide. So we have listed few issues on which research can be done in future. Since many research challenges exist so scope of exploration in this field is quite vast. Hence there is always a need for efficient algorithms.

REFERENCES

[1] Srivastan Laxman and P S Sastry, "A survey of temporal data mining", *Sadhana Vol. 31, Part 2, pp. 173–198, April 2006.*

[2] Sugam Sharma, and Shashi Gadia, "Perl Status Reporter (SRr) on Spatiotemporal Data Mining", *International Journal of Computer Science & Engineering Survey (IJCSES) Vol.1, No.1, August 2010.*

[3] Hong Shua, Xinyan Zhub, Shangping Daic, "Mining Association Rules In Geographical Spatio-Temporal Data", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2. Beijing 2008.*

[4] Xiaobai Yao, "Research Issues in Spatio-temporal Data Mining", *A white paper submitted to the University Consortium for Geographic Information Science (UCGIS) workshop on Geospatial Visualization and Knowledge Discovery, Lansdowne, Virginia, Nov. 18-20, 2003.*

[5] John F. Roddick, Myra Spiliopoulou, "A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research", *SIGKDD Explorations. Volume 1, Issue 1, ACM SIGKDD, June 1999.*

[6] Ying Jin, Jing Dai, Chang-Tien Lu, "Spatial-Temporal Data Mining in Traffic Incident Detection", *Department of Computer Science, Virginia Polytechnic Institute and State University.*

[7] *Monica Wachowicz,* "The Role of Geographic Visualisation and Knowledge Discovery in Spatio-Temporal Data Modelling", *Publications in Geodesy 47,* pp.13-26.