A Knowledge Discovery Framework for Spatiotemporal Data Mining

Jun-Wook Lee*, and Yong-Joon Lee*

Abstract: With the explosive increase in the generation and utilization of spatiotemporal data sets, many research efforts have been focused on the efficient handling of the large volume of spatiotemporal sets. With the remarkable growth of ubiquitous computing technology, mining from the huge volume of spatiotemporal data sets is regarded as a core technology which can provide real world applications with intelligence. In this paper, we propose a 3-tier knowledge discovery framework for spatiotemporal data mining. This framework provides a foundation model not only to define the problem of spatiotemporal knowledge discovery but also to represent new knowledge and its relationships. Using the proposed knowledge discovery framework, we can easily formalize spatiotemporal data mining problems. The representation model is very useful in modeling the basic elements and the relationships between the objects in spatiotemporal data sets, information and knowledge.

Keywords: spatiotemporal data mining, spatiotemporal knowledge discovery, spatiotemporal moving pattern, discovery framework

1. Introduction

In recent times, various technologies for ubiquitous computing and services have been developed, with wireless sensor networks, context-awareness and open service architecture now regarded as the core elements of these services. The ubiquitous service infrastructure with which we can access desired information and services at any time in any place is now being realized.

Most of the services have a strong relationship with objects, events and changes occurring in the real world. As such, the services necessarily require spatiotemporal information and knowledge. With the explosive increase in the generation and use of spatiotemporal data sets, numerous research efforts have been focused on the efficient handling of the large volume of these spatio-temporal sets. However, because of the complexity and enormity of spatiotemporal information, there are many difficulties in extracting as well as handling knowledge. To provide more intelligent services, the real-time, high performance and continuous extraction of knowledge from large-scale spatiotemporal data sets is required.

Spatiotemporal data mining [1, 9] raises many challenging research issues with regard to extracting useful spatiotemporal knowledge from large spatiotemporal data sets. Recently, the non-stop collection of continuously observed data from the real world, real-time support of the ubiquitous applications, and the provision of intelligence and context-awareness make the mining problem much more complex. For example, ubiquitous services such as telematics services, location-based services, surveillance applications, and real-time environment monitoring applications all require the handling of a large volume of spatiotemporal data sets but also the supporting of useful knowledge for real time decision making.

Based on the existing temporal data mining techniques [8] or spatial data mining techniques [4], we may discover spatial or temporal knowledge from the spatiotemporal data set. However, new kinds of spatiotemporal knowledge for the services should be extracted by considering both the temporal element and spatial element of objects [1, 6, 7].

A new framework for this kind of discovery must be supported. First, any new framework for spatiotemporal data mining should define the discovery problem more formally. In this paper, we propose a new foundation model for knowledge discovery after considering the many spatiotemporal models previously proposed. Second, the developing new mining technique should be facilitated based on the available and useful spatiotemporal operations and topological relationships.

Given the remarkable growth of ubiquitous computing technology, mining from the huge volume of spatio-temporal data sets should become a core technology which can provide real world applications with intelligence.

The outline of this paper is as follows. In section 2, we review a number of previous research efforts related to spatiotemporal knowledge discovery and address some of the problems. In section 3, we propose a 3-tier spatio-temporal knowledge discovery framework, and in section 4, we describe the spatiotemporal knowledge model as a foundation model of the framework. In section 5, we describe the implementation model, which includes a

definition of the problem and the architecture of the spatiotemporal knowledge discovery system. In section 6, we show an example of a real mining system, MPMiner, being implemented, and discuss the proposed framework. Finally, we present our conclusion and suggest some further areas of study.

2. Related Works

Until now, many researchers have tried to find efficient mining methods from the large volume of real world data such as historical data, business data and geographical data. As a result, there has been much remarkable growth in terms of the theories and techniques concerning data mining. Temporal data mining techniques extract the relationships or patterns from historical data sets by placing greater emphasis on the temporal element of data; spatial data mining techniques extract useful patterns from spatial data sets. Furthermore, many spatiotemporal data mining techniques have been developed, and may be categorized into three types of approach. The first category of techniques is based on the temporal extension of the existing spatial data mining methods, while the second is based on the spatial extension of existing temporal data mining methods. For example, the spatiotemporal association rule [1] and spatiotemporal clustering [12, 13] belong to these two approaches. A new type of spatiotemporal data mining techniques includes the spatiotemporal evolution rule [1], the spatiotemporal pattern, and the spatiotemporal moving pattern [3].

Most existing spatiotemporal data mining methods are based on the raster-based spatial model, which is extended with the temporal concept [9, 11, 14]. However, this approach suffers from limited modeling and an insufficient understanding of the spatiotemporal phenomenon. To overcome the insufficient modeling power, a new knowledge framework with a spatiotemporal model needs to be designed.

Mennis and Peuquet proposed a Pyramid model [5]. This conceptual knowledge model consists of two components: a data component and a knowledge component. In this model, knowledge is considered as two distinct, yet interrelated components. However, the model is insufficient for use as a base model in implementing a spatiotemporal knowledge discovery system. Abraham [1] also suggested a new knowledge framework for the discovery of a spatiotemporal meta rule. This framework includes a new definition of both the spatiotemporal meta rule and the discovery process. However, it is not easy to apply the knowledge model and process model to other kinds of discovery problem, such as the spatiotemporal moving pattern.

The previous temporal data mining framework of Lee [2] and the framework for spatiotemporal meta rule discovery of Abraham [1] are also short of a foundation model support. The existing frameworks do not satisfy the new requirements of spatiotemporal data mining.

The requirements for a spatiotemporal knowledge discovery framework are as follows. Firstly, the framework should consider the foundation model. Secondly, it should be able to define a new type of spatiotemporal knowledge. Moreover, the spatiotemporal data mining problem must be defined formally. Finally, it should define and support the knowledge discovery process.

The proposed framework consists of a foundation model as well as an implementation model. The foundation model is a 3-tier model for representing spatiotemporal data sets and their operation. This spatiotemporal knowledge discovery framework exhibits more useful features when compared with the previous frameworks. This framework can be used in the existing knowledge discovery problem such as moving pattern mining [3]. Furthermore, it can be utilized in the development of new types of knowledge.

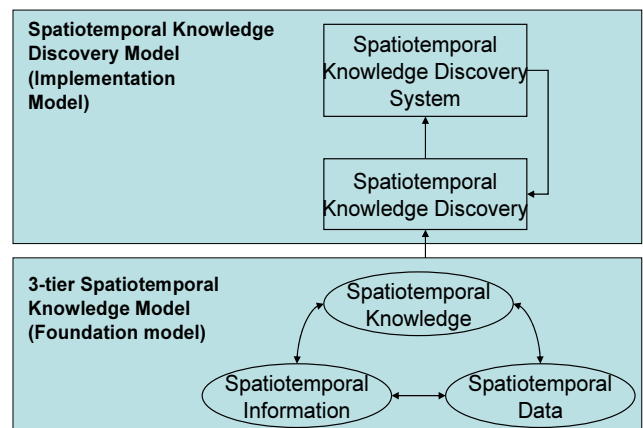## 3. A spatiotemporal knowledge discovery framework

Fig. 1 Knowledge Discovery Framework

A framework is referred to as the set of concepts, models and methodology necessary for knowledge discovery. As shown in Fig. 1, our framework consists of two models - a foundation model and an implementation model.

Foundation model here refers to a 3-tier spatiotemporal knowledge model, which is intended to represent both a set of spatiotemporal objects and a volume of knowledge. The conceptual spatiotemporal knowledge model includes a spatiotemporal data layer, a spatiotemporal information layer and a spatiotemporal knowledge layer. This model supports the representation of data and information, algebraic operation on spatiotemporal objects, and the representation of the extracted knowledge.

As an implementation model, the spatiotemporal knowledge discovery model includes the definition of spatiotemporal knowledge and knowledge evolution, and the spatiotemporal knowledge discovery process. In order to verify our knowledge discovery model, we propose an architecture for the spatiotemporal knowledge discovery system. The proposed system architecture consists of five interrelated layers: a data layer, an information layer, a

spatiotemporal mining layer, a knowledge layer and an application layer. A more detailed description will be given in section 5.

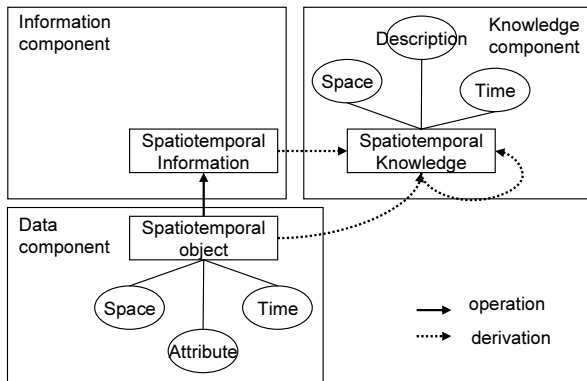## 4. A conceptual 3-tier spatiotemporal knowledge model

Fig. 2. Extended 3-tier model of Spatiotemporal Knowledge

First of all, in order to extract meaningful knowledge over the spatiotemporal phenomenon, the required properties of the foundation model are as follows. Firstly, it must be able to properly describe the identity and existence of spatiotemporal phenomena. Secondly, it must be able to describe the representation of phenomena. Therefore, it must not only include both the static and dynamic elements of spatiotemporal objects, but must also consider the temporal element, spatial element, and attributes as the three main elements of an object. Thirdly, it must be able to represent the multilevel background knowledge for analyzing phenomena. Finally, it must be able to mine new kinds of knowledge using a prior knowledge.

Among the existing knowledge models, we chose Mannis's Pyramid model [5] as the most suitable spatiotemporal knowledge model and enhanced it so as to satisfy our own requirements

The proposed conceptual knowledge model for representing spatiotemporal phenomena actually represents the observed spatiotemporal data, the processed spatiotemporal information, and the discovered spatiotemporal knowledge. As shown in Fig. 2, the data component in this model is composed of representations of the spatiotemporal data that exists in the spatiotemporal universe. The information component concerns the selected and meaningful data derived from the data component by using a generalization or transformation operation in order to extract the meaning. Lastly, the knowledge component represents knowledge objects that are useful to human beings. Knowledge can be directly induced from the set of spatiotemporal objects or indirectly induced from the spatiotemporal information or *a priori* knowledge.

Based on the concept of a value chain and concep-tualization, our conceptual knowledge model for representing spatiotemporal phenomena represents the observed spatio-

temporal data, the processed spatiotemporal information, and the discovered spatiotemporal knowledge distinctly. Compared with Manni's two-tier model of data component and knowledge, by including an information component our model is able to abstract the various operations performed on the spatiotemporal data set. The abstraction of spatiotemporal operations makes it easy to implement the knowledge discovery system in the existing legacy DBMS systems such as relational DBMS and OR DBMS.

5. Spatiotemporal knowledge discovery model

5.1 Definition of Spatiotemporal Knowledge

In order to represent properly the spatiotemporal knowledge described above, a method of representing the temporal qualification on the temporal domain must be given. Lee [2] has given a more complete definition on the time expression of temporal knowledge, including the time expression defined by Chen. For each time expression set, the calendric interval expression set is denoted as CIS, the calendric relation expression set as CRS, the time interval expression set as TIS, and the periodic time expression set as PTS. Let temporal qualification TQ $\in$ TimeExp, where TimeExp={CIS,CIS,TIS,PTS}. Moreover, the way of representing the spatial qualification on the spatial domain must be given.

A spatial qualification SQ of knowledge corresponds to ES in the spatial domain, where it refers to valid coverage of knowledge in the spatial domain. A spatiotemporal knowledge primitive RP is an indivisible unit of knowledge having a temporal qualification, as well as a spatial qualification along with a rule description. Since the spatiotemporal knowledge primitive can be extended from the existing rule descriptions by including the temporal and spatial qualification, knowledge can be defined as follows. Let spatiotemporal knowledge primitive RP be <Rule, TQ , SQ >. Now, existing types of knowledge such as descriptive knowledge, temporal knowledge and spatial knowledge can be redefined by this definition.

Spatiotemporal knowledge R is defined as a compounding of the spatiotemporal knowledge primitives and the unique identifier Rid, as follows. R =< Rid, ri + >, where ri $\in$ RP.

A spatiotemporal knowledge set RS is defined as a set of knowledge discovered, i.e. {Ri | Ri $\in$ R}. A spatiotemporal knowledge snapshot set RST is defined as a set of discovered knowledge at a time point t in the time line T. Whenever a discovery occurs, a new spatiotemporal knowledge snapshot set is generated.

5.2 Spatiotemporal knowledge discovery

Spatiotemporal knowledge discovery consists of discovering a spatiotemporal knowledge set from the given spatiotemporal data set. A spatiotemporal knowledge

discovery is defined as the problem of extracting the spatiotemporal knowledge set RS from all knowledge R=< Rid, ri+> which satisfies the given thresholds from the given spatiotemporal data set DS, using the background knowledge provided.

The proposed knowledge discovery process model is composed of two flows. That is, the development of a mining technique and the spatiotemporal knowledge discovery. The main development processes of knowledge discovery techniques include the setting up of the discovery goal, data discovery, a definition of the discovery profile, and the development of a mining operation. The processes of knowledge discovery consist of 5 steps - task formation, relevant data set formation, spatiotemporal data mining, visualization, and interprettation and evaluation.

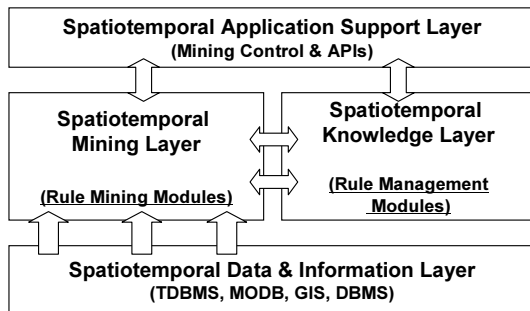### 5.3 System Architecture

Fig. 3. Knowledge discovery system architecture based on the framework

In Fig. 3, the major components and their relationships within our implementation model are depicted. The spatiotemporal data and information layer support the access to the heterogeneous spatiotemporal data set. In this layer, the access request from the higher layer is mapped into the physical access of the each storage. The spatiotemporal mining layer performs the data extraction and transformation, and then discovers rules using spatiotemporal mining operators. Therefore, access to the lower layer must be supported in this layer. The knowledge layer is associated with both the application layer and the mining layer. In this layer, such novel features as the representation of rules, the accessing and maintaining of the valid rules, and the transformation of the rules to the applications are included.

6. Implementation and discussion

6.1 Implementation of a Spatiotemporal knowledge discovery system

In order to show how the proposed framework can be applied to the development of a spatiotemporal knowledge discovery system, we briefly describe the spatiotemporal

moving pattern discovery system, i.e. MPMiner, which is implemented as a semi-tightly coupled structure based on the proposed framework.

As described above, the five layers in the implementation model of the framework correspond to the three layered conceptual spatiotemporal knowledge model of the framework. The system consists of these major modules. As seen in Fig. 4, MPMiner is a moving pattern discovery system that includes the operation MPMine() as the major mining operation and is implemented as a semi-tightly-coupled structure. This structure has the merits of efficient data processing by using a procedure extension of DBMS and the implementation of the information and knowledge component of the proposed framework.
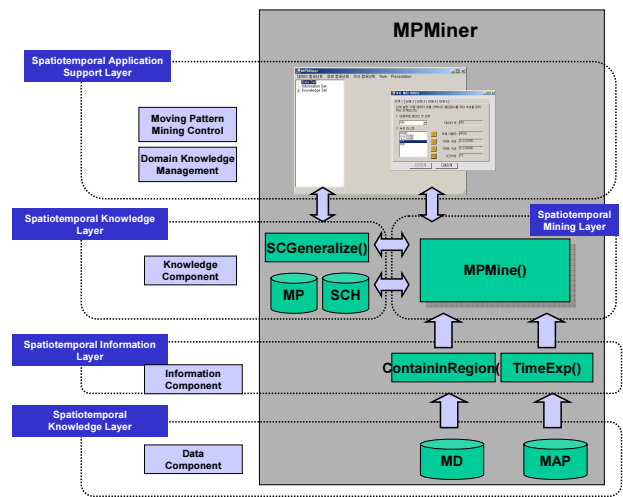

Fig. 4. Implementation of MPMiner

The implementation of this system is divided into the implementation of the framework and the implementation of the mining operations and UI. The implementation of the framework corresponds to the implementation of the three components in the conceptual model. The data component is implemented as the set of relations using Oracle 8i. In Fig. 4, it includes the moving object data set MD and the background map data MAP. As the information component, the major relational algebra operations are expressed as SQL. The spatial and temporal operations are expressed as the stored procedure using the PL/SQL. It includes the TimeExp() function for the processing of time expression and the function SCGeneralize(). In order to represent the discovered knowledge and background knowledge, the knowledge component is implemented as a set of relations in Oracle 8i. HG-Trees for representing a spatial conceptual hierarchy is also implemented as a set of relations by normalization. Also, the discovered spatiotemporal moving patterns are stored as a relation.

6.2 Discussion

As the result of the evaluation, by providing a foundation model which is based on the event-based model and which

represents hierarchically the three related components such as data, information, and knowledge, the proposed framework provides the proper representation of the spatiotemporal phenomena for knowledge discovery. However, some problems still arise where data constraints and integrity are not considered. Regarding background knowledge, the framework supports types of multiple knowledge such as temporal knowledge, spatial knowledge and spatiotemporal knowledge, as well as user defined knowledge. However, the method of representing knowledge and the kinds of operations for knowledge are yet to be included. Concerning the support of knowledge discovery integrated with the knowledge minded already, it supports integrated knowledge discovery. However, the maintenance of knowledge and knowledge evolution has not been sufficiently considered. Finally, regarding implementation in the legacy system, even operation abstraction is possible, although it is limited in some legacy systems. In the future, the framework needs to be extended by considering the problems that have been identified.

7. Conclusions

The advent of ubiquitous computing, the growth of wireless sensor network technology, and the increasing need for context-aware services and intellectualization have led to raised awareness about and the study of knowledge discovery from the huge volume of spatiotemporal data. Most of the recent services seek to provide more personalized and intelligent functions to users. To provide more intelligent services, real-time, high performance and continuous extraction of knowledge from the huge amount of spatiotemporal data sets is required.

In this paper, we discuss a framework for spatiotemporal knowledge discovery that supports the development of new kinds of knowledge such as the spatiotemporal moving pattern. In the proposed framework, it is possible to represent the definition and relationships of spatiotemporal data sets and knowledge by using a foundation model for knowledge discovery. Moreover, the definitions of and model for spatiotemporal knowledge and knowledge discovery process are clearly presented. Finally, we evaluate the characteristics of the proposed framework and present some of the related problems.

The spatiotemporal knowledge discovery framework presented here contains certain merits. However, given the features of the spatiotemporal data sets of recent applications, the size of the data sets are increasing exponentially because of the continuously changing spatiotemporal objects, the relationships of the objects are becoming more complicated because of the convergence of services, the need for real-time analysis has increased, and the semantics of extracted knowledge are continuously evolving.

In recent times, among the various branches of spatiotemporal data mining, a study on stream data mining identify many of the problems and issues to be overcome.

Therefore, in order for it to be applied to the development of new types of knowledge, the proposed framework needs to be refined and enhanced more elaborately.

References

[1] T. Abraham, Knowledge Discovery in SpatioTemporal Databases, School of Computer and Information Science, University of South Australia, Ph. D dissertation,1999.

[2] Lee Y.J., Data Mining Technique for Discovering Temporal Relation Rules, Department of computer science, chungbuk national university of korea, Ph. D dissertation, 2001.

[3] Jeong J.D., Paek O.H., Lee J.W., and and Ryu K.H., "Temporal Pattern Mining of Moving Object for Location-Based Service", In Proc. of International Conference on Database and Expert Systems Applications (Dexa2002), (LNCS2453), 2002.

[4] K. Koperski, J. Han, and J. Adhikary, "Mining knowledge in geographical data", to appear in Communications of the ACM, 1998.

[5] J. Mennis, and D.J. Peuquet, "A Conceptual Framework for Incorporating Cognitive Principles into Geographical Database presentation", International Journal of Geographical Information Science,Vol. 14, No. 6, pp. 501-520, 2000.

[6] Lee J.W., Spatiotemporal Moving Pattern Discovery Technique based on Knowledge Discovery Framework, Department of computer science, chungbuk national university of korea, Ph. D dissertation, 2003.

[7] J.F. Roddick and M. Spiliopoulou, "Temporal data mining: survey and issues", Research Report ACRC-99-007, University of South Australia, 1999.

[8] S.A. Sarabjot, D.A. Bell, and J.G. Hughes, "The role of domain knowledge in data mining", In Proc. of the Int. Conf. on Information and Knowledge Management, pp. 37-43,1995.

[9] Tsoukatos and D. Gunopulos, "Efficient Mining of Spatiotemporal Patterns", In Proc. of the 7th Int. Symp. on Spatial and Temporal Databases (SSTD), 2001.

[10] E. Mesrobian, R.R. Muntz, J.R. Santos, E.C. Shek, C.R. Mechoso, J.D. Farrara, and P. Stolorz, "Extracting Spatio-Temporal Patterns from Geoscience Datasets", IEEE Workshop on Visualization and Machine Vision, Seattle, WA, June, 1994.

[11] E. Mesrobian, R.R. Muntz, E.C. Shek, J.R. Santos, J. Yi, K. Ng, S.Y. Chien, C.R. Mechoso, J.D. Farrara, P. Stolorz, and H. Nakamura, "Exploratory Data Mining and Analysis Using Conquest", IEEE Pacific Rim Conference on Communications, Computers, Visualization, and Signal Processing, May, 1995.

[12] R.T. Ng and J. Han, "Efficient and Effective Clustering Method for Spatial Data Mining", In Proc. of International Conference of Very Large Data

Bases, pp. 144-155, 1994.

[13] R.T. Ng, "Spatial Data Mining: Discovering Knowledge of Clusters from Maps", In Proc. of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1996.

[14] R.E. Valdes-Perez, "Systematic Detection of Subtle Spatio-Temporal Patterns in Time-Lapse Imaging. I. Mitosis", Bioimaging. Vol. 4 , No. 4, pp. 232-242, 1998.

[15] J.F. Roddick and B.G. Lees, "Paradigms for Spatial and Spatio-Temporal Data Mining", Geographic Data Mining and Knowledge Discovery. Taylor and Francis. Research Monographs in Geographic Information Systems. Miller, H. and Han, J., Eds, 2001.

Jun-Wook Lee
He received a Ph.D. degree in Computer Science from Chungbuk National Univ. in 2003. During 2003~2005, he undertook a post doctrol course as a member of spatial information technology center in Electronics and Telecommunications Research Institute. He has been a Senior Researcher in ETRI to develop telematics system and USN middleware since 2005. His research interests include Spatiotemporal Data Mining, LBS, Telematics, Wireless Sensor Network, Sensor Data Mining.

Yong-Joon Lee
He received a Ph.D. degree in Computer Science from Chungbuk National Univ. in 2001. He has been a project leader in Electronics and Telecommunications Research Institute since 2002 and developed e-logistics system and RFID-middleware system. Now, he researches a new middleware system for USN applications. His research interests are in the area of Sensor Data Management, Stream Data Mining, Information Security & Privacy, Workflow.