

A Survey: On Spatial Data Mining

Asmita Bist¹, Mainaz Faridi²

M.Tech Scholar¹, Assistant Professor²

Department Of Computer Science, AIM & ACT, Banasthali University, Rajasthan, India

Abstract The extraction of spatial patterns and characteristics, spatial and non spatial data relationships, and other data features hidden in the spatial database is spatial data mining. The necessity of advance methods for extraction of knowledge from spatial datasets has proved the need in rise of geographic knowledge discovery and spatial data mining as a dynamic research area. There is a dire requirement for productive and powerful techniques and methodology for mining valuable knowledge from spatial datasets of high dimension and unpredictable size. The paper highlights recent work in knowledge discovery and spatial data mining. We evaluated several literatures in characteristics of spatial data, common techniques in spatial data mining, techniques involved in spatial data mining and spatial association rule mining. The survey conclude with various outlooks on the significant work done in spatial data mining and recent research work in spatial association rule mining.

Keywords- Spatial Data Mining, Data Mining, Spatial Database, Knowledge Discovery

I. INTRODUCTION

The geographic issues are complex and spatial scale is very large for experimentation hence several geographic researches are observational rather than experimental. Researchers gather information from complex patterns, testing assumptions with observation and discovering new theories. In recent technology information acquisition techniques, for example, remote sensing, global positioning systems (GPS), location aware surveys, and web based geographic information. The information rich era gives chance to secure new learning and improved knowledge of intricate geographic phenomena, like human interaction with environment, handling real problems and social monetary dynamics. The traditional methods of spatial analysis came into existence when data computation power was not as strong as today.

Traditional analysis strategies have few limitations. Initially existing strategies concentrate on a specific sort of relation model or on a constrained perspective (Miller & Han,2009). If the perspective chosen is not suitable for pattern analysis, it cannot show interesting relationships between data. Second, voluminous data cannot be processed with many traditional methods. Third, recent application requires new techniques to analyze and discover data from complex patterns.

There is necessity for successful as well as gainful procedures to mine unpredicted and unidentified data from voluminous information (e.g., a large number of perceptions), high dimension (e.g., several variables), and complexity (e.g., varied data sources, space-time progression, multivariate connections, implicit and explicit interactions and spatial relations). To overcome these difficulties, revelation of geographic learning and spatial information mining has turned into a dynamic research field, emphasizing on improvement of hypothesis, system, and methods for extraction of knowledge and information from spatial databases.

Spatial data mining has significant roots in knowledge and programming, (for instance, classification, visual analytics, information visualization, association rule mining, clustering) and traditional spatial analysis fields, (for instance, exploratory data analysis, analytical cartography, and spatial statistics). The major objective is to develop effective techniques to extract spatial data from complex patterns. Depending on type of methods of a research the efforts of spatial data mining are categorized under different groups such as geocomputation, geovisualization, spatial statistics and spatial data mining.

Knowledge discovery and data mining involve multiple steps, including data choice, cleaning of data, pre-processing, and transformation, consolidation of previous knowledge, analysis with computational algorithm or visual approaches, interpretation and assessment of the results, formulation or alteration of theories and hypotheses, adjustment to analysis and data strategy, assessment of result etc.

In the literature, discovering knowledge is a multistep procedure where as information gathering is portrayed as use of visual, statistical or computational strategies. The information mining strategy ought to be completed after the above procedure to guarantee significant outcome. In this paper, "geographic information revelation" and "spatial information mining" are both utilized to refer knowledge innovation process

II. AN OVERVIEW OF SPATIAL DATA MINING

A. The Definition of Spatial Data Mining

Spatial data mining implies to extract certain spatial relations or information explicitly from spatial database. It requires incorporating spatial database

technology and data mining, the understanding that can be utilized to discovery of relationship between spatial and non spatial relations, spatial data disclosure of connection between reorganization of spatial database, spatial and non spatial information, streamlining spatial query, organizing of spatial learning base. Spatial data mining is used for image database detecting, remote sensing, navigation and route control, geographic information system, environmental study, and various other fields using spatial information.

B. The Characteristics of Spatial Data

Spatial objects consist of distance properties and spatial location. There is an relationship between neighbouring objects, thus correlation between spatial data is complex (Only direction relations and topological relations accounts, the distance between spatial objects and spatial location is related to measure relationship). Spatial data consists of two types such as non-spatial and spatial attributes. Characteristics of spatial data are following:

1) Voluminous data

Few algorithms cannot be used for calculating large amount of information. Spatial data mining task is to overcome the difficulties created through enormous data by developing effective methodology and efficient algorithms.

2) Scale feature of spatial data

The uniqueness observed of spatial data is different on various levels. Scale characteristic of spatial data is a new emergence in unpredictability of spatial data, and is utilized for investigating steady change law of the quality during the instance of speculation and improvement of data.

3) The imprecision of spatial information

Uncertainty exists in large variety of spatial data, for instance, the ambiguity in spatial relationship, the attribute values of fuzzy and the uncertainty of spatial location.

4) The lack of spatial data

There are several overwhelming external compels to avoid gain or loss of information. The most effective method for information recovery and evaluating the intrinsic circulation constraint of information get to be distinctly one of the challenges in solving intricacy of information.

5) Non-linear association between the spatial attribute

The unpredictable representation of space frameworks, represents the perplexing components of the framework inner function, and is a primary task of data mining.

6) Increase in Spatial dimension

The spatial object properties have increase quickly, in remote sensing field, because of the fast advancement of technology, the quantity of bands have expanded to tens or hundreds, along these lines, how to extract information and revelation learning from multiple dimensional space turns into a different area of study.

III. TECHNIQUES IN SPATIAL DATA MINING

There are different sorts of examples that can be found from databases and can be introduced in a wide range of structures. In view of general information mining it is grouped into following primary classifications: association and co-location method, classification, clustering and outlier detection and trend detection.

1. Associations and Co-relation

After extracting all the relevant information we apply grouping strategies that helps in discovery of characteristic roles. These rules represent spatial items as indicated by their “non-spatial” characteristics. There is a necessity for discovery of spatial rules that associate spatial objects with others. The greatest need in mining association rules is to develop improved techniques for selection of appropriate rules from set of discovering rules.

2. Classification

In the database an object is represented by its characteristics. Classification is a strategy, which is used to discover rules that identify the segment of the database into a clear given arrangement of classes. It is analyzed as predictive spatial information mining, as a model is prepared first as per which the entire dataset is examined.

3. Clustering and Outlier Detection

“Spatial clustering” is an arrangement of spatial objects grouped into classes which are known as clusters. Objects within one group demonstrate a high level of uniformity, while the items introduce in other groups are however many non-comparable as could reasonably be expected. Clustering is an exceptionally well-known method to manage large datasets. Clustering algorithm is categorized into four general classes: grid based method, density based method, hierarchical and partitioning method.

4. Trend Detection

A spatial trend is described as a regular change of one or more than one non spatial characteristics when spatially moving away from an initial object. Therefore, the spatial trend detection method used to discover patterns of the characteristic changes with respect to a specific region of several spatial objects.

IV. TASKS INVOLVED IN SPATIAL DATA MINING

Spatial information extraction is an exploration area that is still under development. During the last decade, due to the far reaching utilizations like GPS innovation, mapping and sharing of spatial online information, location based services, remote sensing, more research spaces have made geographic information of high quality to consolidate spatial data, investigation in different reviews, for example, social analysis. Other than exploration area, private ventures and the public have tremendous interest in both utilizing the incomprehensible knowledge for different application needs and contributing geographic information. Therefore it is well known that in the upcoming years there will be new progressive methods developed to use spatial data mining approaches.

Spatial information mining consists of different tasks, and for each task, various distinctive techniques are regularly accessible, regardless of whether statistical, visual, computational. A set of tasks involved in spatial data mining are multivariate geovisualization, classification (supervised classification), association rule mining, and clustering (unsupervised classification).

1. Spatial prediction and classification

On the basis of properties (characteristic qualities) grouping data items into classes is called classification. Grouping data items is known as supervised classification, and the unsupervised classification is called clustering. In supervised clustering we require a dataset to organize the model, a validation dataset to enhance the arrangement, and a test dataset to review the performance of the trained model. Some methods of classification are linear discriminant function (LDF), decision trees, maximum likelihood estimation (MLE), artificial neural networks (ANN), linear discriminant function (LDF), case-based thinking (CBR), support vector machines (SVM), and closest neighbour strategies.

Spatial classification strategies broaden useful characterization strategies to consider the properties of neighbouring object and their spatial relations. A visual approach for spatial grouping in which the decision tree determined with the conventional algorithm is combined with map visualization to uncover spatial patterns of the classification rules. Decision tree induction has additionally been utilized to evaluate and discover spatial decision behaviour. Artificial neural networks (ANN) have been utilized for an expansive assortment of issues in spatial analysis. Remote sensing usually utilizes classification strategies to group image pixels into named classifications.

“Spatial prediction models” forms a collection of “regression analysis” that estimate the variable of nearest neighbours in foreseeing the reliant variable at

a precise point, for example, the spatial autoregressive models. Thus, spatial regression techniques, for example, SAR frequently include the control spatial weight matrix of an $n \times n$, that is computationally intensive if n is bulky. In this way, recent research endeavours have developed methods for estimated solutions for SAR to handle substantial information datasets.

2. Point pattern analysis, spatial clustering and regionalization

Cluster analysis is utilized for analyzing data; it sorts out data items of similar groups into clusters. There are various clustering methods, for example, pattern recognition, machine learning, data mining and statistics analysis. The strategies involved in clustering can be categorized into two groups: hierarchical and partitioning clustering. Partitioning clustering methods, for example, self-organizing map and K-means, separate an arrangement of data items into a variety of non-overlapping clusters. A data item is assigned to the “nearest” group based on proximity measure.

In hierarchical clustering the data items are arranged in a hierarchy with a sequence of nested grouping. The various hierarchical clustering techniques include the Ward's method, single-linkage clustering, complete-linkage clustering, and average linkage clustering. To consider spatial information few type of clustering analysis have been discussed, including regionalization (grouping with geographic contiguity requirements), point pattern analysis (spatial scan statistics with recognition of problem area) and spatial clustering (spatial points clustering). In Spatial clustering, spatial properties are utilized to define the similarity between clusters, (for example, distances and locations). Spatial Classification strategies can be density based or grid based, partitioning or hierarchical.

The aim of regionalization is to optimize main function by grouping spatial objects into adjacent groups. The clusters should be topographically adjacent in Geographic applications. Techniques of regionalization on the basis of clustering can be grouped as: (1) clustering with a spatially weighted dissimilarity measure, which considers spatial properties as a factor in forming clusters (2) To reorganize clusters into regions, non-spatial clustering is followed by spatial processing (3) contiguity compelled clustering that authorizes spatial contiguity during the clustering process (Guo, 2008).

Point pattern analysis, which is also called "hot spot" analysis (Brimicombe, 2007), focuses on the detection of unusual concentrations of events in space, for example, geographic clusters of disease, traffic accidents, or crime. The general research issue is to verify if there is an excess of observed event points. (e.g. disease incidents) for an area (e.g. within a certain location). To discover spatial clusters a number of scan statistics have been developed, for example,

Openshaw, Cross, and Charlton (1990), the group of space–time scan statistics by Kulldorff (1997) and the geographic analysis machine (GAM) by Openshaw et al. (1987). Increasingly statistics for the detection of spatial clusters are available for non-Euclidean spaces, particularly network spaces.

3. Geovisualization

It concerns with the improvement of hypothesis and new strategies that encourage information development through discovery and analysis of geo-spatial data and using visual tools for consequent communication, knowledge recovery and synthesis [2]. The conventional cartography concentrates on the design and utilizes maps for data correspondence while geovisualization emphasizes on data investigation tools, the innovation of interactive maps, generation of hypothesis and knowledge creation.

Geovisualization is closely related to exploratory spatial data analysis (ESDA), and exploratory data analysis (EDA) which join maps and statistical graphics that depends upon human expert interaction with data, create models, visually identify patterns. To handle huge spatial dataset and understand complex dataset, A few challenges need to be address in geovisualization ,which includes (1) Effectively processing huge dataset (2) to discover complex patterns handling multiple perspectives simultaneously and (3) User interface design and strategies to help the innovation process.[2]

To visualize patterns and process huge spatial data, new methodologies are consolidated with computational techniques (like classification, association rule mining, clustering) which helps users to understand patterns and encapsulate data. [2] It is necessary to consolidate visualization with dimension reduction technique to visualize multiple variables and different perspectives. Example, self-organizing maps (SOM), principle components analysis (PCA) or multidimensional scaling.

4. Spatial Association Rule Mining

Association Rule mining was initially considered to find regularities between objects within large transaction databases. Let $I = \{i_1, i_2, \dots, i_m\}$ are a set of objects (i.e. things purchased in transactions, for example, monitor, mouse, keyboard etc.). Let D is a set of transactions, where every transaction T be a set of objects such that $T \subseteq I$. Let X is a set of items and a transaction T is said to contain X if and only if $X \subseteq T$. An association rule of the form: $X \rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of all transactions in D that contain X also contain Y . The rule $X \rightarrow Y$ holds in the transaction set D if $s\%$ of transactions in D contains $X \cup Y$. Confidence denotes the strength and support indicates the frequencies of the rule (Agrawal et al., 1993).

Spatial association rules can utilize many spatial predicates (e.g., close_to, far_away, overlap, intersected.). It is computationally expensive to consider different spatial predicates in deriving association rules from a large spatial datasets. Another potential issue with spatial association mining is that a large number of rules may be generated and many of them are of common knowledge. Domain knowledge is needed to filter out trivial rules and concentrate on new discoveries.

In traditional data mining the extension of association rules; allude to the relationship among objects of space adjacency graph. Spatial association rules are mine using spatial predicates and properties in a spatial database. The spatial association rule represented as $A \rightarrow B [s\%, c\%]$, A is a set of spatial predicates, B is a set of non-spatial predicates, $s\%$ indicates the support of the rule, $c\%$ indicates the confidence of the rules. It requires analyzing various spatial relationships between spatial objects. Many spatial association mining utilizes stepwise enhancement, to be specific, utilizing an algorithm to extract underlying spatial database, and then on the trimmed database.

V. CHALLENGES IN SPATIAL DATA MINING

1. Data access methods in spatial data are distinct from data in relational database, thus complex spatial objects cannot be analyzed using traditional data mining techniques.
2. Spatial data mining algorithms lacks in efficiency and do not have refined discovery patterns. The possibility of the problem of dimension to be solved and the error patterns increases the search space of algorithm. Therefore we need to design an effective knowledge discovery algorithm to remove unnecessary data and reduce the size of the problem.
3. The main reason for growth in database technology is enhancement and development of spatial data mining and database query language must be developed for efficient spatial data mining.
4. The domain expert knowledge is not utilized efficiently and effectively in knowledge discovery process. The process of spatial data mining is unable to control by users.
5. Knowledge discovery through spatial data mining is limited, as it aims to a specific problem.
6. Recently developed knowledge system is constrained to database field. The discovery of knowledge in broad aspect, an expert system like systems for knowledge discovery, decision support system, an integration system of knowledge base, database, networking ,visualization and other technologies.

The issues described above make knowledge extraction difficult in spatial database as compared to traditional relational database, which bring difficulties in research of spatial data mining.

VI. WORK DONE IN SPATIAL DATA MINING

AGRAWAL ET AL. (1993):-

This proposed a technique for the representation of rule is $A \rightarrow B$. Prior knowledge about regular object set properties are used by the algorithm. All these are non empty subsets of regular item set which must be frequent. The number of candidate set combinations is produced prominently. In traditional data mining, from a transactional database frequent data sets are mine using Apriori algorithm. A-priori algorithm established a milestone in Association rule mining.

Ng and Yu (1994):-

This proposed a technique for mining of frequent, strong and discriminating distinctiveness of clusters which are based on thematic maps. The algorithm chooses themes for searching common characteristics of clusters whose values are almost similar. It projected steps for utility values of properties of clusters. Only the properties of thematic maps are analyzed.

J. Han et al. (Han and Fu) (1995):-

This proposed an algorithm for utilization of multi-level logic structure. The algorithm utilizes a multi-level logic structure to mine cross level schema and abstract target rules. Complex logic rules are described well in this algorithm.

Ester et.al (1995):-

This proposed a technique which used ID3 algorithm which uses neighbourhood graph technique which classifies the non spatial properties of classified and neighbouring objects.

Koperski and Han (1995):-

This proposed a multi level technique for spatial rules. Spatial rules like “not disjoint, close_to”, “Intersects, Inside, Contain, Equal “to represent spatial topology. The final rules are represented as “is_a(x, house) A close_to(x, beach) \rightarrow is_expensive(x) (90%)”. Spatial association rules that describe relationship among objects based on spatial neighbourhood relations.

Krzysztof Koperski and Jiawei Han (1995):-

This proposed Discovery of Spatial Association Rules in Geographic Information Databases which is an efficient method for mining strong spatial association rules in geographic information databases is proposed using a two step spatial computation technique. A two step spatial mining method facilitates mining strong spatial Association rules at multiple concept levels by a top down progressive deepening technique.

Fayyad et al (1996):-

This proposed FOCAS: a low level data image processing system that selects object and creates attributes like areas, magnitudes, intensity, orientation, etc. It deals with image database. To identify images of galaxies and stars decision tree methods were used for classification. It analyzed 3TB images of sky. Used for astronomical application.

Inokuchi et al. (2000):-

This proposed AGM algorithm which at previous level apriori algorithm is used to combine sub-graphs and in next level it generates all candidates.

Zaki (2001):-

This proposed SPADE algorithm which decomposes the problem of mining into various small sub problems. In memory sub- problem are solve independently.

Michihiro Kuramochi and George Karypis (2001):-

This proposed FSG algorithm in which previous level frequent sub-graphs are mine using Apriori approach. In the next level all candidates are generated.

Liu Junqiang et al (2002):-

This proposed TD-FP-Growth algorithm for mining frequent patterns and association rules by using two methods: 1. to push multiple minimum support 1.to push minimum confidence. 2. To push multiple minimum support. It is a "Single layer" Boolean type algorithm which is based on "prefix tree"(FPT-Generate). TD-FP-Growth seeks the FP-tree in the top-down order, instead of the bottom order of already proposed FP-Growth.

Donato Malerba (2003):-

This proposed uses the upgrade of Koperski's method in multi-relational data mining approach for mining spatial association rules in Census Data. A logic-based technique for spatial association rule mining in geo-referenced statistical data. The application to census data of Manchester Stockport shows that the expressive power of computational logic enables us to tackle applications that cannot be handled by statistical spatial analysis. The experimentation is the first step towards the fulfilment of objectives of the SPIN (Spatial Mining for Data of Public Interest) managed by Eurostat. Project to put to practical use the timely, cost-effective dissemination of statistical information over the Internet.

L.K. Sharma et al (2005):-

This is implemented using thematic map data. Useful for mining multiple level "positive & negative" spatial association rule of multi-space. For the significant sets of relations and objects strong spatial association rules are obtained.

Appice and Buono (2005):-

This proposed it uses Multi level spatial association rules. Summarizes about advantages obtain from the knowledge obtained on spatial data. We obtain "multi-level" spatial association rules.

Huang et al (2008):-

It implemented spatial temporal patterns. The relationship between temperature and sea salt. It provides temporal information through temporal constraints.

Verhein et al (2008):-

This proposed a Mining spatial temporal patterns a source-thorough fare sink model describing patterns while leaving charts without connections unexplained. Describes the trajectory patterns

Bembenik et al (2009):-

This proposed a method to mine spatial association rule: FARICS. Delaunay diagrams are used to define the objects of neighbourhood

Lee and Chan (2009):-

This proposed a TI-lists structure, to identify trajectory patterns using GBM algorithm and a TI-lists approach is used to find Trajectory Patterns.

R. Parvath (2010):-

Proposed a technique to find spatial association rules which is an effective method for discovering NeuroImaging. The propose system makes easier for the physicians in the diagnosis of Neuro imaging and the results of the decision support system have completely matched with those of the physician's decision.

VII. CONCLUSIONS

This survey shows the significance of spatial data mining and proves it a promising research area that helps in knowledge discovery from huge spatial databases. It can help in understanding the relationship among spatial and non spatial attributes and build a knowledge base of spatial attributes. In recent years, new techniques and several algorithms are projected in order to explore diverse forms of facts from spatial database and mention their strengths and weakness. Spatial data mining is utilized in various areas and has achieved greater results. Spatial data mining will help in declaring one's public approval or support for the growth of science and will boost human skills in order to understand and transform the world.

VIII. ACKNOWLEDGMENT

The author would like to thank to our survey partner, Mrs.Mainaz Faridi at Banasthali University, who share her ideas. The author would like to thank the anonymous reviewers, who provide their valuable and detailed comments to help us improve the work

and Mrs. Mainaz Faridi who helped us in the survey work.

IX. REFERENCES

- [1] Lu W. Han J.et al. *Discovery of general knowledge in large spatial databases*, In *Proc.Far East Workshop on Geographic Information Systems*.Singapore, 1993, 275-289
- [2] Han Jiawei, Kamber M. *Concept and Technology of Data Mining*. Fan Ming, Meng Xiaofeng, etc. Translate.Beijing:China MachinePress,2001.
- [3] N. Sumathi, R. Geetha and Dr. S. Sathiya Bama,"*Spatial Data Mining Techniques Trends and its Applications*", Journal of Computer Applications, Vol.1, No.4, Oct-Dec2008.
- [4]Raymond T. Ng and Jiawei Han," *Clarans: A Method For Clustering Objects For Spatial Data Mining*", IEEE Transactions On Knowledge And Data Engineering, Vol.14, No.5, Sep-Oct 2002.
- [5] R. Agrawal and R. Srikant, *Fast algorithms for mining association rules*. In: *Proceeding of the 20th Very Large DataBases conference*, pp 487-499, 1994
- [6]Deepti Sisodia,Lokesh Singh,Sheetal Sisodia and khushboo Saxena."*Clustering Techniques: A Brief Survey of Different Clustering Algorithms*",International Journal of Latest Trends in Engineering and Technology (IJLTET),Vol.1 Issue 3 Sept 2012.
- [7] Koperski K. Han J, et al. *An efficient two-step method for classification of spatial data*.In:Proc.Int'l Sympon Spatial Data Handling SDH '98, Vancouver, BC.Canada, 1998.
- [8] MILLER H J.HAN Jia-we.i *Geographic data mining and knowledge discovery* [M].London: Taylor and Francis, 2001:3-32.
- [9]Ester M, Kriegel H P, Sander J. *Spatial Data Mining: A Database Approach* In: Scholl M V, ed. *Proceedings of the 5thInternational Symposium on Spatial Databases (SSD'97)*.Berlin: Springer-Verlag, 1997.
- [10] D. Guo and J.Mennis," *Spatial Data Mining and geographic knowledge discovery - An introduction* ", Computers, Environment and Urban Systems 33, 2009.
- [11] R. Agrawal, T. Imielinski, and A .Swami. *Mining Association Rules between Sets of Items in Large Databases*. In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data, pp.207-216, Washington, D.C., May 1993.
- [12] K. Koperski and J. Han. *Discovery of Spatial Association Rules in Geographic Information Databases. n Advances in Spatial Databases*, In *Proceedings of 4th Symposium,SSD'95*.(Aug.6-9,Portland,Maine).Springer-Verlag,Berlin,1995,pp.47-66
- [13] M. Ester, H-P. Kriegel, and X. Xu. *Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification* .In Proc 4th Int. Symp. on Large Spatial Databases (SSD'95),Portland, Maine, August 1995.
- [14] R. Ng, and J. Han. *Efficient and Effective Clustering Method for Spatial Data Mining* In Proc.1994 Int. Conf. VLDB, pp. 144-155,Santiago , Chile, Sept. 1994
- [15] GeoMiner: *A System Prototype for Spatial Data Mining*, Jiawei Han, Krzysztof Koperski and Nebojsa Stefanovic, 1997
- [16] *Exploring Spatial Association Rule Mining for Geo Decision Support System*, Ranjana Vyas, Lokesh Kumar Sharma and U.S Tiwary, Journal of Computer Science 3 (11): 882-886,2007 ISSN 1549-3636,2007 Science Publications.

[17] R.T. Ng and Y. Yu. *Discovering Strong, Common and Discriminating Characteristics of Clusters from Thematic Maps* . Proceedings of the 11th Annual Symposium on Geographic Information Systems pp.392-394, 1997

[18]M.S. Chen J. Han, and P.S Yu. *Data mining: An overview from a database perspective*. IEEE Transactions on knowledge and Data Engineering,8: 866-883,1996