ISSN 2047-3338

Spatio-Temporal Database Using Fuzzy Clustering

K. Kamaraj[1] and Dr. C. Chandrasekar[2]
SSM College of Engineering, Komarapalayam, Namakkal, Tamil Nadu, India
Computer Science, Periyar University, Salem, Tamil Nadu, India
kamaraj41@gmail.com, ccsekar@gmail.com

***Abstract*– Given a large spatio-temporal data of events, where each event consists of the following fields. Event-Id, time, location, event type. A wide range of database applications manage time-varying data. In contrast, existing database technology provides little support for managing such data. The research area of temporal databases aims to change this state of affairs by characterizing the semantics of temporal data and providing expressive and efficient ways to model, store, and query temporal data Temporal database management is a vibrant field of research. In this paper new method for fuzzy itemsets based on of the temporal database manipulations in terms of fuzzy clustering. It also explains the semantics, Patterns using fuzzy clustering. The fuzzy approaches also extract additional values hidden information in comparison with the usual crisp approaches. By using fuzzy clustering queries based on information about two fuzzy item sets , expressions our clients side are supported and are accessing relational database in the same way as the classical SQL.**

***Index Terms*– Query Operators, Temporal Databases, Fuzzy Sets, Membership, Clustering and C-Means Algorithm**

I. INTRODUCTION

IN many applications areas such as having business geographical, engineering and medical diagnosis etc., Sometimes there is huge unstructured numerical data in the form of database. The data analysis is one of new trends in data management in order to gain information from the data being calculated. At present the well known methods are using as statistics, machine learning Neural Network and then data are being useful for exploratory data. A spatiotemporal database is a system that manages both space and time information, such as biological Databases, wireless communication networks, and processing of objects with uncertainty. Basically, spatiotemporal databases are a generalization of spatial databases, which include features of variation coordinates in space, e.g., with plate tectonics, or of otherwise temporal variation of spatial data, e.g. of moving objects, residing temporally in spatially defined locations.

This is somewhat different from areas like CAD databases (solid modeling etc.) where geometric entities are composed hierarchically into complex structures, although the issues are certainly related [2].

Topics of research are the Analysis & Retrieval of spatial and spatiotemporal data. There are three main classes of query languages devoted to spatial databases: Textual languages (natural, SQL and extensions), Graphical languages (QBE) and Visual languages. Since these languages are not sufficient to work with spatiotemporal databases [4]. The updations are made with the existing visual query languages to manipulate with the temporal databases. Visual query with temporal database provides graphical representation of query analysis and performance of a query in temporal databases.

Most applications of database technology are temporal in nature. Examples include financial applications, record-keeping applications such as personnel, medical-record, scheduling applications such as airline, train, and hotel reservations and project management and scientific applications such as weather monitoring [1]. Temporal database management is a vibrant field of research, with an active community of several hundred researchers who have produced some 2000 papers over the last two decades [7].

Fuzzy set was first proposed by Zadeh in 1965. This theory intends to represent and manipulate date and information that possess non-statistical uncertainty fuzzy set allow a degree of membership to be associated with a value in a set. And it can be used to handle both vague and imprecise data.

II. TEMPORAL DATA SEMANTICS

Before considering temporal data models and query languages, we examine, in data model-independent terms, the association of times and facts, which is at the core of temporal data management.

The Main Goals of Temporal Database:

- Identification of an appropriate data type for time
- Prevent fragmentation of an object description
- Provide query algebra to deal with temporal data
- Compatible with old database without temporal data

The *transaction time* of a database fact is the time when the fact is current in the database [5]. Unlike valid time, transaction time may be associated with any database entity, not only with facts [3]. For example, transaction time may be associated with objects and values that are not facts because they cannot be true or false in isolation [6]. Thus, all database entities have a transaction-time aspect.

A. Fuzzy Membership Function

A membership function degree μF~(x) € [0, 1], an element X€ U is said to be in fuzzy set F if and only μF~(x) >0 and to be full member if μF~(x =1 . a triangle fuzzy number is specified .

$$\mu_{F}\tilde{}(x) = \begin{cases} 0 & , x \le a \\ \frac{(x-a)}{(b-a)} & , a \le x \le b \\ \frac{(c-x)}{(c-b)} & , b \le x \le c \\ 0 & , c \le x \end{cases}$$

The attributes (a, b, c) with a< b< c determine the x co-ordinates of the three corner under lying triangle membership function.

III. FUZZY CLUSTERING ANALYSIS

Fuzzy Cluster analysis can be essentially be categorized as one domain of the data analysis in main aim is partitioned a given set of data (or) objects into cluster (sub set of group. This partition should have the properties [9] homogeneity within clusters and heterogeneity between clusters. Since, it will be concerned with data in the form of crisp measurements only.

They would real valued vectors $X = (x_1, x_2, \ldots, x_n) \in R^n$, the Euclidean distance between data can be used as measure of the dissimilarity. The set of an objects $X= (x_1, x_2, \ldots, x_n) \in R^n$, is assigned into C fuzzy clusters. The partition of set n objects (patterns) into C clusters 1≤i≤C , is expressed by an n * c matrix U (or) Uik, where Uik € [0,1] the member slip degree of datum X_K [8].

The C-Means fuzzy clustering:

$$\sum_{K=i}^{n} u_{ik} > 0 \qquad V\ I \in (1,2,\ldots,c)$$

$$\sum_{i=1}^{c} u_{ik} = 1 \qquad v\ k \in (1.2\ldots n)$$

Fuzzy C-means algorithm as a fuzzy version of hard C means clustering is introduced by Dunn and improved by introduction the fuzzfer (m) by bezdera the FCM recognizes spherical clouds of points cluster here is represented by its centre. Main issue in fuzzy cluster analysis is to obtain the optimal assignments of data to clusters:

$$J(X,U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (U_{ik})^m d^2 (V_i \ X_k)$$

Then

$X =\{x_1, x_2, \ldots\ldots x_n\} \in R^n$ is the data

C = The no. of fuzzy cluster

$U_{ik} \in [0,1]$ is membership degree of datum χ_k is the cluster i

$V_i \in R^p$ is the prototype

$(V_i \ ,\chi_K)$ be the Euclidean distance between prototype V_i datum χ_k. Thus,

$d_{ix} = \|\chi_k - v_i\|$ Weighted with the membership degrees is used for minimizing

$$V_i = \frac{\sum_{k=1}^{n} (U_{ik})^m \chi_K}{\sum_{k=1}^{n} (U_{ik})^m}$$

$$U_{ik} = \frac{1}{\dfrac{\sum_{j=1}^{c} d^2 (v_i, x_k)}{d^2 (v_i, x_k)^{2/m-1}}}$$

Then, $\|V^{(t-1)} - V^{(t)}\| \le \epsilon$ weighted with their membership degrees us used for minimizing.

This iteration produced is proceeding until successive optimization.

IV. QUERY

Chittaro and Combi have proposed three alternative visual metaphors for querying temporal intervals [11]. The authors based the expressivity of their visual language on Allen's classification of the relations that may hold between two intervals.

The example query covers the entire USA country database manipulation with visual query in temporal database. The system manipulates STATES, COUNTRIES, CITIES, ROADS and LAKES. The scenario is that the user is interested in the relative impact of main of the highways on jobs, income and the general economy of places for blacks and whites. They start by looking at the number of mobile phones, black and white populations of small towns or small to medium cities. An example query from s set of similar ones might be for all cities with total population < 50,000 within 4 state region of PA-WVMD-VA and that within 5 kilometers – give me the city names black population white population mobile homes at each city.

A. Formulating and Configuring Queries

The system can get inputs from a numeric range query design, which represents an atomic query, e.g., "*select POP1990 from COUNTRIES, where POP1990 between 1,000 and 50,000*".

This query returns the numeric information's for the manipulations.

select *CITIES.NAME, CITIES.WHITE, CITIES.BLACK, CITIES.MOBILEHOME, CITIES.POP1990,ROADS.NAME, CITIES.STATE_NAME* ***from*** *ROADS, CITIES*

where *((CITIES.STATE_NAME IN ('Maryland', 'Pennsylvania', 'Virginia', 'West Virginia') AND (CITIES.POP1990) <= 50000.0) AND SDO_WITHIN_DISTANCE(ROADS.GEOM,*

CITIES.GEOM, 'distance=5') = 'TRUE')

The query statement is automatically extracted from the large database (e.g., Table 1).

V. FUZZY CLASSIFICATIONS USING FUZZY QUERYING

Fuzzy clustering analysis of the numerical data that contain set of the object X= (x_1 ,x_2, …x_n) € R^n, each of which is represented by vectors X= (x $_1$, x_2, x_n) € R^n. Each of these vectors being feature or attribute vector comprised crisp values of all attribute {x_j}.

In order to construct the fast and efficient fuzzy classifier, it is suggested to accomplish classification of the numerical data using fuzzy query processing. Reasons for this suggestion are to support the form of relational database representation of the data to be classified and the existence of standard and the existence of standard any SQL tools. The SQL' s supports quickly searching and retrieving with respect to crisp data in according to crisp query values.

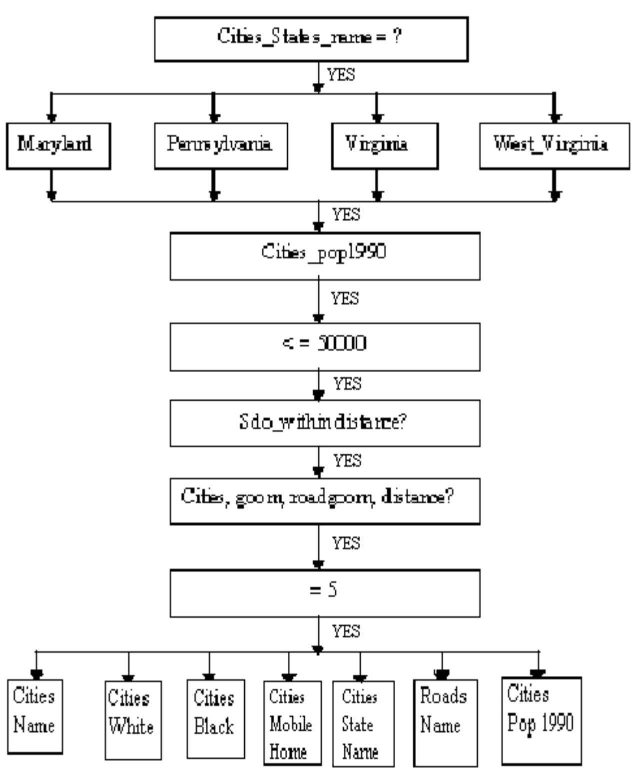

Fig. 1. Transition Diagram of the System

VI. CONCLUSION

At the core of the matter, temporal database queries are often large and complex [10]. Because of this added complexity, it is not only more important, but also more challenging, to optimize temporal database queries. The use of fuzzy querying for classification provides the fast and efficient access of the data that is best dealing with huge data sets optimize temporal database queries.

REFERENCES

[1] T. Abraham and J. F. Roddick. Survey of Spatio-Temporal Databases. *GeoInformatica*, (1):61–99,March 1999.

[2] Ahlberg, C. and Shneiderman, B., Visual Information Seeking: Tight coupling of dynamic query filters with starfield displays. in Proceedings of ACM's SIGCHI Conference, (1994), ACM Press.

[3] An overview of interval encoded temporal mining involving prioritized mining, fuzzy mining, and positive and negative rule mining .C. Balasubramanian et al. / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 1 (3) , 2010, 163-168

[4] Ankerst, M., Jones, D.H., Kao, A. and Wang, C., Data Jewel: Tightly Integrating Visualization with Temporal Data Mining, in ICDM Workshop on Visual Data Mining, (2003).

[5] R. T. Snodgrass. *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufmann Publishers 2000

[6] Catarci, T., M. Costabile, et al. (1995). "Visual Query Systems for Databases: A Survey." Journal of Visual Languages and Computing.

[7] Egenhofer, M. and W. Kuhn (1999). Interacting with Geographic Information Systems. GeographicalInformation Systems: Principles, Techniques, Applications, and Management. D. Rhind, Wiley: New York.

[8] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms,New York: Plenum Press, 1981.

[9] D. M. Rocke and J. Dai, Sampling and Sub-sampling for Cluster Analysis in Data Mining: With Applications to Sky Survey Data, Data Mining and Knowledge Discovery, 7, 215–232, 2003

[10] R. T. Snodgrass. *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufmann Publishers 2000.

[11] Chittaro, L. and Combi, C. (2003) Visualizing queries on databases of temporal histories: new metaphors and their evaluation. Data & Knowledge Engineering, 44 (2), 239-264.

K. Kamaraj is a Professor/MCA, SSM College of Engineering, Komarapalayam, Namakkal (DT), Tamil Nadu, India. He received his UG/PG degree from University of Madras. He is pursuing his Ph.D. from Anna University of Technology Coimbatore. He presented 11 papers in National level journals/conferences/Seminars and three in international conferences. His research area is Temporal Databases using Data Mining.

Dr. C. Chandrasekar received his Ph.D. degree from Periyar University, Salem. He has been working as Reader at Dept. of Computer Science, Periyar University, Salem, Tamil Nadu, India. His research interest includes Wireless Networking, Mobile Computing, Computer Communication and Networks. He was a Research guide at various universities in India. He has been published more than 50 technical papers at various National/International Conference and Journals.

Table 1: Database Example

CITY	WHITE	BLACK	CITY. MOBILE	CITY. POP	CITY.ROAD	CITY.NAME	STATE
Maryland	Yes	No	2245723	45723	Miami City Road	Georgia	Georgia
Pennsylvania	No	Yes	4572482	42841	Green Valley	South Carolina	South Carolina
Virginia	Yes	No	2479914	37421	Blue ridge	North Carolina	North Carolina
West Virginia	No	Yes	4445213	43325	Pan-America	Pan-America	North America
Maryland	Yes	No	2246745	37890	Green Valley	Pan-America	North America