CHAPTER 3

PROBLEM STATEMENT AND RESEARCH METHODOLOGY

3.1 PROBLEM DEFINITION

The evident increase in computerization and collection of large size documents requires proper analysis. The extraction of useful information from the huge databases is regarded as the initial stage in analysis. But the information in temporal database applications are of a time varying nature. Traditional database techniques offered little support to such time varying data. The record of time referenced data in the recording and scheduling applications depends upon the temporal database. In general, the non-procedural query language called Standard Query Language (SQL) is used for the development of applications that access the information from temporal database. The querying and manipulation operations are not supported by SQL since the hidden relations identification between the sequences and subsequence events is difficult.

The mining of casual relations is effectively performed by temporal sequence mining algorithms with the continuous real valued elements collection namely, the time series. The multivariate time-series data analysis requires temporal patterns which capture the relation between two instant events such as before, at the same time and after. The relationship identification is more complex if relations have time duration. The multivariate nature of data and the irregular time sampling makes pattern mining a challenging task with the presence of noisy data, high dimension and feature correlation leading to unsuitability of traditional temporal mining algorithms.
The multi-label temporal patterns discovery identifies the point based sequences effectively. The extraction of interesting patterns from the multi-label representation generates too many patterns for low minimum support which leads to high dimension. The storage of history of database and the re-evaluation of query in the temporal data mining requires a relevant algorithm to handle the rigid names. The temporal data uncertainty introduce the problem of probabilistic query evaluation. Matrix manipulation allows the suitable location evaluation in probabilistic measure. But the data analysis over the temporal data is absent. The acquiring of potential information from the updating models by the probabilistic measure contains faults and low speed. The dimensionality reduction is the major concern in the time-series data mining.

The classification of multi-variant temporal data requires the pattern mining approach. The creation of multi-temporal patterns by using this approach are irrelevant in the classification process. The prediction of noise objects, core objects and the adjacent clusters are difficult in the temporal data mining processes. The evolution of density based clustering algorithms effectively deals with the noise objects identification. But it suffers due to two limitations i.e., capture of noise prints during the clustering of diverse densities and irregular cluster properties.

The data mining of transactional data base requires a set of association rules. The conversion of time expressions to association rules requires the proper controlling. The continuous addition and updation of temporal database required the updation of rules. Hence, re-running of temporal mining algorithm is tedious which degrading the formerly discovered rules. Traditional mining techniques cannot deal with the temporal association rules. The prediction of time-series in the diverse domain of temporal rules is difficult. An efficient subsequence discovery are represented as queries introduced in the problem of searching. The
discovery of exact matches in the clustering model based on approximate matches.

Time-series data mining techniques reveal the complexity in the evaluation of approximate matches. The similarity measurement and high dimensionality are the two problems addressed in time-series mining approaches. The time-series data representation and the employment of original structure are used to solve the above problems. But it is difficult in traditional temporal mining techniques. The lack of conformance between the data acquired in future to the present model introduced the anomalous behaviour in time series. The polynomial time-series subspace representation with the Principal Component Analysis (PCA) basis effectively detects the anomalous behaviour by incorporating the similarity measurement concept. The generalization abilities from the machine learning algorithms are poor. But the subspace representation avoids the over fitting in machine learning algorithms which improved the generalization. The similarity measures inclusion of the time series mining leads to an increase of processing and storage cost. The reduction of dimensionality is easily achieved by subspace representation. However, the maintenance of vital characteristics are poor for a specific dataset.

The study of traditional temporal data mining techniques conveyed the problems in query manipulation in temporal database. The extraction of interesting patterns from the large database is the challenging task in query manipulation schemes. With regard to timely factors, the temporal facts existence in the data base leads to more execution time. Moreover, the results obtained from the traditional algorithms are irrelevant to the user query with slow response. The time-series data mining algorithms such as fuzzy querying, n-tuple timestamp in traditional works were not responsible for an intelligent data retrieval system. The time varying nature, large dimension, huge search space,
and too many pattern generation are the limitations observed in traditional temporal models of SQL, and ORM, which limited the processing speed and efficiency as well. To overcome these limitations, an intelligent data retrieval system with enhanced query manipulation scheme is required.

3.2 OBJECTIVES

The described problems in the temporal database were modelled as objectives of the proposed user-assisted genetic-based image retrieval system creation and TSQL3mining query manipulation designing as given in following sections.

3.2.1 Main Objective

- To develop a framework for improving image retrieval and query manipulations in temporal database.

3.2.2 Specific Objectives

- To develop an user friendly image retrieval system for temporal databases by using an Event Matching Agent (EMAGT) algorithm
- To design TSQL3MINE (Temporal SQL3 Mining) for effective query manipulations in Temporal Database using object oriented techniques
- To compare the existing SQL and ORM with the proposed TSQL3MINE regarding execution time, processing speed, resource efficiency, latency, precision and recall.

3.3 OVERALL ARCHITECTURE OF PROPOSED METHODOLOGY

The proposed methodology of designing the TSQL3MINE query manipulations is graphically represented in Figure 3.1. The architecture consists of two major modules namely, Event Matching Agent algorithm (EMAGT)
based user assisted neural system and temporal SQL3MINE query approach. EMAGT based user assisted system enhances the computation of multimedia query processing. Two sets such as training and validation set are used for evaluation on the basis of similarity measurement. The sequential actions such as reproduction, crossover and mutation are performed to select the best individuals from the temporal data base.

The query processing and navigation is made easy by the TSQL3MINE approach in the next process. The introduction of new keywords which utilizes the operators such as ON, ANY, BETWEEN, AS and JOIN extends the SQL3MINE to temporal (T) SQL3MINE. The similarity calculation between the generated keywords and temporal database is performed on the basis of ASCII values of semantic key words.
Figure 3.1 Architecture of Proposed Research Work
3.4 RESEARCH METHODOLOGY

The research work improves the data retrieval and query manipulation system in temporal database. The two major phases of proposed work are Event Matching Agent based image retrieval and TSQL3MINE query manipulation. EMAGT includes an intelligent image retrieval system with a user friendly interface, which allows the user to retrieve images from a temporal database. The temporal databases aim the manipulation and modelling of different temporal facts with two timely attributes basis namely, valid and transaction time. In SQL standard, a major development is Database management system (DBMS) which propose transaction-time databases. The research work emphasizes that the legacy query results remain unchanged when other queries run on the databases. It also enhances the query manipulation in temporal databases.

The temporal database is used to store the time related data. In the initial phase of the research work, an efficient AGT approach is proposed to perform the event matching. Before retrieving the data, the system is trained with the event based data. Initially, the input video file is converted into multiple frames. All the frames are provided a time range. The image processing technique applied to each video frame and the RGB values, histograms, hue, saturation, intensity, time information, etc. are obtained from the image. The features that are extracted are stored in the temporal database. The image retrieval is based on the events and the user provides the query in natural language query. A few set of images is extracted from the database using natural language query. On applying the image processing technique to the query image, multiple features of the query image are extracted. By exploiting the image features, the SQL query is framed. The image that exactly matches the requirements is then fetched from the database. The proposed genetic framework provided optimal results than the traditional systems.
The second phase of the research work proposes new keywords to provide temporal extension to SQL3. In the Temporal Semantic Mining process, the new keywords are used to define different versions of temporal operators to SQL. The objects are defined and a dictionary is created to provide new keywords. The object oriented operations are performed in the proposed method. The new functions and methodologies are used to perform the query operations. The query processing is simplified by the proposed Temporal Semantic Mining approach. The Semantic concepts are used to improve the relevancy calculation in query processing and also to compute the similarity measures. The proposed technique assures that the legacy query results remain the same when other queries run on the Database (DB). The Semantic method allows easy processing and navigation. The proposed semantic enhances query manipulations in the temporal database and achieves improved time efficiency, search efficiency, processing speed and accuracy when compared with other existing techniques such as (SQL) Standard Query Language and Object Role Modeling (ORM) approach.