Design and Development of Software Framework for Spatio-Temporal Data Analysis and Mining

Prepared By

Ruchi Sharma

12MCEC27

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Design and Development of Software Framework for Spatio-Temporal Data Analysis and Mining

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

Master of Technology in Computer Science and Engineering

Prepared By

Ruchi Sharma

(12MCEC27)

Guided By

Prof K P Agrawal

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Certificate

This is to certify that the Major Project Report entitled **"Design and Development of Software Framework for Spatio-temporal Data Analysis and Mining"** submitted by **Ruchi Sharma (Roll No: 12MCEC27)**, towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this thesis, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof K P Agrawal
Guide & Associate Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Vijay Ukani
Associate Professor
Coordinator M.Tech - CSE
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Undertaking for Originality of the Work

I, **Ruchi Sharma**, Roll. No. **12MCEC27**, give undertaking that the Major Project entitled **"Design and Development of Software Framework for Spatio-temporal Data Analysis and Mining"** submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Nirma University, Ahmedabad, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date: May 2014

Place: Ahmedabad

Endorsed by

Prof K P Agrawal

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof K P Agrawal**,Associate Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

My deepest thank you is extended to **Prof Vijay Ukani**, PG CSE - Coordinator, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for support and continual encouragement throughout the Major Project.

It gives me an immense pleasure to thank **Dr Sanjay Garg**, Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed to **Dr K Kotecha**, Director, Institute of Technology, Nirma University, Ahmedabad for the motivation he has extended throughout course of this work.

I extend my sincere thanks to SAC-ISRO, Ahmedabad for providing the data set.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, and Ahmedabad for their special attention and suggestions towards the project work.

The blessings of God and family members make the way for completion of Project. I am very much grateful to them.

- Ruchi Sharma

12MCEC27

Abstract

Space and time are the basic dimensions of our existence and thus needed to be explored in order to reveal hidden information and useful knowledge. Spatiotemporal data are generated in various fields. Such as agriculture, defence, meteorology, crop sciences, medicine, transportation etc. These datasets are growing rapidly with the use of geographic location based devices and satellite images. Data are captured in various formats at multiple levels of granularity both in space and time. The increased volume of datasets demand research in the field of spatiotemporal data mining and analysis.

In the thesis, a framework has been proposed and developed for the analysis and mining of spatiotemporal data. It has been designed as two subsystems i.e. Automated and Customized. Automated system is meant for multi-dimensional analysis and visualization. A Data warehouse is also created for the use of Automated System. For performing all the data mining tasks manually, customized system has been developed, which is an extension of open source data mining tool, WEKA.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Objective

- To design and develop a software framework for analyzing and mining geographic spatio-temporal data.
- To design and create a data warehouse for online analytical processing on earth observation data.

1.2 Motivation

Spatiotemporal data are generated in various fields. Such as agriculture, defence, meteorology, crop sciences, medicine, transportation etc. These datasets are growing rapidly with the use of geographic location based devices and satellite images. Data are captured in various formats at multiple levels of granularity both in space and time. The increased volume of datasets demand research in the field of spatiotemporal data mining and analysis.

1.3 Scope of the Work

There are several challenging issues, that have not been resolved and implemented. Work is needed to be done in following areas :

- Support for heterogeneous Spatial-Temporal datasets.
- Resolving scale effect i.e. spatial resolution and temporal granularities before data mining.

- Effective visualization of spatio-temporal data and output patterns after mining.
- Effecient Techniques for spatio-temporal clustering, classification and assciation mining.
- Development of a generalized framework that can provide support for heterogeneous spatio-temporal data input, provide preprocessing techniques to resolve scale effect, provide facility to embedd effecient data mining techniques and components to visualize output in effective and easy to understand way.

1.4 Thesis Organization

The thesis report is organized as follows:

In the **first chapter**, introduction about the work is given.

Second chapter covers concepts related to the domain i.e. spatio-temporal data mining and data warehousing, GIS etc.

In the **third chapter**, literature survey has been covered.

Chapter four is all about proposed system design i.e. what the system is, what are the proposed functionalities.

Fifth chapter covers all information related to the implementation. It includes description about tools and libraries used, datawarehouse design & creation, system implementation details etc.

In the **sixth chapter**, developed use interface and its usage have been elaborated.

In the **last chapter** the thesis has been concluded and future scope is mentioned.

Chapter 2

Domain knowledge

2.1 Types of Data

Data exists in one of the three forms ie. structured, semi-structured and unstructured.

Structured Data

Data that follows strict format or fixed structure. Data stored in the relational databases comes under this category, as every record in a table follows predefined schema and constraints.

Semi-structured Data

Data that may have certain structure but not all information collected may follow identical structure. Some attributes may exist in some entities but not in others.

Unstructured Data

Data that does not have identifiable structure comes under this category. For example images, videos, texts and documents etc.

2.2 Data Warehousing

Data Warehouse

Accumulation of transactional data over large periods of time after doing required preprocessing can be referred as a data warehouse.

Data warehousing can be done for serving different motto for an organization. It can be used for fulfilling one of three purposes, as follows.

1. **Archiving**
   Large amounts of data is stored for long periods of time but need not be queried or processed.

2. **Data Mining**
   Datawarehouse can be created for accumulating large amount of data each day, for mining hidden information and patterns.

3. **Analytics**
   Highly structured data is stored for the use of analytical processing tools to generate reports and graphical visualizations.

2.3 Spatio-temporal Data and its Mining

2.3.1 Spatio-temporal Data

Data that takes space and time dimensions in its considerations can be called as Spatio-temporal data. It can be classified according to the kind of changes happening over time:

1. Changes in existence, i.e. appearance and disappearance.

2. Changes of spatial properties: location, shape or/and size, orientation, altitude, height, gradient and volume.

3. Changes of thematic properties expressed through values of attributes: qualitative changes and changes of ordinal or numeric characteristics (increase and decrease).[1]

C1assification of such data types, based on the two dimensions:

- *the temporal dimension* describes to which extent the evolution of the object is captured by the data. On the basis of this, there can be three different cases. In the first case, only static view of each object is available. It is applicable when objects do not evolve over time. In second case, each object can change its state, but only its most recent value is being captured. In this case, we lack any knowledge

about its past history. In third case, whole history of changed states of all objects is stored.

- *the spatial dimension* describes about the mobility of objects. There can be two cases, the objects considered can be at a fixed location or they can move, i.e., their location can change with respect to time.

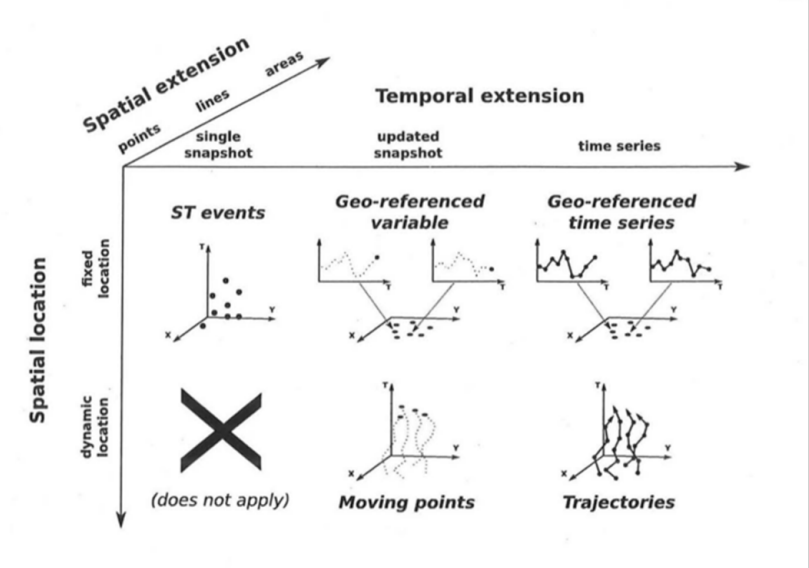

Figure 2.1: Context of ST mining [2]

2.3.2 Spatio-Temporal Data Mining

Extraction ofimplicit knowledge, spatial and temporal relationships, or other patterns not explicitly stored in spatiotemporal databases is called spatio-temporal data mining. Spatio-temporal data needs separate data mining algorithms, as algorithms designed for classical datasets are not appplicable to those due to high auto-correlation among data over space and time dimensions.

Data mining techniques may include, spatio-temporal version of:

- Clustering
- Classification
- Association rule mining
- Time series analysis

2.3.3 Domain of Applicability

Spatio-temporal data mining ans analysis can be helpful in various domains. Some of the possible applications are mentioned below:

- Crop sciences: soil quality changes,harvesting,seasonal grasshopper infestation,land usage management.
- Transportation: fuel efficient routes,tracking vehicle movement,traffic control, traffic planning, vehicle navigation.
- Forestry: tree planting planning, forest fires, hydrology patterns, canopy development,forest growth.
- Medicine: supervising developments in embryology, patients cancer developments.
- Meteorology: Moving storms, tornados, movement of precipitation areas,developments of high pressure areas, droughts.
- Biology: animal movements, relocation and extinction.
- Geophysics: earthquake histories, volcanic activities and prediction. [3]

2.4 GIS

Geographic Information System (GIS), is a category of software systems. It includes functionalities like analysis, modification and visualization of geographic data. Many open source libraries are available for implementing basic GIS functionalities. One can make use of the available libraries and APIs to develop spatio-temporal analysis software as per the requirement.

GIS softwares can be categorized in various classes, on the basis of the functionalities that they offer.

The table shows various categories of GIS software and their corresponding functionalities. [4]

GIS task vs. GIS software	query/select	storage	exploration	create maps	editing	analysis	transformation	creation	conflation
Desktop GIS									
- Viewer	●	●	●	○					
- Editor	●	●	●	●	●		○	●	
- Analyst/ Pro	●	●	●	●	●	●	●	●	●
Remote Sensing Software		●	●	○	●	●	●		
Explorative Data Analysis Tools	●	●	●	●	○	●	●		
Spatial DBMS	●	●				○	●		
Web Map Server	●		●	●	○			○	
Server GIS / WPS Server	●	●		●		●	●		●
WebGIS Client									
- Thin Client	●		●						
- Thick Client	●	●	●	●	●	●		●	
Mobile GIS	●	●	●		●			●	
GIS Libraries		●		●		●	●		●

Figure 2.2: GIS software vs Functionalities

Chapter 3

Literature Survey

There are few number of spatio-temporal data analysis softwares or tools available. As the result of exhaustive literature survey, we have found some relevant tools that are summarized in the table below.

Software	**Language**	**Area**	**Category**	**Library Used**
GeoDMA	C++	Image Processing, Data Mining	-	TerraLib
GAV(GeoAnalysis Visualization)	Csharp	Visualization Analytics	Visualization Framework and ToolBox	DirectX
S-TVQE	Java, XML	Query Handling	-	-
CubeView	Java, XML	Image Processing, Data Mining, Visual Analysis	Web-Based ST-Mining Tool	-
GeoSTAT	Java, XML	Exploratory Data Analysis	Web-Based ST-Mining Tool	-
Weka GDPM	Java	Data mining and visualization	open source data mining tool	-

Table 3.1: Survey of Spatio-temporal Data Analysis Tools

Paper [10] discusses about an analytical tool, GeoDMA for processing geo-spatial images. It classifies the data in the images into predefined objects. It integrates data mining techniques, multi-temporal feature handling and image analysis tools. The tool is implemented in c++ and uses TerraLib library.

Framework proposed in this paper [19] is based on visual analytics. It advocates use

of sophisticated visualization modules in analytical software to interpret given data or information from multiple perspectives. It introduced an enhanced parallel coordinate component, various information visualization and geographical visualization components. The framework is developed in c# and uses DirectX library.

New Visual Query System is presented in this paper [5], which supports querying spatio-temporal data. It provides a graphical user interface to choose and apply spatial, temporal and classical filters. The system was only designed for querying spatio-temporal data. It lacks functionalities to answer complex queries, that can only be answered through data mining.

The paper [6] presents application of DB-SCAN algorithm to discover regions with similar characteristics of sea-water. System schema for the purpose of analytics and mining was presented. The system scheme described about the architecture and organization of input, data storage according to usage, data analysis methodologies used and forms of outputs.

In the paper [7], a software is described that takes space and time as input parameters for each measure. The system is meant for the analysiss of traffic data. The data is preprocessed before loading into the data warehouse module. It provides Online Analytical Processing(OLAP) operations and multi-dimensional analysis for data visualization. It aso provides data mining analysis techniques such as classification, clustering and outlier detection. The output is displayed in the form of maps and charts. The whole system presented is web-based.

The paper [8], presents a spatio-temporal data analysis approach, where spatio-temporal clustering in combination of visualization components for temporal geo-spatial data is used. The user interface that is presented in the paper offers components for selction and fitering of spatial, temporal and thematica data. The software lacks other data mining techniques like association rule mining, time series analysis etc. and also it does not offer functinality for performinig OLAP operations.

Data warehouses are essential tools in data analysis and decision making. Data cube

operator simplifies burden of writing multiple OLAP queries. The paper [9] presents an idea of MapCube, which is an spatial version of data cube operator. It takes input in the form of maps, classical data and presents output after OLAP cube processing in the form of maps and graphs.

Chapter 4

Proposed System Design

Spatio-temporal geographic data can be analyzed in three ways:

- **By Visualization**

  Data can be visualized over space or time dimension and then can be interpreted by us human beings. Over space dimension, classified or clustered geographical data can be visualized in the form of maps. Over time dimension data can be plotted as charts or graphs for getting the trend.

- **By OLAP or Multi-Dimensional Analysis**

  Data can be summarized at multiple levels of granularity, over space as well as time dimension through OLAP cube operations. This sort of analysis can be helpful to answer many spatio-temporal queries.

- **By Data Mining**

  Online analytical processing can resolve simple queries but to get answers of complex queries like trend discovery, dependency analysis and future predictions; data mining techniques like clustering, classification, time series analysis etc. are needed to be applied.

In our proposed system, we have designed and developed two separate systems for analysis and named them as:

- Automated System
- Customized System

Automated System is named so, because the user does not need to do anything rather than selection of dimensions and parameters from the user interface. Backend process will be started executing automatically after that and results will be displayed in the form of maps and graphs.

Customized System is meant for data analysts, where everything has to be done manually i.e. preprocessing, applying data mining technique and selection of visualization.

4.1 Automated System

Input to the automated system is a combination of space, time and measure. In space dimension there are three levels of granularities: country, state and district. Time dimension has three levels of granularities: Year, quarter and month. There are three measures: vegetation index, rainfall and temperature. Appropriate data will be retrieved through

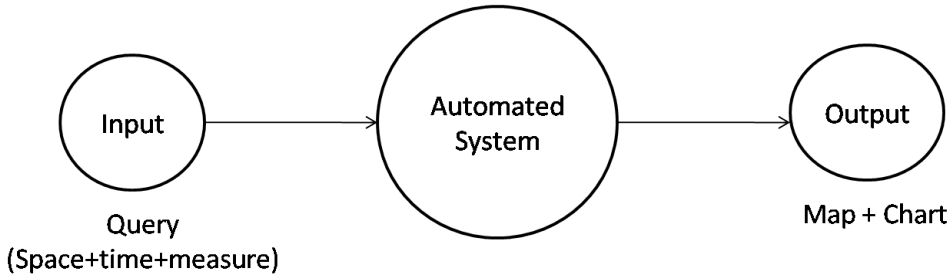

Figure 4.1: Automated system's Input Output

OLAP server and output will be displayed in the form of maps and charts. Data flow of the system is shown in the figure. At first, multi-dimensional query will be fired through the user interface, which will be passed to the OLAP server for processing. OLAP server uses logical schema cube, for mapping relational database to multi-dimensional data model. Retrieved information is then displayed through graphs and maps.

4.2 Customized System

Customized system can be thought to be consisting of several modules according to their functionality. The modules are:

- Input
- Preprocessing

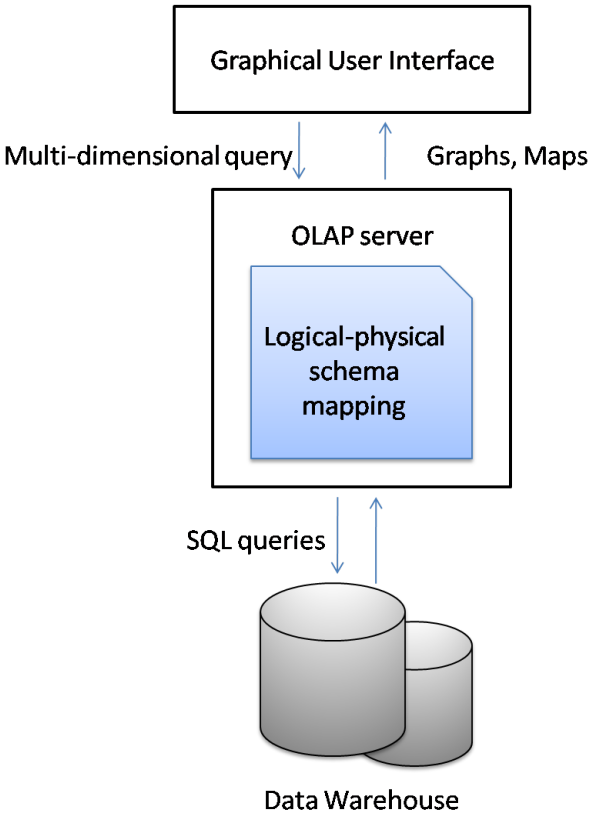

Figure 4.2: Data Flow diagram for Automated System

- Data mining technique
- Visualization

Input

One year spatio-temporal data can be given to the system in any of the following formats:

- csv
- arff
- Database

Multi-year data can be there in separate files. So, input module would facilitate multiple file selection.

Preprocessing

System will facilitate all the pre-processing tasks that are incorporated in WEKA. For example:

- Manual data modification

- Range classification
- Attribute subset selection etc.

Data Mining Techniques

Data mining techniques developed in R, are integrated in the customized system and separate window for parameter setting is also displayed.

Visualization

Spatio-Temporal knowledge after data mining can be visualized in the form of maps and graphs for interpretation.

Chapter 5

Implementation

The chapter covers description about the system architecture, working schema and tools & libraries used for the implementation.

5.1 Tools and Libraries Used

Tools and libraries used in the software implementation are all open source and written in Java. These are as follows:

Open Souce Library or Tool	**Purpose of Use**
Eclipse IDE	Development of software
PostGIS	Spatial Database Management and Operations
GeoTools	Spatial Data formats support
Geomondrian	Spatial Data Cube creation and operations
Weka	Data Mining Techniques source code
JfreeChart	Graph Plotting

Table 5.1: Relevant Open source tools and libraries

A **rich client platform (RCP)**

With a RCP, programmers can build their own applications on existing computing platforms. The platform allows the seamless integration of independent software modules like graphic tools, spreadsheets and mapping technologies into a software application easily.

PostGIS

PostGIS is an extension to the PostgreSQL object-relational database system, which allows GIS (Geographic Information Systems) objects to be stored in the database. It includes support for GiST-based R-Tree spatial indexes, and functions for analysis and

processing of GIS objects.

GeoMondrian

Mondrian is an OLAP engine written in Java. It executes queries written in the MDX language, reading data from a relational database (RDBMS), and presents the results in a multidimensional format via a Java API. GeoMondrian is a "spatially-enabled" version of Mondrian. It provides a consistent and powerful support for geospatial data. It also provides geo-extensions to MDX.

Geotools

It is an open source (LGPL) Java code library which provides standards compliant methods for the manipulation of geospatial data, for example to implement Geographic Information Systems (GIS). The library implements Open Geospatial Consortium (OGC) specifications as they are developed.

JFreeChart

It is an open source Java chart library for implementing and displaying several types of charts.

Weka

It is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

5.2 Automated System

The Automated system has been made up of 4 layers:

- **User Interface** has been made up of SWT (swing widgets toolkit) controls using eclipse IDE and is built on RCP framework.
- For **Data Storage**, data-warehouse had been created, where physical databases are stored in postGIS and logical data cube is created in XML.
- Geo-Mondrian is used as an **OLAP server** for doing multi-dimensional analysis.

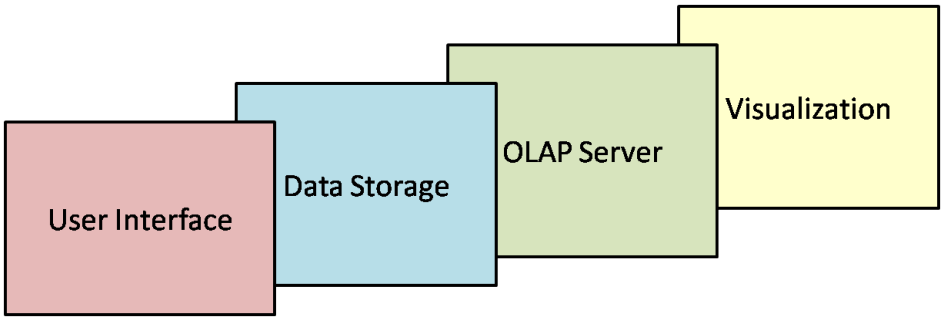

Figure 5.1: Layers of Automated System

- **Visualization layer** consists of map and chart visualization modules, which is implemented with the help of open source libraries, i.e. GeoTools and JFreeChart respectively.

5.2.1 Working Scheme

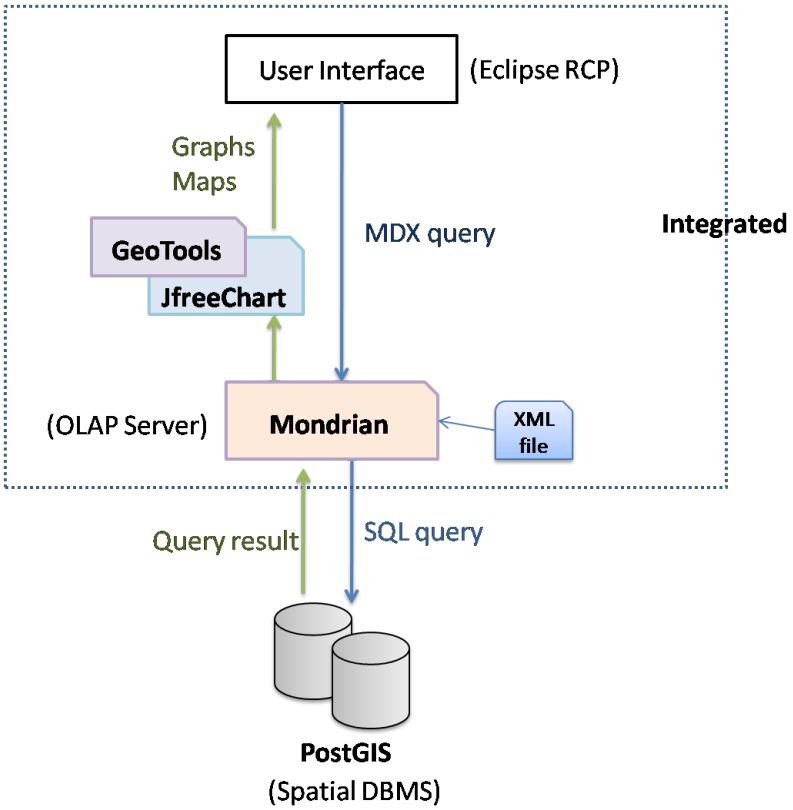

Figure 5.2: System flow diagram

The automated system works as follows. User selects relevant parameters and granularities of space and time dimensions. Selected combination of space-time dimensions and measures is mapped to pre-stored MDX (multi-dimensional expressions) query. MDX query is further passed to the OLAP server, there with the help of logical cube schema,

MDX queries fetch multi-dimensional data from the data-warehouse.
Retrieved data at the desired level of summarization and granularity is then visualized using Geotools and JFreeChart libraries.

5.3 Customized System

Customized system is an extension of WEKA for supporting, spatio-temporal data mining techniques. Extended Functionalities to WEKA, that were missing prior:

- New Panel for spatio-temporal data mining
- Multiple file selection
- Parameters setting for techniques
- Integration of R to Weka

5.4 Data Warehouse Design and Creation

Data warehouses generalize and consolidate data in multidimensional space. For multi-dimensional analysis through the automated system, spatio-temporal data warehouse had been created.
The process followed for the data warehouse creation is as follows: Firstly we designed a snowflake schema for modeling multi-dimensional data for data warehouse. Then the actual data was preprocessed for cleaning, noise removal and splitting. Preprocessed data then stored in the postGIS for use. After that, logical schema was written for multi-dimensional analysis of actual data through OLAP server.

5.4.1 Conceptual Model

5.4.2 Multidimensional Data Model

The snowflake schema has been adapted for modeling actual data into multidimensional database for efficient OLAP analysis. Data is stored as a fact table and multiple dimension tables.
For our data, fact table is having 3 measures and 2 dimensions. Measures can be categorized and then be accessed on the basis of dimensions.

- **Measures:** Vegetation Index, Rainfall and Temperature.

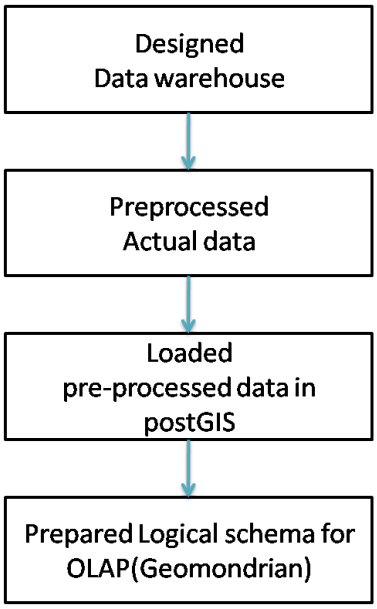

Figure 5.3: Procedure followed for creating data warehouse

- **Dimensions:** Space and Time.

Space dimension is normalized, thus splitted into 2 separate tables for district and state.

5.4.3 Physical Database

The data is stored as fact and dimension tables in postGIS. Data collected is 10 year data, from 2002 to 2011, for vegetation, rainfall and temperature for all grids of india. Data is taken from SAC, ISRO Ahmedabad. The actual single year data tables had been normalized and splitted in the form of fact and dimension tables. Each fact table consists of 29,97,061 rows.

5.4.4 Logical Data Cube Schema

Mondrian uses the concept of a schema to map from the logical data structure used for analysis to the physical structure used in the data warehouse. So, for the purpose logical schema is created in XML. It consists of 10 cubes, one for each year in the single schema file. Multiple levels are defined for varied granularities in space and time dimensions.

Logical Schema in XML

$<?xmlversion = "1.0"?>$

$< Schemaname = "Automated2" >$

$< Cubename = "VEGI\_2002\_2003" >$

$< Tablename = "fact\_2002\_2003" / >$

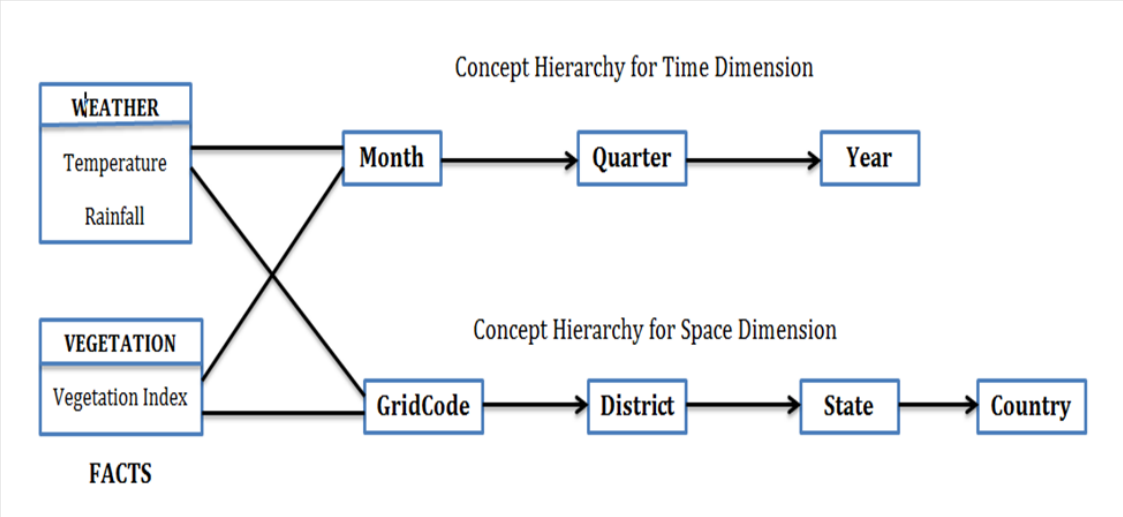

Figure 5.4: Multi-dimensional conceptual model

$< Dimensionname = "Time" foreignKey = "record\_id" >$

$< HierarchyhasAll = "false" primaryKey = "record\_id" >$

$< Tablename = "time\_2002\_2003" / >$

$< Levelname = "Year" column = "year" uniqueMembers = "false" / >$

$< Levelname = "Quarter" column = "QUARTER" uniqueMembers = "false" / >$

$< Levelname = "Month" column = "MONTH" uniqueMembers = "false" / >$

$< /Hierarchy >$

$< /Dimension >$

$< Dimensionname = "Space" foreignKey = "gridcode" >$

$< HierarchyhasAll = "false" primaryKey = "gridcode" primaryKeyTable =$
"$space\_dist$" $>$

$< joinleftKey = "dist\_id" rightKey = "dist\_id" >$

$< Tablename = "space\_dist" / >$

$< Tablename = "space\_state" / >$

$< /join >$

$< Levelname = "Country" table = "space\_state" column = "country" uniqueMembers =$
"$false$" $/ >$

$< Levelname = "State" table = "space\_state" column = "state" uniqueMembers =$
"$false$" $/ >$

$< Levelname = "Dist" table = "space\_state" column = "dist\_name" uniqueMembers =$
"$false$" $/ >$

$<Levelname="GridCode"table="space\_dist"column="gridcode"uniqueMembers=$
$"false"/>$

$</Hierarchy>$

$</Dimension>$

$<Measurename="NDVI"column="vi"aggregator="avg"/>$

$<Measurename="RainFall"column="rainfall"aggregator="avg"/>$

$</Cube>$

$<Cube>$

.

.

.

$/Cube$

$</Schema>$

5.4.5 Multi-dimensional Expressions

GeoMondrian uses MDX queries for multi-dimensional analysis of data. Possible number of queries through user interface are mentioned in the figure.

VI for Year 'A'

- **For District 'P'**

  SELECT {[Time].[Year].[A]} ON COLUMNS, {[Space].[Dist].[P]} ON ROWS FROM [VEGI_A];

- **For State 'Q'**

  SELECT {[Time].[Year].[A]} ON COLUMNS, {[Space].[State].[Q]} ON ROWS FROM [VEGI_A];

VI for Quarter 'X' for Year 'A'

- **For District 'P'**

  SELECT {[Time].[Quarter].[X]} ON COLUMNS, {[Space].[Dist].[P]} ON ROWS FROM [VEGI_A];

- **For State 'Q'**

  SELECT {[Time].[Quarter].[X]} ON COLUMNS, {[Space].[State].[Q]} ON ROWS FROM [VEGI_A];

VI for Month 'X' for Year 'A'

- **For District 'P'**

  SELECT {[Time].[Month].[X]} ON COLUMNS, {[Space].[Dist].[P]} ON ROWS FROM [VEGI_A];

- **For State 'Q'**

  SELECT {[Time].[Month].[X]} ON COLUMNS, {[Space].[State].[Q]} ON ROWS FROM [VEGI_A];

Figure 5.5: MDX queries

Chapter 6

User Interface and Usage

The chapter describes about the user interface of the overall system, its utilities and usage. As mentioned previously, there are 2 subsystems developed i.e. 'Automated' and 'Customized'. So, seperate user interfaces for the mentioned subsystems will be displayed.

6.1 Start Window

At first, start window will be splashed over the screen, asking for an option to be selected between the two. Depending on the choice of the user, user interface for the selected subsystem will be displayed next.

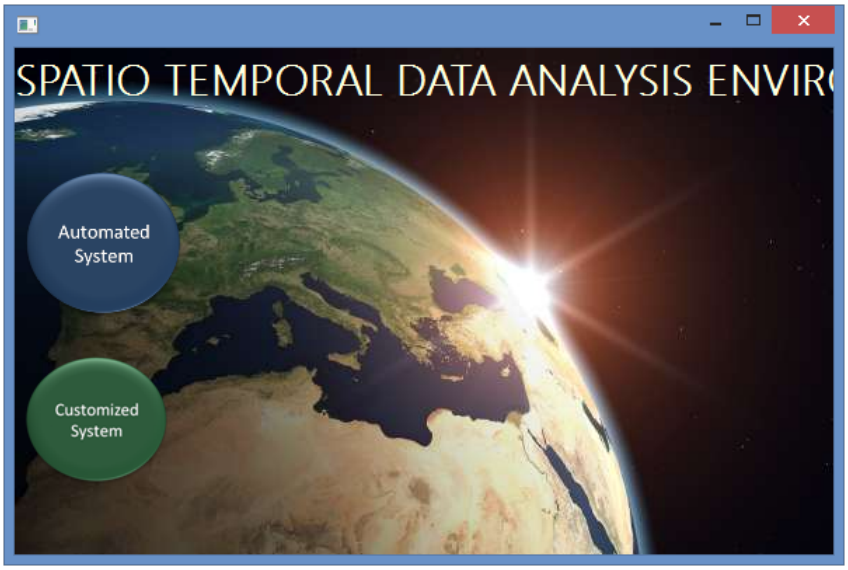

Figure 6.1: Main window for selection

6.2 Automated System

It offers three groups of selection controls: Measures, space dimension and time dimension.

Group 1 – Measures offer three parameters to select from, as vegetation index, rainfall and temperature.

Group 2 – Time dimension has three levels of granularity i.e. Year, quarter and month. Furher selection of particular time will be displayed upto selected level of granularity.

Group 3 – Space dimension also offers three levels of granularity to select from, i.e. Country, state and district. Choices displayed for the lower level will depend on the higher level selected.

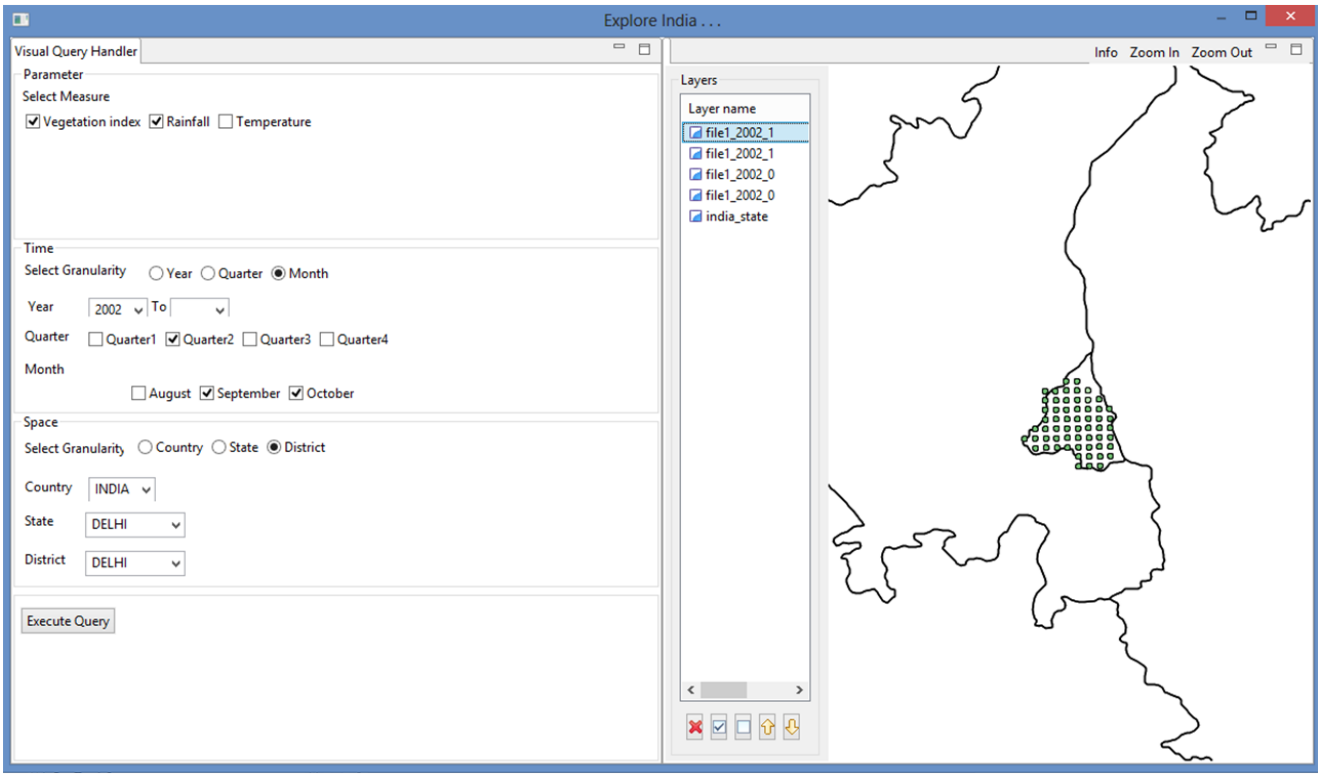

Figure 6.2: User Interface for Automated System

Output is displayed in the form of Maps in the cascaded panel and chart for time series

is diplayed in a pop-up window as shown in the figure 6.3.

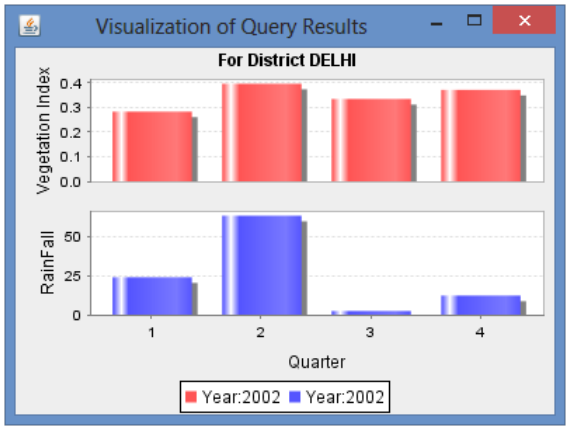

Figure 6.3: Chart for selected parameters over time dimension

Map window can be zoomed in and Layers of map can be made visible or invisible according to user's wish. As shown in the figure 6.4.

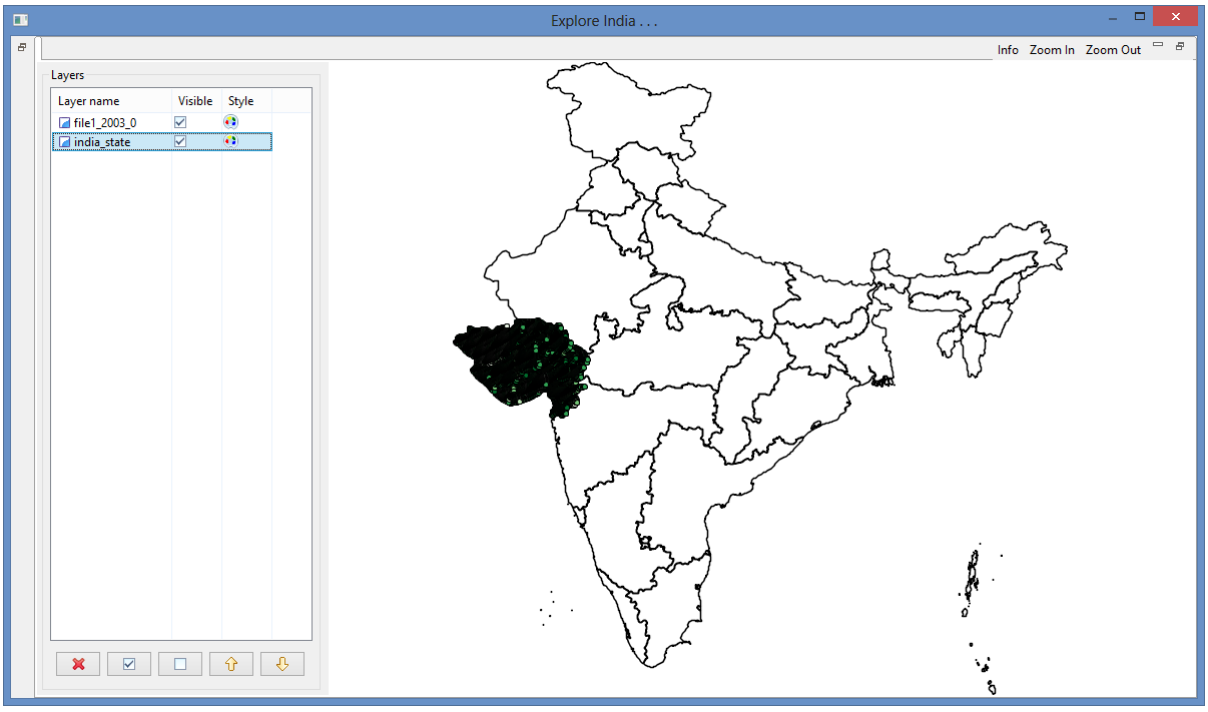

Figure 6.4: Map Panel and layers utility

Toolbar of the map panel facilitates some functionalities i.e. zoom-in, zoom-out and information display at the selected point of location. Figure 6.5 shows zoom-in map for Gujrat district.

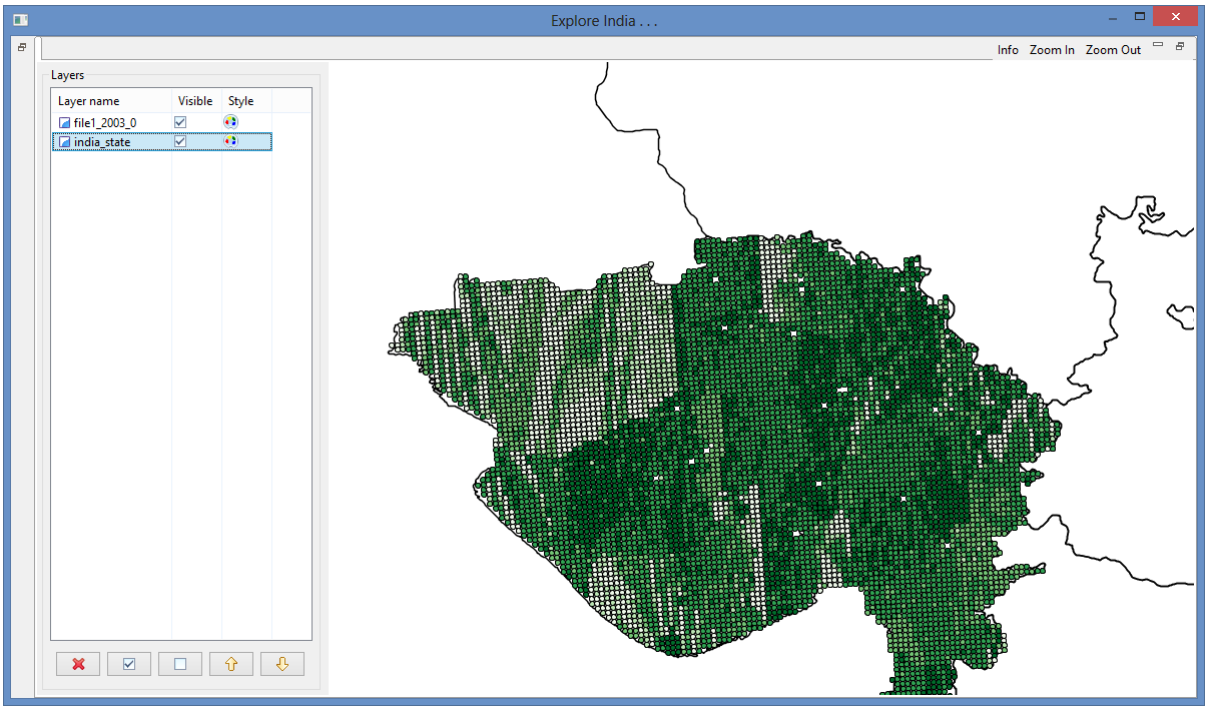

Figure 6.5: Zoom-in Map display

6.3 Customized System

As customized system is an extension to Weka, new panel is added for aptio-temporal data mining. The panel incorporates three subpanels:

1. For Input data selction
2. To display multiple files that are selected
3. For data mining technique selection

On the selection of file mode for data input, file chooser window will be pop-up. The file chooser provides multiple file selection facility that was lacking otherwise in Weka. This utility is beneficial for spatio-temporal data, as spatio-temporal data is often arranged in separate files year-wise.

Files that are selected for further processing are displayed in the text area as shown in the figure. There are two options provided for the selection of data mining technique.

- Spatio-temporal clustering
- Spatio-temporal time series analysis

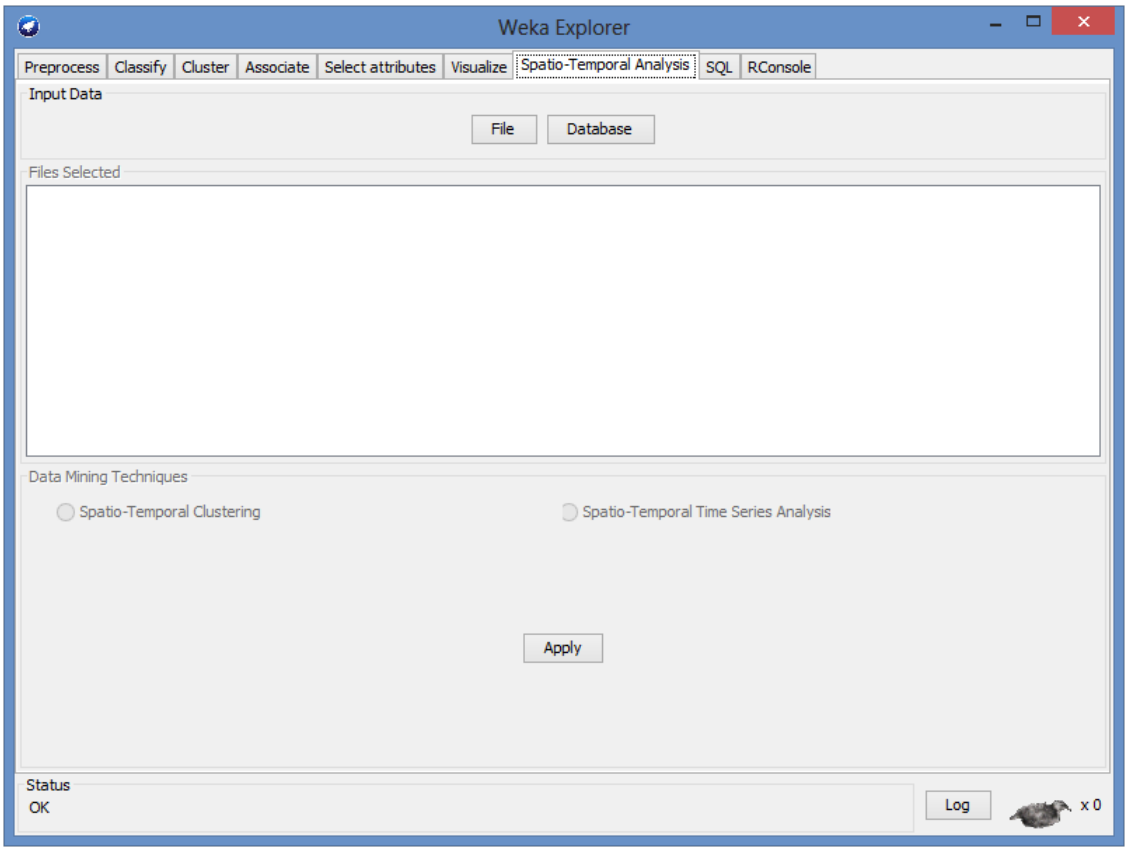

Figure 6.6: Added Spatio-temporal data analysis panel to Weka

On the selection of particular technique, pop-up window will be opened up for parameter selection for that technique, as shown in the screenshot.

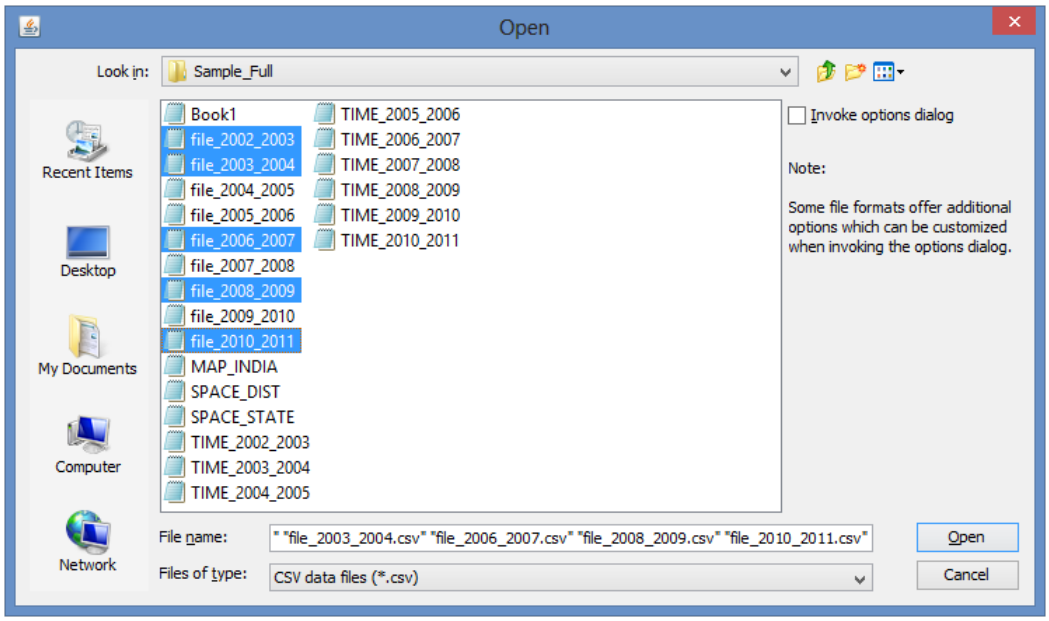

Figure 6.7: File Chooser, showing multiple file selection utility

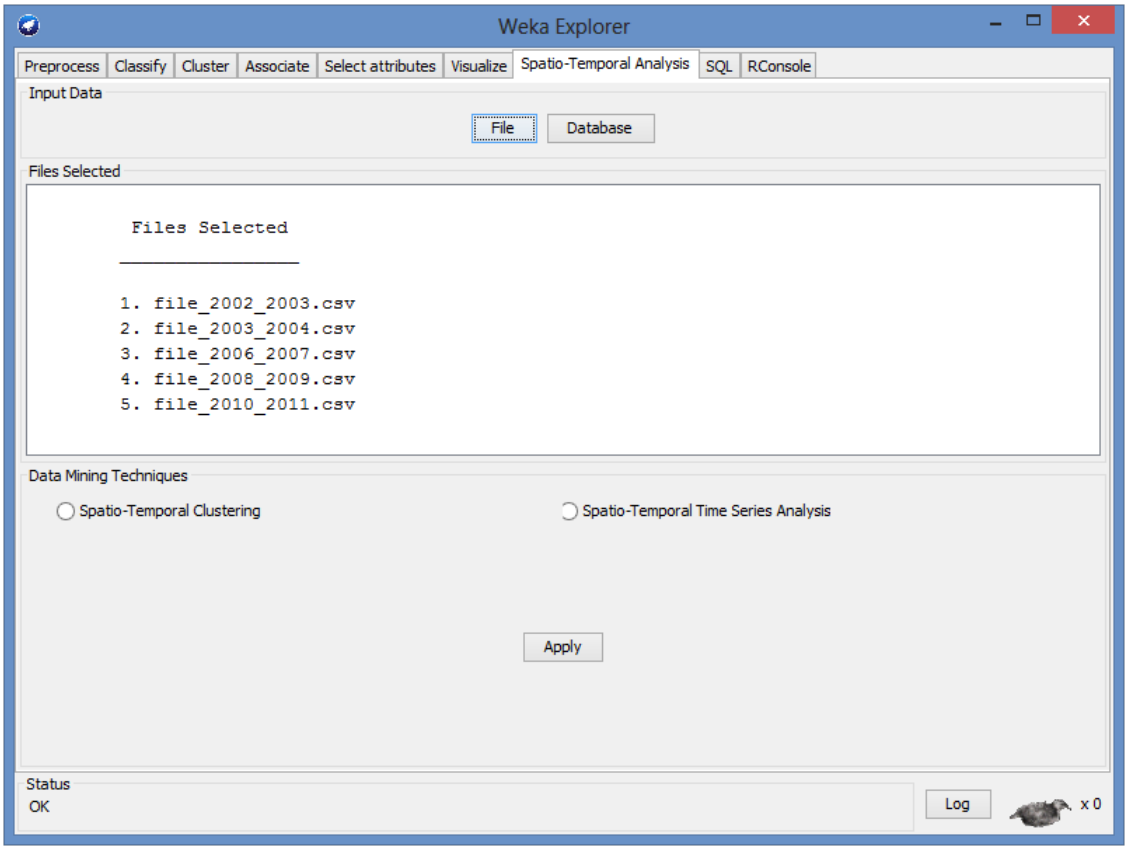

Figure 6.8: Displaying selected files

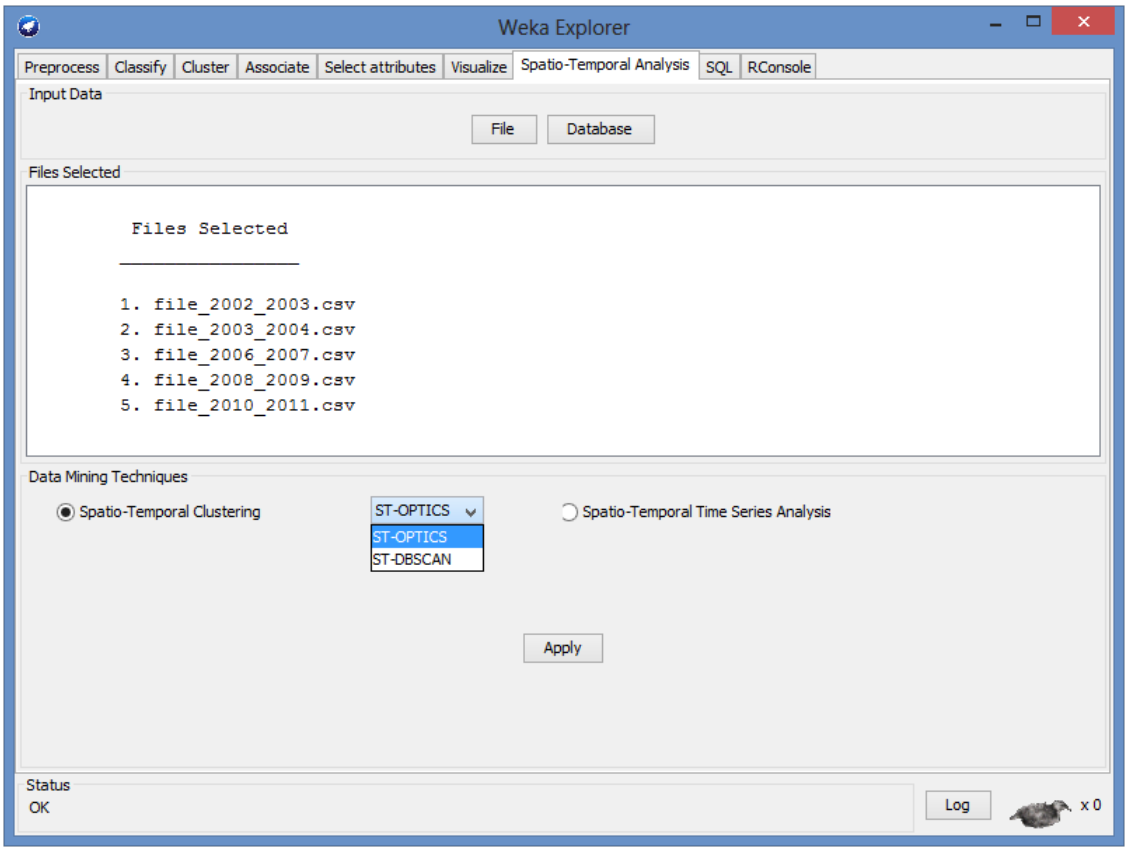

Figure 6.9: Selection of data mining technique

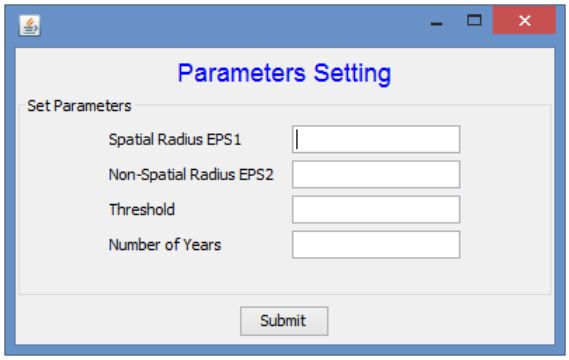

Figure 6.10: Parameter setting window for spatio-temporal clustering

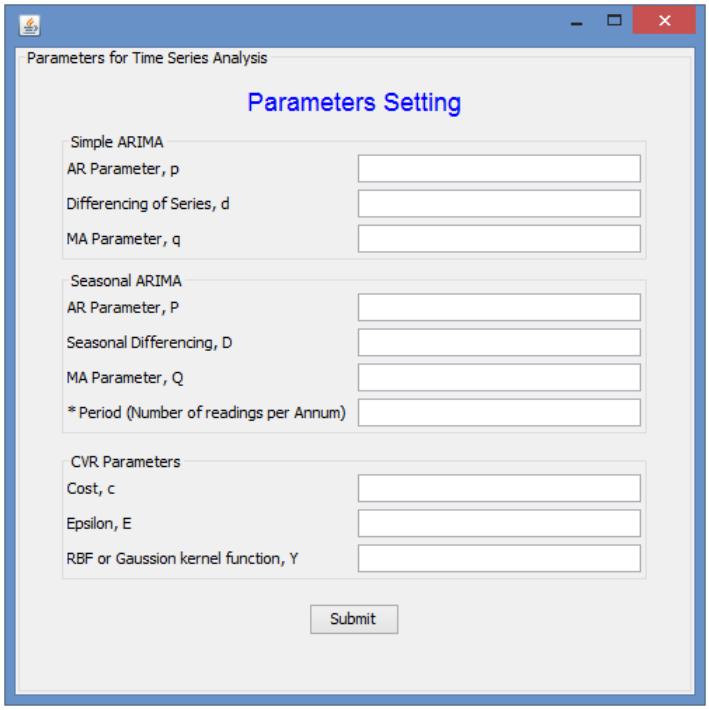

Figure 6.11: Parameter setting window fo spatio-temporal time series analysis

Chapter 7

Conclusion and Future Scope

In this thesis, software for analysis of spatio-temporal data has been designed and developed. For designing the framework, existing frameworks, tools and research papers related to spatio-temporal data warehousing were surveyed. Open source libraries that could be used were explored for the implementation.

On the basis of literature survey, the framework has been designed. It has been designed as two subsystems i.e. Automated and Customized. Automated system is meant for multi-dimensional analysis and in this system user intervention is only once for the selection of query parameters. A Data warehouse has been created for the use of Automated System. For performing all the data mining tasks manually, customized system has been developed, which is an extension of open source data mining tool, WEKA.

Work can be extended for further optimization of automated system and for adding additional functionalities to the customized system.

Bibliography

[1] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages and Computing*, vol. 14, no. 03, 2003.

[2] S. Kisilevieh, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal c1ustering," 2010.

[3] K. V. Rao, "Spatio temporal data mining : Issues , tasks and applications," *International Journal of Computer Science and Engineering Survey (IJCSES)*, vol. 3, 2012.

[4] S. Steiniger and A. J. Hunter, "The 2012 free and open source gis software map a guide to facilitate research, development and adoption," *Department of Geomatics Engineering, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4*, 2012.

[5] V. M. B. Cavalcanti, U. Schiel, and C. de Souza Baptista, "Querying spatio-temporal databases using a visual environment," in *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM, 2006.

[6] D. Birant and A. Kut, "St-dbscan: An algorithm for clustering spatial-temporal data," *Data Knowledge Eng.*, vol. 60, Jan. 2007.

[7] S. Shekhar, C. Lu, R. Liu, and C. Zhou, "Cubeview: a system for traffic data visualization," in *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, 2002.

[8] M. G. de Oliveira and C. de Souza Baptista, "An approach to visualization and clustering-based analysis on spatiotemporal data." *JIDM*, vol. 4, 2013.

[9] X. T. S. C. S. Shekhar, C. T. Lu, "Map cube : A visualisation tool for spatial data warehouses," 2002.

[10] T. Sehn, L. Maria, G. Fonseca, and G. amara, "Computers & geosciences geodma geographic data mining analyst," *Computers and Geosciences*, vol. 57, 2013.

[11] M.-t. Kechadi and M. Bertolotto, "Mining spatio-temporal datasets : Relevance , challenges and current research directions," no. February, 2009.

[12] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung, "Mining, indexing, and querying historical spatiotemporal data," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '04. ACM, 2004.

[13] A. N. M. B. Rashid and A. Hossain, "Challenging issues of spatio-temporal data mining," vol. 3, 2012.

[14] H. Alouaoui, S. Turki, and S. Faiz, "Mining spatiotemporal associations using queries," in *Information Technology and e-Services (ICITeS), 2012 International Conference on*, 2012.

[15] G. Andrienko, D. Malerba, M. May, and M. Teisseire, "Mining spatio-temporal data," *Journal of Intelligent Information Systems*, 2006.

[16] X. yong Hou, L. Han, M. Gao, X. li Bi, and M. ming Zhu, "Application of spatio-temporal data mining and knowledge discovery for detection of vegetation degradation," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, vol. 5, 2010.

[17] J.-w. Lee and Y.-j. Lee, "A knowledge discovery framework for spatiotemporal data mining," vol. 2, 2006.

[18] S. Rathee and A. Yadav, "Survey on spatio-temporal database and data models with relevant features," *International Journal of Scientific and Research Publications IJSRP*, vol. 3, 2013.

[19] J. Jern, Mikael Franzn, "Geoanalytics exploring spatio-temporal and multivariate data," *Proceedings of the information visualizations IEEE*, 2006.