Ain Shams University
Faculty Of Computer
& Information Science
Computer Science Department

Using Spatio-Temporal Relationships For Data Fusion And Mining Of GIS

A Thesis Submitted To The Department Of Computer Science, Faculty Of Computer & Information Science, Ain Shams University, In Partial Fulfillment Of The Requirement For The Degree Of Master Of Computer And Information Sciences

By :

Ahmed Abd El Rahman Abd El Lateef Osman
B.Sc. In Computer Science
Faculty Of Computer & Informatics, Suez Canal University.
Demonstrator At Information System Department,
Faculty Of Computer & Informatics,
Suez Canal University- Ismailia - Egypt

Supervised By:

Prof.Dr./Ahmed Hamad
The dean, Faculty Of Computer & Information Science. October University

Prof .Dr/ Mohammed Hashem
IS Dept., Faculty Of Computer and Information Sciences.
Ain Shams University

Prof.Dr./M.Helmy A.Mahran
The Previous Dean, Faculty Of Computer And Informatics, Suez Canal University, Professor Qassim University – Saudai Arabia

Faculty Of Computer and Information Science
Ain Shams University
Cairo ٢٠٠٧

Abstract

Many different database locations have to be searched to get complete geographical data about an object. The process of collecting geographical data from all resources and stores them in only one location is called Data Fusion. Similarity search algorithm is the most known algorithm for performing this kind of data fusion. This algorithm is based on searching and comparing in all resources sequentially which increase its complexity and wastes a lot of time, effort and space.

In this thesis, we propose a proposed algorithm named tree algorithm to minimize the fusion complexity of GIS data. The proposed algorithm is based on the simplicity of searching and comparing sorted trees. The resulting database can enhance the GIS description and reduce the search complexity.

This research introduces an evaluation of the proposed algorithm and compares it with the similarity search algorithm.

The fusion process may results in large database after combining different databases in only one. Mining these large databases to extract knowledge from it is a complex process. The most popular data mining technique -association rules technique- has proved very efficient in mining theses large databases. However, if the database to be mined has descriptive fields, as is the case in GIS databases, the technique may lead to wrong results. Some GIS fields have a special nature that prevents their categorization (the process which the association rules technique depends on), and thus are not amenable for manipulation by the

normal association rules technique. To overcome this deficiency, we propose in this research a modification to this technique.

The modification is a new algorithm that depends on skipping the categorization steps in the normal association rules technique. The algorithm uses instead some statistical methods, such as the correlation coefficient.

The modified technique is applied to a real GIS database---the demographical database of the USA. When compared with the expected results, the results obtained by the modified technique have proved twice as accurate as those obtained by the normal technique.

To give the results some practical value, they have been used to develop a tiny Decision Support Systems (DSS). Namely, the extracted rules are incorporated in the decision making of GIS-related problems.

Decision support systems play a very important role in our every day life, since no one can live without taking decisions at home, work even in street. These decisions must be based on perfect background and experience and previous studies in order to be valuable and correct. Decisions with no pre-calculation and studies may lead to bad results. Different data have to be searched to get complete view about any object which consumes a lot of time. Data fusion is the process that combines the data found in many different locations in one location so that all required data could be found in one place.

Data mining process is most used to get perfect data and knowledge about some thing, usually in the process of pre-

studying before the decision making process. Association rules technique is the most famous technique for data mining as it generates strong rules from large databases. These generated rules can then be used as pre-study knowledge that help in the decision support process.

In this thesis different GIS databases are fused in only one. Association rules technique is then applied in order to generate strong rules which are then combined in the decision support rules to induct new rules to be used in the decision making process. These extracted rules are then compared to those from experts systems.

Table of Contents

List of Figures

List of Tables

Acknowledgement

In the name of Allah, the Beneficent, the Merciful: "**And whatever of comfort you enjoy, it is from Allah…"** *Al-Nahl (٥٣).* First and foremost, I humbly give my deep thanks to **Allah** for giving me the opportunity and the strength to accomplish this work.

I would like to thank **Prof. / Mohamed Essam Khaleifa** - the dean of the faculty of Computer and Information Sciences, Ain Shams University- for his support. My great thanks to **Prof. / Hamed Nassar,** the dean of the faculty of Computers and Informatics, Suez Canal University, for his encouragement and support.

I would like to express my deep appreciation and thanks to my supervisors: **Prof. / Mohammed Hashim** and **prof. / Ahmed Hamad** and **prof. / Helmy Mahran,** for their great help and encouragement during the execution of this work. I would like to thank **Prof. / A. B. Salem, Prof. / M. E. Ghonemy** for their advice and support**.**

I would like to thank Eng./ **Ahmed Magdy Ghanem,** Eng./ **Amr Mostafa a**nd Eng/ **Mohamed Azab** for their help and advice. Finally, I would like to express my deep thanks to all my family and my colleague friends who were encouraging me strongly during the execution of this work.

Chapter ١

Introduction

The emergence of remote sensing, scientific simulation, telescope scanning, and other survey technologies has dramatically enhanced our capabilities to collect spatio-temporal data, however the explosive growth in data makes the management, analysis and use of data difficult and expensive. In decision support application with spatio-temporal data, it is important to study the temporal relationships of the parameters that influence the decision.

The geographical information system is a computer system capable of capturing, storing, analyzing, and displaying geographically referenced information; that is, data identified according to location. Practitioners also define a GIS as including the procedures, operating personnel, and spatial data that go into the system [١،٤،٦].

Data fusion is the process of combining data and knowledge from different locations in only one location with the aim of maximizing the useful information content. It improves reliability or discriminate capability while offering the opportunity to minimize the data retained.

The fusion process may results in large database after combining different databases in only one. Mining these large databases to extract knowledge from it is a complex process. Data mining combines techniques including statistical analysis, visualization, decision trees, and neural networks to explore large amounts of data and discover relationships and patterns that should solve problems.

In this thesis, we proposed three algorithms. The first one is the tree search algorithm for fusing different GIS databases in only one database. The second one is a modification to the normal association rules technique to mine the descriptive and special nature fields in the GIS databases. The third is a new frame work for making a decision support rules for the GIS data.

The three algorithms are testing using a real data – the GIS data of USA- which obtained from the internet. We choose the USA data because it is highly descriptive, correct and very real.

The three algorithms are implemented using C# programming language.

١,١ Thesis Objective

The first objective of this thesis is to make analysis for geographical data using data fusion by fuse geographical databases founded in diferent locations in only one location with a proposed technique which is then compared to the standard fusion technique. The second objective of this thesis is mining the fused geographical databases whether the special nature geographical data or the traditional geographical data. The last objective is to invest concept description and rule discovery for geographical information system data to aid in the decision making process.

١,٢ Thesis Organization

This thesis is organized as follows:
Chapter two illustrates the different defenitions of the geographical information system, data fusion and the mining of gegraphical data. It produces the different algorithms and techniques which have been previously used in the proceses of fusion and mining of geographical data and its main problems.

Chapter three proposes an improved technique for making data fusion process of geographical data. This technique is called tree search algorithm as it based on storing data in binary tree.

This chapter continues by comparing the results of the tree search algorithm with the results of the old similarity search algorithm after applying them on the same data set. The proposed algorithm provides more suitable results using less number of comparisons and time than the old algorithm.

Chapter four introduces a new algorithm for modifing the numerical association rule mining technique to be suitable for using in mining the special cases geographical data by using some statistical measures such as corelation statistical measure to generate association rules describing these data.

Chapter five a new proposed framework for combining the process of fusion and mining of the geographical data. This framework is based on fusing the geographical data founded in diferent places in one place using the propose algorithm in chapter three then use the numerical association rule algorithm on the fused data in order to generate association rules from these data which are then augmented in the decision support system to output new asoociation rules which can then be used in the decision making process.

Chapter six illustrates the experiments which have been done and their output results. This chapter illustrates the data set which has been used.

This chapter continue by comparing the results of the modified association rule technique which proposed in chapter four with the results of the traditional association rule technique after applying them on the same data set, it is founded that the proposed technique provide a promising results with less miss ratio, wrong ratio and time than the traditional association rule technique.