Applying spatial intelligence for decision support systems

Amira Idrees  
_Faculty of Computers and Information Fayoum University, Fayoum, Egypt_, ami04@fayoum.edu.eg

Mohamed H. Ibrahim  
_Faculty of Computers and Information, Fayoum University, Egypt_, mhi11@fayoum.edu.eg

Recommended Citation

Available at: [https://digitalcommons.aaru.edu.jo/fcij/vol3/iss2/22](https://digitalcommons.aaru.edu.jo/fcij/vol3/iss2/22)
Applying spatial intelligence for decision support systems

Amira M. Idrees \textsuperscript{a}, Mohamed H. Ibrahim \textsuperscript{a,*}, Ahmed I. El Seddawy \textsuperscript{b}

\textsuperscript{a} Faculty of Computers and Information, Fayoum University, Egypt
\textsuperscript{b} Department of Information Systems, Faculty of Business Administration, Arab Academy for Science Technology & Maritime Transport (AASTMT), Cairo, Egypt

Received 2 November 2018; revised 3 November 2018; accepted 5 November 2018
Available online 10 November 2018

Abstract

Data mining is one of the vital techniques that could be applied in different fields such as medical, educational and industrial fields. Extracting patterns from spatial data is very useful to be used for discovering the trends in the data. However, analyzing spatial data is exhaustive due to its details as it is related to locations with a special representation such as longitude and latitude. This paper aims at proposing an approach for applying data mining techniques over spatial data to find trends in the data for decision support. Basic information considering spatial data is presented with presenting the proposed approach aiming to be applied in the Egyptian organizations to prove its applicability.

Copyright © 2018 Faculty of Computers and Information Technology, Future University in Egypt. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Data mining techniques; Spatial OLAP; Spatial data warehouse; GIS

1. Introduction

Geographic Information Systems (GIS) can be defined as the information systems that process spatial data. GIS is able to produce different types of trends in spatial data represented in maps with various quality levels. This variation usually allows the user to find useful relations among the discovered trends [7]. Users could gain numerous benefits from applying GIS by reaching the useful information about different locations and the relations among these locations which is a result of analyzing image data that is represented in maps. Spatial data is presented in different layers (as shown in Fig. 1) forming the map such as the buildings, entities and their boundaries [7].

The remaining structure of the paper is as follows: Section 2 discusses the main definitions that are related to spatial data and business intelligence and the relation between both paradigms. Then Section 3 focusses on the main approaches for applying data mining overs spatial data. Moreover, Section 4 demonstrates the related work while Section 5 presents the proposed approach for integrating spatial data with different sources for developing spatial intelligence for enhancing the findings of the decision support systems. Finally, Section 6 provides the experimental case study discussion and the conclusion is presented in Section 7.

2. Background

This section discusses the main definitions that are related to spatial data and business intelligence and the relation between both paradigms.

2.1. Vector and Raster data

Spatial data can be classified into Vector and Raster data. The following aspects provide a simple definition for both classifications.

Vector data provides representation to the places and entities as either dots or lines as well as polygons. On one hand, the dot provides the exact map place for the entity, the line represents the
relation between these entities and the polygon represents more complex relations. On the other hand, raster data divides the map into adjacent cells which provide a storage capability to each cell targeting to store all details in the map as shown in Fig. 2. These information can be stored with linking to its corresponding part such as the height as shown in Fig. 2. The extracted information is useful for finding patterns among the map components such as medical users as depicted in Fig. 2.

2.2. Spatial data in business intelligence

Data warehouse is the main repository that is essential for Business Intelligence (BI) solutions. Therefore, developing a spatial data warehouse can provide a further step in BI by applying spatial data analysis for supporting business intelligence. OLAP is one of the most common techniques that could be applied on different types of repositories which consequently could lead to adopt decision support systems. OLAP that is based on spatial data provides a concrete recognition for the useful information in different fields such as medical and educational fields [5]. Moreover, providing results to the OLAP query should depends on both spatial data and other types of data for fully reliable results. Additionally, although data mining is traditionally applied in data repositories has also been targeting to provide the useful trends in GIS data.

According to the research by Ref. [4]; building data warehouses over GIS data has been emerged with different researchers such as in, consequently, developing efficient queries through cube relations among spatial data can then provide the required decision support as depicted in Fig. 3.

The cube for spatial data includes many dimensions and measures that match geometry features. These features usually represent the entities defining the members of the field. Spatial dimensions are then integrated with the ordinary type of dimensions for providing different levels of analysis. However, spatial dimensions may depend on both vector data for discrete source and raster data for continuous source. The analysis is then applied according to the provided type. Fig. 4 presents the different classes of spatial data representing the dimensions in the data warehouses.

Many different measures as well as Extract, Transform, and Load (ETL) techniques could also represent spatial data with the same concept. For instance, ETL provides different aggregations on spatial data. This situation leads to an essential process to ensure the applicability, consistency and reliability of the spatial data to be in use. This leads to different issues such as the spatial data integration for decision support. The defined spatial boundaries should be also determined. This obviously clarifies the complexity in building spatial data warehouses. Therefore, one of the most challenging points is to apply the traditional BI techniques over the GIS data. Although these difficulties may hinder the required investigation, however, using cartographic sources may provide a clear identification and contribute in the solution.
3. Spatial data mining

Earlier researchers have investigated applying the data mining techniques over the spatial data. Different targets were on focus such as remote sensing or customers’ support. Recently, different architectures are proposed from different points of view such as in Refs. [9,10] as shown in Fig. 5.

Focusing on data mining tasks and techniques, segmentation provides a classification for all entities according to criteria defined by the decision maker. Dependency analysis is able to predict the classes in the repository. Deviation and analysis are responsible for classification input data into two main classes clustering and classification. The first class determines and describes relationships between data items. The second discovers rules in order to allocate data items to existing classes. Deviation and outlier analysis can classify the user as good or malicious users. Finally, trend decision provides the most suitable result according to the discovered trends.

According to Ref. [15]; analyzing spatial data for finding the relations and monitoring the spatial behavior is a core problem in spatial data mining. Different algorithms are proposed for this direction to extract different types of relations such as k-means, constraint-based, and Bayesian for clustering, FP-Growth for associations, Genetic algorithms, and Naïve Bayes for classifications.

4. Literature review

Some applications are presented by different researchers, for example, a research by Hernandez and his colleagues [11] applied GIS in the medical field for monitoring the disease direction, while another study by ML and her colleagues [16]
focused on applying online Analytical Processing (OLAP) techniques over spatial data in the University of Pittsburgh.

There are different data mining techniques that have been applied. For example, a classification algorithm was presented in Ref. [2] to provide a reliable management in urban problems for systems planning. The spatial data mining succeeded in extracting the required oil relationships but had a lack in the accuracy for the real data.

In addition Ref. [3], proposed a method for recommendations to the user with considering the user's location. The main contribution is that the system learn from the provided results which enriched the systems information, however, these extracted information should be automatically validated before usage in the system. Moreover, a research in Ref. [14] demonstrated a system for decision support that integrated the spatial data with the experts' perspectives in facilitating the street-works with maintaining the activities of all stakeholder, however, the research idea was not implemented as it is demonstrated in a workshop presentation.

In Ref. [13], a DSS was presented for water network management in the governorates. The proposed system was based on spatial data. The proposed system proved the positive impact of the interference of spatial data in providing the decision maker with an effective decision, however, the research highlighted the requirement of integrating other factors for enhancing the system effectiveness which is the focus of the current study. In the same context, a study presented by Ref. [18] presented different techniques for multi-criteria decision support systems and argued the effectiveness of integrating spatial data. The research of Moghadam [18] discussed the effectiveness of applying the proposed solution in the energy sector and presented how this integration will provide a positive impact in decision making.

5. Proposed approach

The proposed approach aims to provide an architectural design for spatial data mining that could be applied over different domains. Integrating spatial database with other sources is a challenging direction as providing analysis for the integrated source may have many obstacles due to the variety of data nature.

5.1. BI and DSS integration

Integrating spatial data techniques with the BI tools generates additional roles for BI for extracting the required patterns or the expected events and situations. This combination provides the ability of the proposed approach to apply both statistical and mining techniques for the required target. In the proposed approach, data analysis methods are enriched through the spatial data aggregations. This phase is challenging phase due to the essential requirement of extracting routines from integrated types of data.

Table 1 provides a brief comparison which highlights the benefits of integrating decision support systems paradigm and business intelligence. It is clear that the integration of both paradigms overcome different sparse functions which is beneficial in the decision making.

5.2. Dealing with heterogeneous data

Aggregations on heterogeneous data in the proposed approach has emerged as a result of the need for many

<table>
<thead>
<tr>
<th>Patterns discovery</th>
<th>Output visualization through spatial view</th>
<th>Statistical analysis</th>
<th>Spatial data manipulation</th>
<th>Integrated data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical methods *</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge discovery systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIS tools</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Proposed approach</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
functions’ requirements in different fields. For example, the discovery of a disorder such as brain tumor can be a challenging task in the data mining field [6]. The high dimensionality of the data in addition to the requirements of different type of data such as rays, statistics, reports and analysis to reach a concrete decision. These types should be all taken into consideration as each type of this data contributes in the analysis result. Including spatial data type raise the challenging in the data heterogeneity analysis in the data mining tasks as it requires specific methods for pre-processing data including filtering and managing inconsistency, to data management, and finally the output visualization. The proposed approach integrates the different data sources into unified subject-oriented structure aiming to enrich the mining task to be able to provide efficient analysis with the required data for accurate results.

5.3. The architectural model for the proposed approach

The proposed approach targets to support the decision makers with a wide view of the current situation from different perspectives due to the different data nature. Applying business intelligence tools over the spatial data warehouse is also one of the main targets that should be considered which consequently provide concrete and accurate decision support. Fig. 6 presents the proposed approach for mining spatial data. The proposed approach aims at adopting BI solutions to be able to be applied over the spatial data warehouse in order to find trends in business. The main contribution in the proposed approach is the possibility of integrating the other sources with spatial data for better decision support recommendations. Fig. 6 illustrates the main architecture of the proposed approach.

The proposed methodology of the research maintain the qualitative approach which is performed by analyzing the current situation and provides the related work of this research represented in the previous proposed approaches of earlier researches. According to the different previous researches, although integrating different types of data is not a recent research point, however, it has been found that few researchers have considered including spatial data as one of these types due to the diversity in the data nature. Moreover, integrating business intelligence solutions for a system with spatial data is also challenging as BI depends mainly on numeric data which highlights the essential integration requirement in the proposed approach.

This research aims to provide a general approach for decision support which is based on different types of data including spatial data. Therefore, the following research plan is generalized while it can be applied in different domains with identifying the general phases according to the domain nature. The research plan is identified in the following phases:

Phase one considers identifying the main focus in the field with gathering the required data, this phase is very critical as the gathered data is considered the main pillar for the success of the proposed methodology. Phase two considers developing the spatial database, integrating the data sources, and building the spatial data warehouse. Phase three focuses on selecting the required data mining model while phase four focuses on defining the dimensions and measures, identifying the key performance indicators, then selecting the analysis tool for spatial data. Phase five targets provide recommendations through different views such as reports and domain map for decision support.

Table 2

<table>
<thead>
<tr>
<th>Object ID</th>
<th>County</th>
<th>Disease rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baltimore City</td>
<td>236.9</td>
</tr>
<tr>
<td>2</td>
<td>Washington</td>
<td>202.5</td>
</tr>
<tr>
<td>3</td>
<td>Wicomico</td>
<td>254.6</td>
</tr>
<tr>
<td>4</td>
<td>Somerset</td>
<td>294.4</td>
</tr>
<tr>
<td>5</td>
<td>Queen Anne's</td>
<td>168.7</td>
</tr>
<tr>
<td>6</td>
<td>St. Mary's</td>
<td>167.8</td>
</tr>
<tr>
<td>7</td>
<td>Montgomery</td>
<td>110.7</td>
</tr>
<tr>
<td>8</td>
<td>Prince George's</td>
<td>172.5</td>
</tr>
<tr>
<td>9</td>
<td>Kent</td>
<td>150.9</td>
</tr>
<tr>
<td>10</td>
<td>Howard</td>
<td>125.2</td>
</tr>
</tbody>
</table>

Source: Ref. [7].

Fig. 6. Structure of the proposed approach for applying spatial data mining.

Fig. 7. Map distribution of heart disease (ESRI).
Source: Ref. [7].
6. Experimental case study

The components of the proposed approach were applied through the suitable tools for each phase. The case study was applied on a benchmark dataset [7] targeting to prove the applicability of the proposed approach. Spatial data and numerical-textual data were downloaded for Heart Disease Mortality rate in Maryland. The main aim of the case study was to demonstrate a clarified map for the heart disease which provides the expected trend of the disease. KPIs were successfully identified which provided comprehensive reports for the disease distribution. Table 2 provides a sample of the data which includes the disease distribution over the territories with demonstrating the distribution in the map. More resources are then provided which included the total population, age distribution, and residential level with respect to the year (Fig. 7).

Reports are divide the given set to have both training and testing Data, the main aim was to generate reports that are matching with the expected output. The system succeeded to predict the disease distribution in 2015 with 85% of accuracy. Fig. 8 presents the distribution of 2012—2015 which clarifies that the distribution of 2014 reached 14% of the total death rate and 15% in 2015 was an increase of 1% which leaves an error rate with the predicted percentage with 15%. However, it is clear that a requirement of more experiments with larger dimensionality would prove the applicability of the proposed approach.

7. Conclusion

While mining heterogeneous data is a challenging field, however, it could provide effective solutions for different sectors. Moreover, including spatial data in the dataset increases the challenge that the approach could face. This paper presented an overview for different algorithms and approaches which are applied on spatial data with demonstrating the issues faced in the field. The paper also proposed an approach that can be applied for spatial data mining in different fields such as medical, educational, and other fields. The main challenges that this research focused on is as follows: integrating spatial data with other data sources, dealing with heterogeneous data in all phases of the proposed approach, and visualizing the results in an interactive method. The proposed approach has been evaluated through a benchmark dataset of heart disease mortality, and succeeded in predicting the mortality rate with an accuracy of 85%. The error percentage was due to the minor dimensions of the attributes which could be overcome with using real dataset. As the proposed approach was applied using supportive tools, therefore, future work can target to build an application with a user interface. Moreover, more experiments should be applied in different fields with higher dimensionality of the dataset.

References

