Applying Spatial Intelligence for Decision Support Systems

Amira M. Idrees
Faculty of Computers and Information, Fayoum University, Egypt

Mohamed H. Ibrahim
Faculty of Computers and Information, Fayoum University, Egypt

Ahmed I. El Seddawy
Arab Academy for Science Technology and Maritime Transport (AASTMT), Cairo, Egypt

ABSTRACT

Data Mining is one of the vital techniques to be applied in different fields such as medical, educational and industrial fields. Extracting patterns from spatial data is very useful to find 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.

Keywords: GIS, Data Mining Techniques, Spatial ETL, Spatial OLAP, Spatial Data Warehouse

1. INTRODUCTION

Geographic Information Systems 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 allow the user to find useful relations among the discovered trends (ESRI, http://support.esri.com/en/knowledgebase/GISDictionary/, 2015). Users gain benefits from applying GIS by reaching information about different locations and the relations among these locations which is related directly by analyzing image data represented in maps. Spatial data is presented in different layers (see figure 1) forming the map such as the buildings, entities and their boundaries (ESRI, http://support.esri.com/en/knowledgebase/GISDictionary/, 2015).

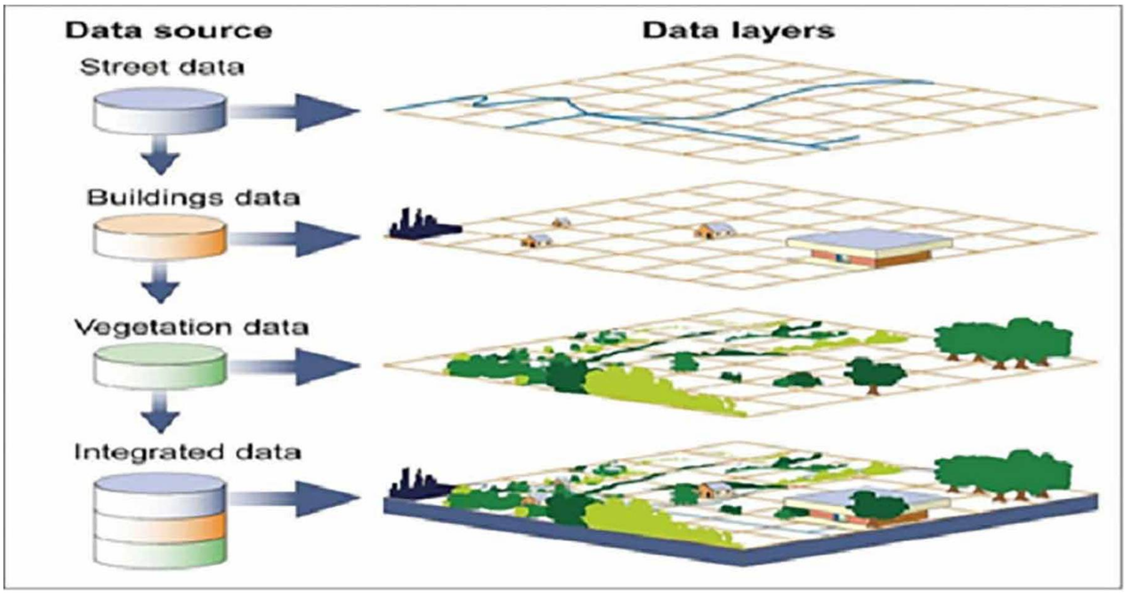

Figure 1 Layers for Spatial Data

1.1 Vector and Raster Data

Spatial data can be classified into Vector and raster data. The following points provides a simple definition for them.

Vector data provides representation to the places and entities as either dots or lines as well as polygons. The dot provide the exact map place for the entity, the line represents the relation between these entities and the polygon represents more complex relation. 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 about this part in the map. Information about this part can be stored with linking it to its corresponding part such as the height. These information is useful for finding patterns among the map components such as medical users (IBEG, 2011), (see figure 2).

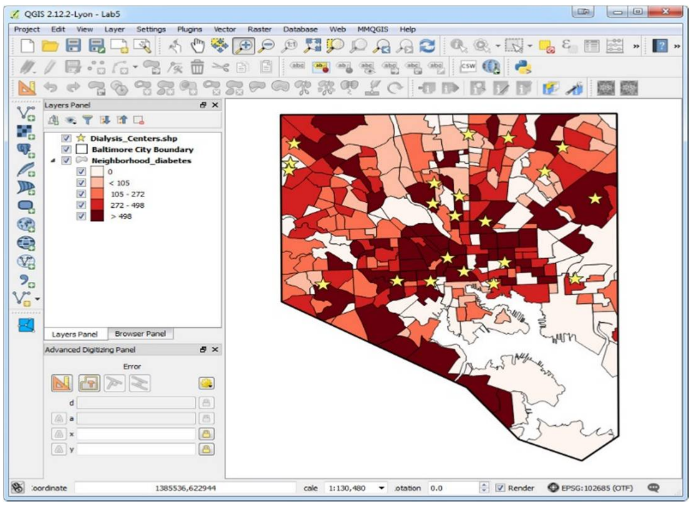

Figure 2 Medical Map

1.2 APPLYING SPATIAL DATA IN BUSINESS INTELLIGENCE

Data warehouse is the main repository that is essential for business intelligence solutions. Therefore, developing a spatial data warehouse can provide a further step in business intelligence by applying spatial data analysis for supporting business intelligence. OLAP techniques can be applied on different types of repositories which consequently could lead to adopt decision support systems. OLAP that is based on spatial data can provide a concrete recognition for the useful information in different fields such as medical and educational fields (Bimonte, S., Pinet, F., Miralles, A., Papajorgji, P., 2014). However, providing results to the OLAP query should depends on both spatial data and other types of data for fully reliable results. Moreover, although data mining is traditionally applied in data repositories has also been targeting to provide the useful trends in GIS data (Mithal, Nipun Garg , Surabhi, 2010).

Building data warehouses over GIS data has been emerged with different researchers such as in (Bédard.Y,Han.J, 2009), consequently, developing efficient queries through cube relations among spatial data can then provide the required decision support (see figure 3).

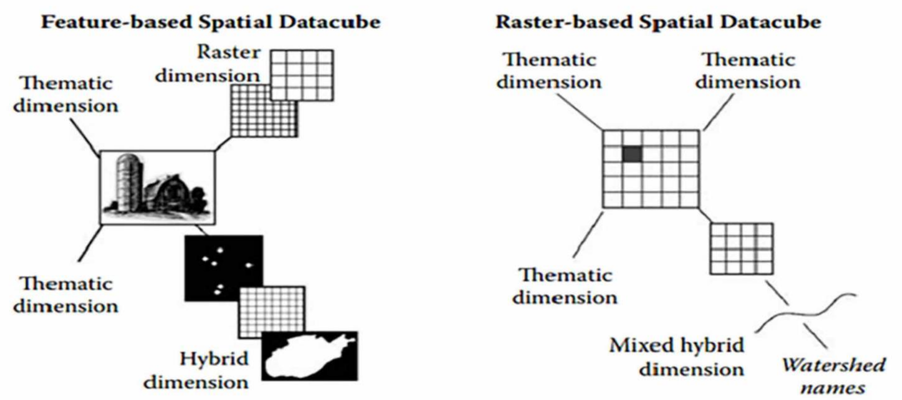

Figure 3: feature-based and raster-based Spatial Cubes, (Bédard.Y,Han.J, 2009)

The cube for spatial data includes 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 to provide different levels of analysis. However, spatial dimensions may depend on vector data as discrete source or raster data as continuous source. The analysis is then applied according to the provided type. Figure 4 presents the different classes of spatial data representing the dimensions in the data warehouses.

On the other hand, measures and ETL can also represent spatial data with the same concept. 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 clarify the complexity in building spatial data warehouses. So, one of the most challenging point 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.

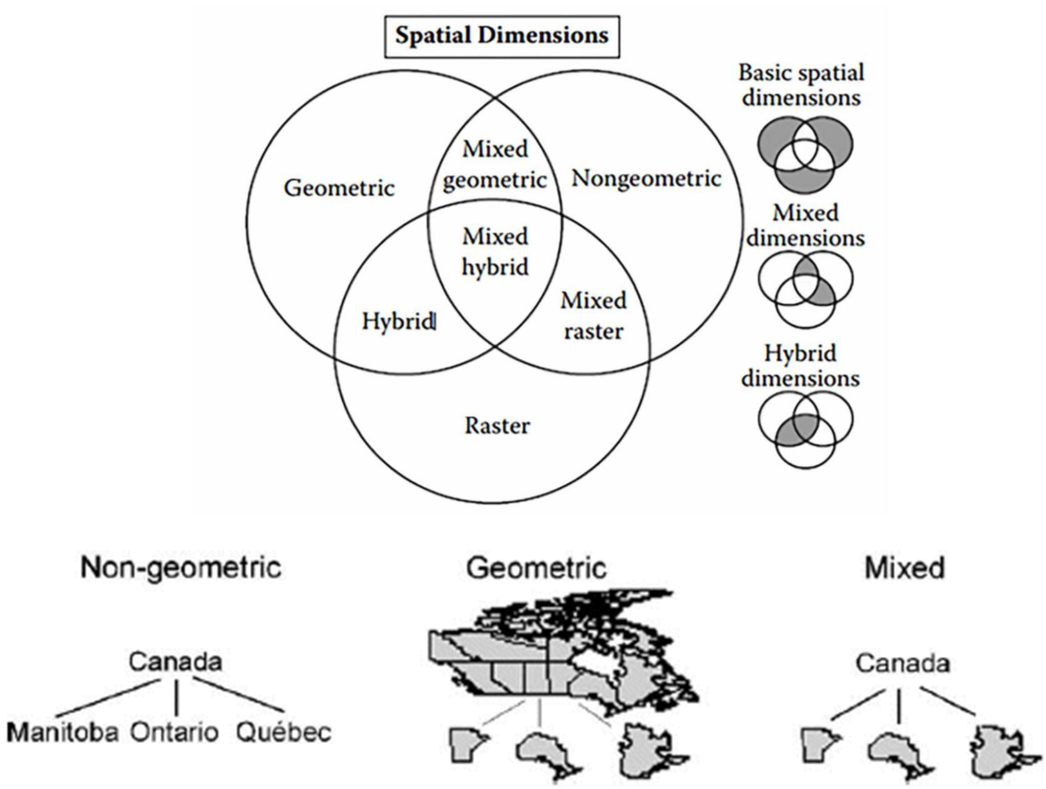

Figure 4: spatial data cubes with different examples of spatial dimensions. (Bédard.Y,Han.J, 2009)

2. DATA MINING OVERS SPATIAL DATA (SPATIAL DATA MINING)

Different researchers have investigated applying the data mining techniques over the spatial data (Diansheng Guo, 2009). Different targets were on focus such as remote sensing or customers' support (Diansheng Guo, 2009). Different architectures are proposed such as in (Hemalatha.M, Naga Saranya.N, 2011), (Jassar, K. K., Dhindsa, K. S, 2015), see figure 5.

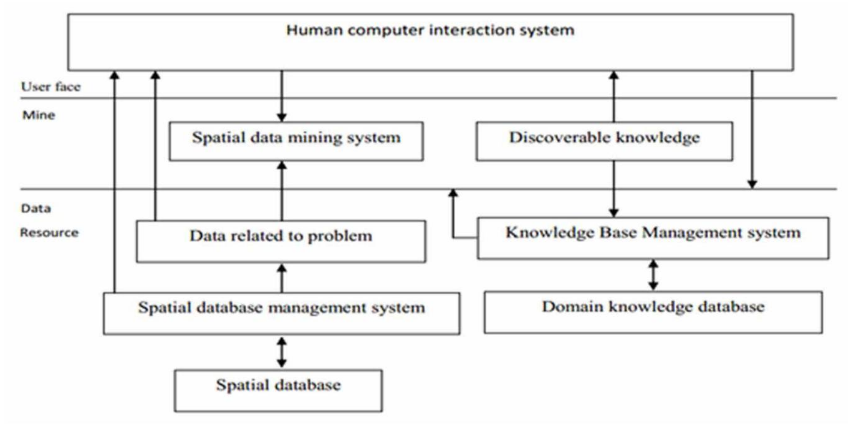

Figure 5: Structure of spatial data mining (Hemalatha.M, Naga Saranya.N, 2011) (Jassar, K. K., Dhindsa, K. S, 2015)

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 is 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 (Martin Ester, 2001), 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, and Naïve Bayes for classifications.

3. LITERATURE REVIEW

Some applications are presented by different researchers, for example, a research by Harnandez and his colleagues (Harnandez, et al. , 2013) applied GIS in the medical field for monitoring the disease direction while another study by ML and her colleagues (ML, et al., 2013) focused on applying OLAP techniques over spatial data in the University of Pittsburgh.

On the other hand, different data mining techniques have been applied. For example, a classification algorithm was presented in (Cai et al., 2014) 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.

Moreover, Bao, Zheng, and Mokbel (2012) proposed a method for recommendations to the user with considering his location. The main point is that the system learn from the provided results which enriched the systems information, however, these extraction information should be automatically validated before usage in the system.

4. PROPOSED APPROACH

The proposed approach aims to provide a conceptual design for spatial data mining that can be applied over different domains especially in medical domain which is the current focus in many nations. Developing spatial medical database is very challenging as well as integrating with other sources. Provide an analysis for the integrated source can have many obstacles due to the variety of data nature. Currently, there has been very minor number of research that focused on applying spatial data mining over medical data although it is a very critical field to consider. 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. The following figure presents a proposed approach for mining spatial data especially in medical field.

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. Figure 6 presents the main architecture in the proposed approach.

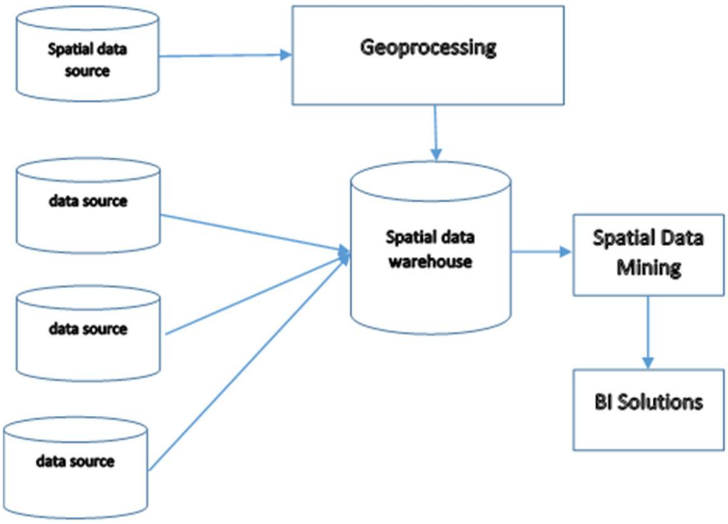

Figure 6: Structure of the proposed approach for applying spatial data mining

The main steps of the proposed framework is to apply geoprocessing over the spatial data sources with integrating the other sources for building a unified structure data warehouse. The following are the main steps for the proposed approach.

1. Collecting spatial data for the entity.
2. Build spatial data source
3. Determine other data sources
4. Apply geoprocessing over the spatial data source
5. Integrate different types of data sources to the spatial data warehouse
6. Apply data mining techniques
7. Apply business intelligence solutions to present the recommendations in spatial interface such as maps.
8. Provide recommendations through different views such as reports and maps

5. EXPERIMENTAL CASE STUDY

In this section, applying a web application for medical sector is performed for proving the applicability of the proposed approach. This application has many tasks including creating, updating, and handling the spatial medical data (see figure 7). Medical maps are created are manipulated with identifying the key medical indicators for decision support. The case was applied on a benchmark dataset (Machine Learning Repository). Due to integrating the spatial data with other numerical data, the application provided higher accurate percentage with 7% of accuracy level. This percentage was evaluated by exerts who compared the produced key indicators generated from the developed web application before and after the data integration.

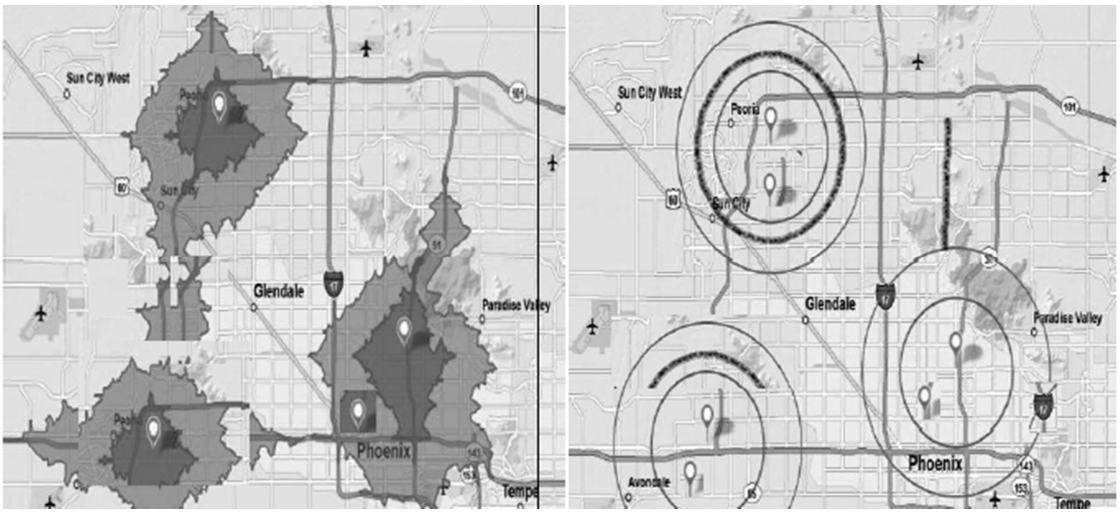

Figure 7: Derived Queries

CONCLUSION

This paper demonstrated basic information of spatial data, spatial data warehouse, spatial data mining, and spatial OLAP. The 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 presented a proposed approach that can be applied for spatial data mining in different fields such as medical, educational, and other fields. Due to the experts' evaluation, the proposed approach enhance the key indicators identification accuracy. However, future work can target to provide more experiments in different fields.

REFERENCES

Cai, Y., Wang, X., Hu, K. & Dong, M. (2014). A data mining approach to finding relationships between reservoir properties and oil production for CHOPS. Computers & Geosciences Vol. 73.

Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. F. (2012). A survey on recommendations in location-based social networks. GeoInformatica, V(212)

Bédard.Y,Han.J. (2009). Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery. Taylor & Francis Group.

Bimonte, S., Pinet, F., Miralles, A., Papajorgji, P. (2014). Spatial data warehouses and SOLAP. GeoInformatica, pp. 269-272.

Diansheng Guo, J. M. (2009). Spatial data mining and geographic knowledge discovery—An introduction. United States.

ESRI. (2012).

ESRI. (2015). http://support.esri.com/en/knowledgebase/GISDictionary/. Retrieved 2015, from http://support.esri.com.

Harvey, M.,Han, J. (2001). Geographic data mining and knowledge discovery: An overview. In Geographic Data Mining and Knowledge Discovery.

Hemalatha.M, Naga Saranya.N. (2011). A Recent Survey on Knowledge Discovery in Spatial Data. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2.

(2011). IBEG.

Jassar, K. K., Dhindsa, K. S. (2015). Comparative Study of Spatial Data Mining Technique. International Journal of Computer Applications.

Jaime Harnandez, Ignacia Nunez, Antonella Bacigalupo, and Pedro E. Cattan, (2013). Modeling spatial distribution of Chagas disease vectors using environmental variables and people's knowledge, Geography.

Martin Ester, H.-p. K. (2001). Algorithms and applications for spatial data mining. France: Geographic data mining and knowledge discovery .

Mithal, Nipun Garg , Surabhi. (2010). Spatial Data warehouses

Miranda ML, Casper M, Tootoo J, Schieb L., (2013). Putting chronic disease on the map: building GIS capacity in state and local health departments. Previous Chronic Disease.

Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/Statlog+(Landsat+Satellite)