Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.9, 2012

Applying Association Rules and Co-location Techniques on Geospatial Web Services

Eman ElAmir[1*] Osman Hegazy[2] Mohamed NourEldien[3] Amr H. Ali[4]

1. Faculty of Computers and Information, Cairo University, PO box 12613.1 ,Giza, Egypt
2. Faculty of Computers and Information, Cairo University, PO box 12613.1 ,Giza, Egypt
3. Faculty of Computers and Information, Cairo University, PO box 12613.1 ,Giza, Egypt
4. Faculty of Engineering, Benha University, Surveying Engineering Department, Egypt

* E-mail of the corresponding author: eman.elamir@gmail.com

Abstract

Most contemporary GIS have only very basic spatial analysis and data mining functionality and many are confined to analysis that involves comparing maps and descriptive statistical displays like histograms or pie charts. Emerging Web standards promise a network of heterogeneous yet interoperable Web Services. Web Services would greatly simplify the development of many kinds of data integration and knowledge management applications. Geospatial data mining describes the combination of two key market intelligence software tools: Geographical Information Systems and Data Mining Systems.

This research aims to develop a Spatial Data Mining web service it uses rule association techniques and correlation methods to explore results of huge amounts of data generated from crises management integrated applications developed. It integrates between traffic systems, medical services systems, civil defense and state of the art Geographic Information Systems and Data Mining Systems functionality in an open, highly extensible, internet-enabled plug-in architecture. The Interoperability of geospatial data previously focus just on data formats and standards. The recent popularity and adoption of the Internet and Web Services has provided a new means of interoperability for geospatial information not just for exchanging data but for analyzing these data during exchange. An integrated, user friendly Spatial Data Mining System available on the internet via a web service offers exciting new possibilities for spatial decision making and geographical research to a wide range of potential users.

Keywords: Spatial Data Mining, Rule Association, Co-location, Web Services, Geospatial Data

1. Introduction

As the applications of spatial databases grow, spatial data mining has been developed to discover interesting, previously unknown patterns in spatial databases. The demand for processing massive spatial data is increasing rapidly, particularly in science (GIS, ecology, etc.), engineering (i.e., traffic control) and industry (i.e., GPS navigation and mobile/sensor network). Consequently, it is necessary to develop efficient spatial data mining techniques to help domain experts discover useful knowledge from the given databases.

A co-location pattern is a group of spatial features/events that are frequently co-located in the same region. For example, human cases of West Nile Virus often occur in the regions with poor mosquito control and the presence of birds. For co-location pattern mining, the studies often emphasize the equal participation of every spatial feature. As a result, interesting patterns involving events with substantially different frequency cannot be captured. (Huang et al. 2006)The objective of co-location pattern mining is to find frequently co-located subsets of spatial features. For Example, a co-location "{traffic jam, police, car accident}" means that a traffic jam, police, and a car accident frequently occur in a nearby region. To capture the concept of "nearby", the concept of user-specified neighbor-set was introduced. A neighbor-set L is a set of instances such that all pair wise locations in L are neighbors. (Huang et al. 2006)Spatial co-location pattern mining is similar to association mining. A spatial association rule is a rule of the form "A B" where A and B are sets of predicates and some of which are spatial ones. In a large database many association relationships may exist but some may occur rarely or may not hold in most cases.(Kumar et al. 2012)

There are two main approaches in data mining the first is based on quantitative reasoning, which computes distance relationships during the frequent set generation. It has the advantage of not requiring the definition of a reference object, but has some general drawbacks such as deal only with points and do not consider non-spatial attributes. For spatial objects represented by lines or polygons, their centroid is extracted. This process may lose information and generate non-real patterns (e.g. the Mississippi River intersects many states as a multi-line object, but is far from the same states by considering its centroid).(Science 2006)

The second approach is based on qualitative spatial reasoning and considers distance and topological relationships between a reference geographic object type and a set of relevant feature types represented by any geometric primitive (e.g. points, lines, and polygons). Spatial association rules are pruned using minimum support and confidence, but rule interestingness is not addressed. All spatial relationships are extracted from the database and transformed into a deductive relational database, which is non-trivial for real applications. Well known rules are partially pruned with declarative bias defined by the user, but in post processing steps, after the generation of frequent sets and association rules. (Science 2006)

Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence, which is the conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules. Association analysis is commonly used for market basket analysis. For example, it could be useful for a video store manager to know what movies are often rented together or if there is a relationship between renting a certain type of movies and buying popcorn or pop. The discovered association rules are of the form: P -> Q [s,c], where P and Q are conjunctions of attribute value-pairs, and s (for support) is the probability that P and Q appear together in a transaction and c (for confidence) is the conditional probability that Q appears in a transaction when P is present. For example, the hypothetic association rules:

Rent Type(X, "game") AND Age(X, "13-19") -> Buys(X, "pop") [s=2%, c=55%]

This would indicate that 2% of the transactions considered are of customers aged between 13 and 19 who are renting a game and buying a pop, and that there is a certainty of 55% that teenage customers who rent a game also buy pop.

In this paper we propose a model of discovering rule association geospatial data mining solving this disparate data and systems problem, and then we showcase this model through a geospatial data mining system exposed through web services, applied to crisis management of road accidents.

The rest of the paper is structured as follows. In section 2 we explain why we used spatial data mining and the methods used of the research. Section 3 describes the case study, rule association of spatial data mining techniques used to correlate data of crises management concerned with road accidents. In section 4 we provide the research results of the case study, present and discuss the results. Section 5 gives the conclusions and future work.

2. Why Spatial Data Mining?

Since data mining becomes a cornerstone in many data processing and knowledge discovery seekers, and spatial data are getting huge and need to be deeply searched to get any simple information or any valuable knowledge. (Ng 1994)It became an urgent need to identify spatial patterns and spatial objects those are potential generators of patterns, to identify information relevant for explaining the spatial pattern, hiding irrelevant information and to present the information in a way that is intuitive and supports further analysis. Because spatial data is not identically distributed in the space, data properties are location dependent, the local trends can sometimes contradict the global trends and spatial data is heterogeneous, there was an urgent need for spatial data mining.(Patterns 2010)

Spatial data mining can answer some critical questions, characterize or predict effects of human activity on the environment. In our geospatial data mining web service we will apply rule association techniques, co-location mining, and then we will apply regression analysis on the results to prove the results of our predictions are relevant to data revealed by regression this is illustrated at Figure 1.

2.1 Spatial Association Rules

Towards achieving a higher level of efficiency and competitiveness in manufacturing operations, the Association rule-based approaches focus on the creation of transactions over space so that a priority like algorithm can be used. Transactions over space can use a reference-feature centric approach or a data-partition approach. The reference feature centric model is based on the choice of a reference spatial feature and is relevant to application domains focusing on a specific Boolean spatial feature, e.g., incidence of cancer. Domain scientists are interested in finding the co-locations of other task relevant features (e.g., asbestos) to the reference feature.(Lin & Lim 2009)

In the real world, spatial classes usually have refined subclasses. For example, the “road” class can be refined to “highway,” “interstate,” “street,” etc. Spatial relationships may also be refined. For example, a “close to” relationship can be refined into “intersects,” “touches,” “within,” etc. Given concept hierarchies on spatial classes and/or spatial relationships, multiple-level association rules mining can also be applied. [6] The mining of spatial association rules takes the following steps: 1- Extract spatial objects from the spatial database, 2- For each object of the reference class, discover its specified spatial relationship with spatial objects of other classes, 3- Mine spatial association rules.(Lin & Lim 2009)

Our proposed model was applied on a case study that uses the proposed model to integrate between agencies involved in the response to road accidents and enable them to exchange the data needed to support each of them in addressing and handling road accidents. The mining of spatial association rules requires five inputs: 1) a spatial database, 2) a reference class (“accidents” , 3) a set of classes whose relationships with the reference class are interesting to users (“road,” “buildings,” “fire stations,” “traffic units,” and “medical services” , 4) a spatial relationship type (“close to”

relationship between the reference class and other classes , 5) a minimum support threshold and a minimum confidence threshold.

2.2 Spatial Co-location Mining

Co-location rules are models to infer the presence of Boolean spatial features in the neighborhood of instances of other Boolean spatial features. Co-location rule discovery is a process to identify co-location patterns from large spatial datasets with a large number of Boolean features. The spatial co-location rule discovery problem looks similar to, but, in fact, is very different from the association rule mining problem because of the lack of transactions.(Shekhar et al. 2004)
Given a finite set of Boolean spatial features and their instances, spatial co-location mining seeks to discover a set of features whose instances are frequently co-located in close proximity. Spatial co-locations represent a subset of features whose instances are frequently located together in spatial neighborhoods. Spatial co-location patterns may yield important insights for many applications.(Yoo et al. 2006)
The co-location mining procedures starts with spatial datasets of our interest civil defense, medical services, traffic agency and the roads departments and work with conjunction Main application. We refer to main application here as any interested decision making agency will have ability to deal with all those entities having the authority to view, analyze and access data and synchronize requests and knowledge between different entities helping them perform their work in a better way.

In our case study we use Boolean spatial features of road accident, fire cases this predicts the relationship of loss of more lives when we had no coordination between civil defense, and medical Services when moving to solve accident and fire happened in nearby location. We try to materialize the relationship of the neighborhood keeping in mind that no relation is missed, duplicate relations are minimized and the cost is kept as cheap as possible, then candidate co-locations are generated. The filtration phase then applied and then co-location neighborhood rules are generated, we can back to step of materializing relationships to repeat the procedures again and generate new rules.(Yoo et al. 2006)

2.3 Regression Analysis

This is well known oldest statistical technique that is utilized by the data mining users. Essentially, regression makes use of a dataset to develop a mathematical formula which fits the data. So whenever you want to use the results for predicting future behavioral patterns, all you need to do is just take the new data, and apply it to the formula that has been developed, and you will get your prediction. In our case study we use geographically weighted regression model to be used of predictions of results. Geographically weighted regression is a local regression model where one regression equation is calculated for each feature and relationships are allowed to vary across. The aim of using this model is to prove the concept that our used model using rule associations and correlation techniques have valid assumptions when we reveal same results using regression modeling. The features participating and sharing in the model will include all features from civil defense, traffic agency and medical services including all variables of influence from all of them.

3. Web Services Data mining for Crises Management in "Alexandria, Egypt"

The case-study was applied on a real-life case of one of the most crowded cities in Egypt with high population density: the area of Attareen in Manshia district, Alexandria City. The case applied on different disasters happing at same time, road accidents along with fire at nearby accident location.(ElAmir et al. 2012)
Accident analysis plays an important part in the strategy to reduce road accidents. In the past, the main analysis tools available to the road safety engineer were paper maps allied to databases like Excel spread sheets. Accidents were identified on the map using road segments or an area-based location. This was very time consuming process and lacked accuracy.
However, nowadays Geographical Information Systems have revolutionized the whole framework of accident processing and analysis. In diagnosing the cause of traffic accidents, traffic engineers, planners and decision makers often need as much information as possible before deciding the appropriate countermeasure.
In our analysis apart from producing the most accurate prediction from statistical figures, the need to visualize information geographically is always essential. Also data need to be prepared to be able to detect relations between different datasets of interest we used ArcGIS applications to prepare needed vector layers, raster images , tables and information to build geo-database at each interested entity and a centralized geo-database at the Main application.(ElAmir et al. 2012) Figure 4
Our solution is based on collecting information by police officers via handheld devices per accident location and uploads this data to the accident application to Main Application through web services which can access web services of medical that includes ambulance dispatch application to complete the accident information and upload it for reviewing and decision making by roads department to analyze the roads that needs enhancement to avoid more and similar traffic accidents. The solution provides an intelligent crises management for traffic accidents and fire fighting solutions that provides the Ministry of interior and civil defense with all related geographic information for traffic accidents and fires that usually occurs in same areas. Web Servers used for hosting custom developed web applications that will be

consumed directly by the end users and web services that will be consumed by the handheld applications for data synchronization activities.

In Figure 5 we show the interaction between Main Web Service, Civil Defense, Traffic and medical Web Services. The scenario illustrated is as follows:

1- An Accident happens and fire at same location. Traffic web service and civil defense inform Main Web Service

2- Main web service traffic, and civil defense informing them that there are conflicts in their way to solve the issue and at same time request needed information for both of them from medical web service

3- Medical web services send needed information to Main Web Service

4- Main Web Service synchronizes gained information from medical and sends it to both traffic and civil defense.

The centralized geodatabase at main application storing the required vector and images layers for ongoing accidents registration and processing activities along with medical services and civil defense daily data updates. The Main application not only directing requests between interested entities it also synchronize needed knowledge and hidden relationships extracted by geospatial data mining applying rules association techniques and co-location mining (Ester et al. 2001) on those accumulated huge amount of data processed though our proposed model. The rules revealed from our proposal model

Acc(x) -> fire(y) an accident happen at same time with a fire

Acc(x) ->Acc (y) two accidents happen in same time in a nearby location

Acc(x) ->Road cut(y) an accident followed by a Road cut

Acc (Loc, “X”) AND Acc (Loc,”Y”) -> Loss (P) AND Loss ($)

The probability of loss in people and money related to locations of the two accidents on same time

Acc (Loc, “X”) AND Fire (Loc,”Y”) -> Loss (P) AND Loss ($)

The probability of loss in people and money when an accident happens at same time with a fire

Acc (Loc, “X”) AND Road cut (Loc,”Y”) -> Loss (P) AND Loss ($)

The probability of loss in people and money an accident followed by a Road cut

This case study indicate that 9% of the road accidents considered are of speed limits in rush hour, and who are causing another accident to happen, and that there is a certainty of 55% of the accidents cause loss of people and money are the ones who followed by road cut.

After all predictions and results revealed by our system we apply the most commonly used statistical method which is geography weighted regression.

Input feature class: Roads

- Dependent variable: Road Accidents
- Explanatory variables: Speed Limits, Road Conditions, Age, Number of lanes
- Output feature class: GWR Results
- Kernel type: ADAPTIVE (This means the spatial context used to solve each local regression analysis is a function of a specified number of neighbors. Where feature distribution is dense, the spatial context is smaller; where feature distribution is sparse, spatial context is larger.)(Huang et al. 2003)

Because the dependent variable is number of accidents, the explanatory variables might be things like speed limits, roads conditions, number of injured people, percentage of damage, etc.—the things that could possibly contribute to a high accidents volume.

Geography Weighted Regression (GWR) also outputs coefficient surfaces, which offer a way to visualize where you have strong and weak relationships between the dependent and independent variables. Coefficient surfaces may help a community focus remediation efforts where they will be most effective.

4. Discussion of Results

After accumulating huge amount of data, we performed geo-data mining analysis and discover the hidden information and relations between data; we had applied the rule association techniques to help solve problems later and have better scenarios solving road accident problems and have better understanding of real cause of the problem. Also we used co-location techniques in conjunction with rules association over web environment.

Considering literature and technical publications on geospatial information, infrastructures, technology, standards and interoperability, data mining, GIS, this research has explored and investigated the potential of applying Web Services as a new approach to a geospatial data infrastructure. The research used specifications endorsed by the Open Geospatial Consortium (OGC) and the access, visualization, evaluation and discovery of geospatial data with Conformity with the following measurements in dealing with geographical data:

- WFS/WMS :data: Vector Data
- WMS :digital images: Images

- WCS/WMS :satellite images: Raster Imagery

The display of digital maps and Images and digital elevation models using methods compatible with OGC and display tools users should be enabled to determine coordinates of any point on the map, measure distances and angles between points, measure areas and track positions using GPS devices. The system provides ability to amend the method of cartography, according to the Scale-dependent styling ability to deal with different types of data (documents, movies, pictures, any electronic content) .Extraction of reports and statistics in the cartographic form (black spots - the distribution of the accidents on the road - the numbers of the incidents classified according to location, history, type of accident...) to reach recommendations along with conducting data and geographic inquiries of accidents.Figure 7
Using application prototypes, this research has reviewed and assessed past trends in geographic information interoperability, and explored Web Services as a new approach to interoperability, in the context of Geographical Data Infrastructure in addition to data mining techniques for intelligent data analysis. We had developed an interoperable web service that can handle intelligent analysis of geospatial data from different heterogeneous sources to ease of use for decision makers. We applied the case study on disaster management of road accidents; the sample date explored in Alexandria, Egypt. We have created a multivariate geography weighted regression model that can helped gain further insight into the data, their relationships, how they vary across space, and what may be coming in the future.

5. Conclusion

One analysis can often lead into others, parameters for tools may change, criteria for analyses can evolve, or you may want to perform additional visual analyses on the results to make them more meaningful or easier to interpret. Data mining in such huge temporal changing geo-databases is a promising area of future research. Decision makers are interested in learning hidden rules describing the process of crises management handling and it's affecting factors on other disasters. Spatial co-locations represent a subset of features which are frequently located together in geographical space.(Ng 1994) Co-location pattern discovery presented with rules association techniques over web services was a challenge since spatial objects are frequently changing and embedded in a continuous space, whereas classical data is often discrete on tables and spreadsheets. (Wang et al. 1997)Geography weighted regression predicted accidents and loss volumes for the future. This allows agencies not only to anticipate accidents and fire response and medical needed demands for the future, but also provides a way to measure how effective remediation policies are. Our model expand the benefits from GIS, data mining, web services, statistical methods from being an isolated islands working alone to an integrated intelligent solution yielding new opportunities to serve the planet in a proper way.

Suggested future works include gaining the knowledge from diverse data sources without being able to access it directly, data will be collected in a data warehouse virtually and then hidden knowledge from those different sources will be extracted, and feedback analysis is only available for the interested users to aid in decision making. While technology being rapidly changing instead of hosting the data at each organization servers we aim moving all data into cloud computing. Cloud computing furnishes technological capabilities commonly maintained off premises that are delivered on demand as a service through standard internet protocol. Using cloud computing we can reduce the capacity and handle issues of privacy and authentication to determine who have the ability to get the data and analysis results, ability to apply rest of geospatial data mining techniques such as clustering , spatial outliers at the cloud (Kouyoumjian 2011) will be investigated to get all benefits of technology updates to our proposed integration as future work.

References

ElAmir, E. et al., 2012. Integrating Web Services with Geospatial Data Mining Disaster Management for Road Accidents. GeoInformatica- An International Journal, 1(2), pp.1–11.

Ester, M., Kriegel, H. & Sander, J., 2001. Algorithms and Applications for Spatial Data Mining. Geographic Data Mining Knowledge Discovery, pp.1–32.

Huang, Y., Pei, J. & Xiong, H., 2006. Mining Co-Location Patterns with Rare Events from Spatial Data Sets. GeoInformatica, 10(3), pp.239–260. Available at: http://www.springerlink.com/index/10.1007/s10707-006-9827-8 [Accessed September 4, 2012].

Huang, Y., Xiong, H. & Shekhar, S., 2003. Mining Confident Co-location Rules without A Support Threshold. New York, pp.0–4.

Kouyoumjian, V., 2011. GIS in the cloud. In Esri ArcNews. Redlands USA: Esri.

Kumar, G.K., P.Premchand, P.P. & Gopal, T.V., 2012. Mining Of Spatial Co-location Pattern from Spatial Datasets. International Journal of Computer Applications, 42(21), pp.25–30. Available at: http://research.ijcaonline.org/volume42/number21/pxc3877994.pdf.

Lin, Z. & Lim, S., 2009. Optimal candidate generation in spatial co-location mining. Proceedings of the 2009 ACM

symposium on Applied Computing - SAC '09, p.1441. Available at: http://portal.acm.org/citation.cfm?doid=1529282.1529604.

Ng, R.T., 1994. Efficient and Effective Clustering Data Mining Methods for Spatial. Proceedings of the International Conference on Very Large Data Bases, 129(7057), pp.144–155.

Patterns, C.S., 2010. VisTracks TM Marketing Research & Analytics The Cloud – Do More with Less IT Cost Breaking Free from Legacy. , pp.1–2.

Science, C., 2006. Enhancing Spatial Association Rule Mining in Geographic Databases. Knowledge Creation Diffusion Utilization, Vania Bogo(October), pp.1982–1989.

Shekhar, S., Zhang, P. & Huang, Y., 2004. Trends in Spatial Data Mining. In S. Shekar, ed. Science. Minneapolis: AAAI/MIT, p. 363.

Wang, W., Yang, J. & Muntz, R., 1997. STING : A Statistical Information Grid Approach to Spatial Data Mining 1 Introduction 2 Related Work. , pp.1–18.

Yoo, J.S., Member, S. & Shekhar, S., 2006. A Joinless Approach for Mining Spatial Colocation Patterns. IEEE Transactions on Knowledge and Data Engineering, 18(10), pp.1323–1337.

Eman ElAmir is a GIS Technical Consultant at Esri Northeast Africa Egypt since 2003. She is *Ph.D.* student at faculty of computers and information, Cairo University, Egypt from 2007 till present. The major fields of study are geographical information systems, data mining and web Services.

Osman Mohamed Hegazy is Professor of Information (Data Engineering) since 1992 at Information Systems, Cairo University, Egypt. Bachelor degree in Electronic Engineering, Faculty of Engineering, Cairo University, Egypt, 1964, Ph.D. in Computer Engineering, university of Leicester, England 1977. The major fields of study are data engineering, cloud computing and computer/ human interface.

Mohamed NourEldien is an Associate professor at Information Systems, faculty of computers and information Cairo University, Egypt Since 2002. *Ph.D.* in GIS and Information systems from Louis Pasture University, France, 2001.The major fields of study geographical information systems, data mining, Ecommerce and E-Business.

Amr H. Ali is an Associate professor at surveying engineering, faculty of Engineering- Shoubra Benha University, Egypt *Ph.D. in Geodesy & Surveying from Technical University Graz, Graz, Austria*. The major fields of study are geographical information systems, data mining, geodesy and surveying.

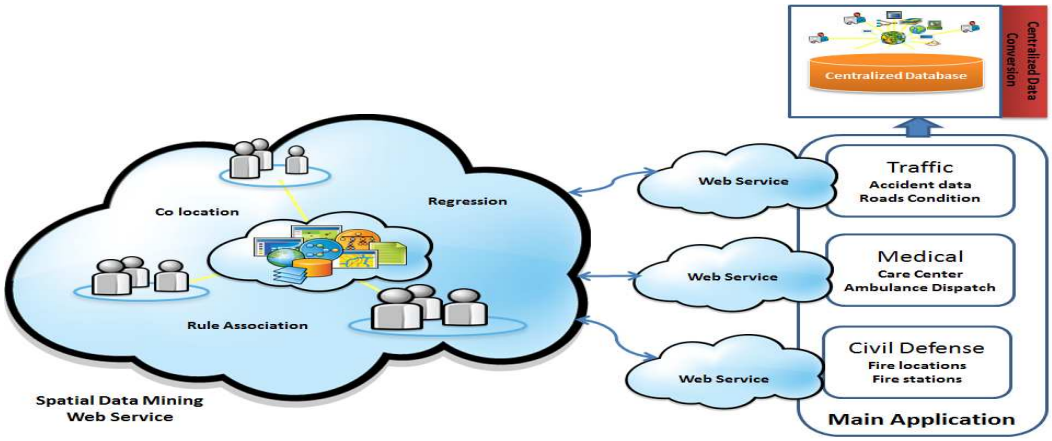

Figure 1Geospatial data mining proposed analysis

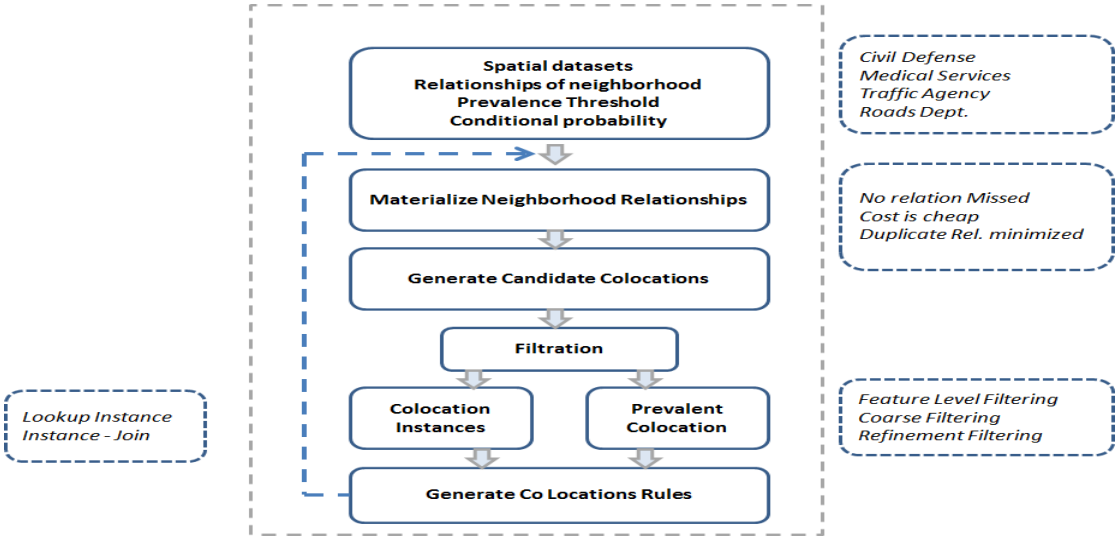

Figure 2 Co-location Mining Procedures

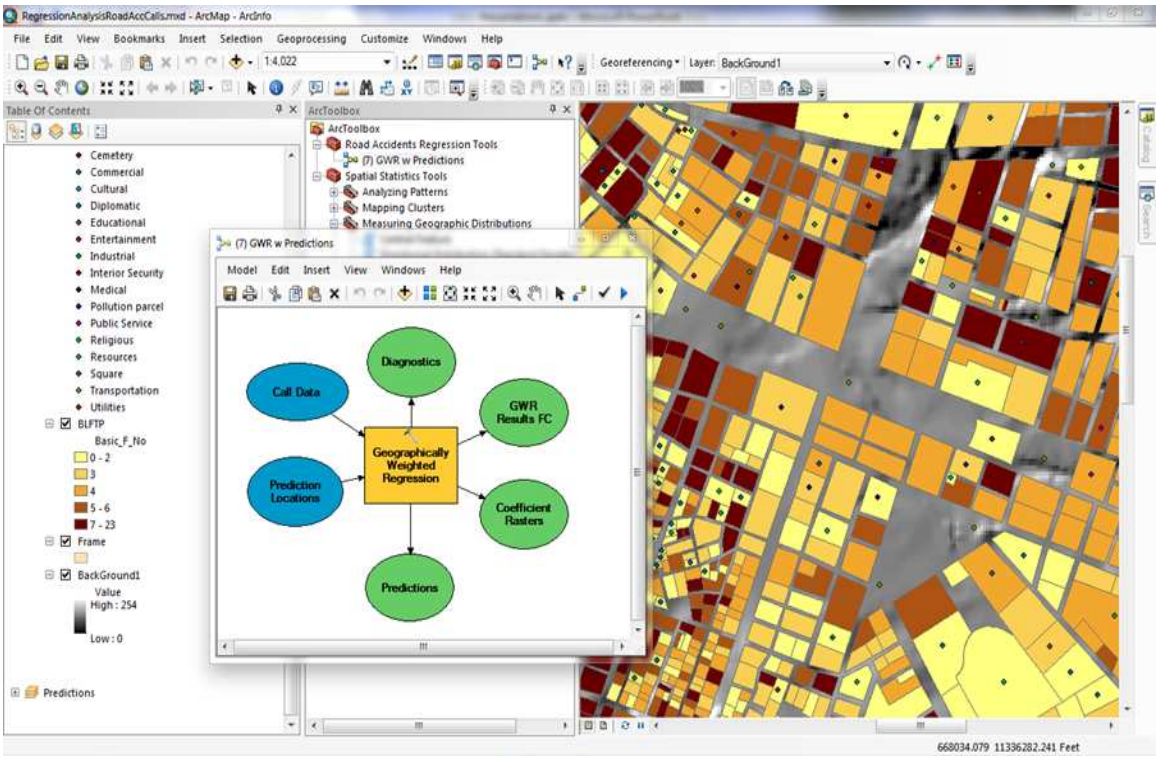

Figure 3 Geography Weighted Regression

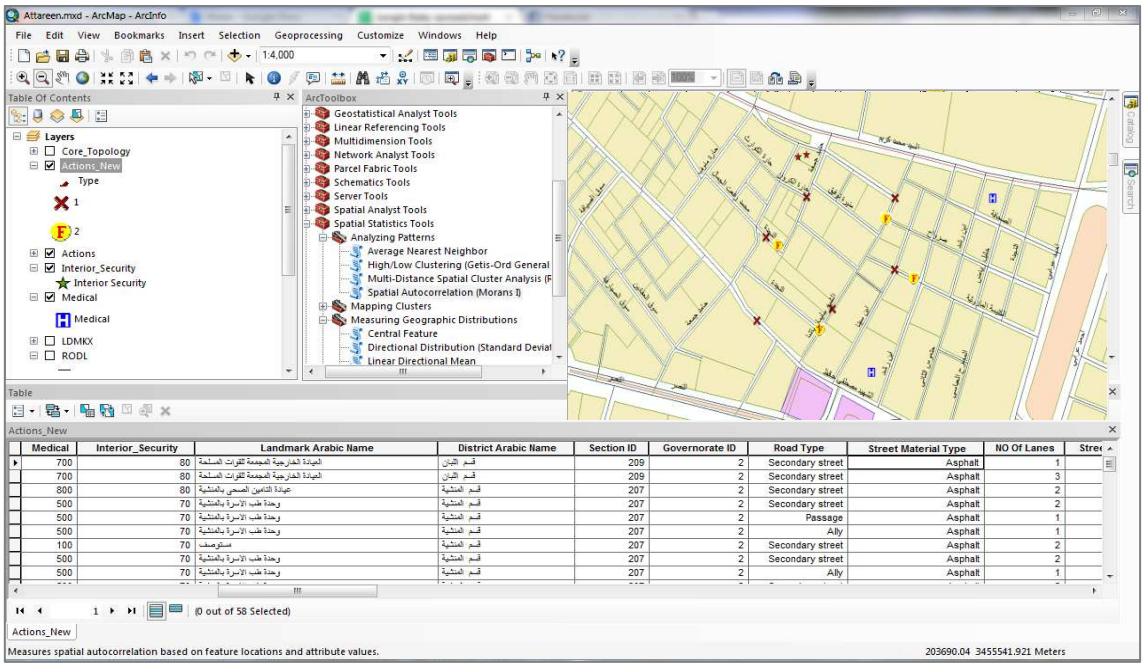

Figure 4 Datasets Preparation for Spatial Association Rules

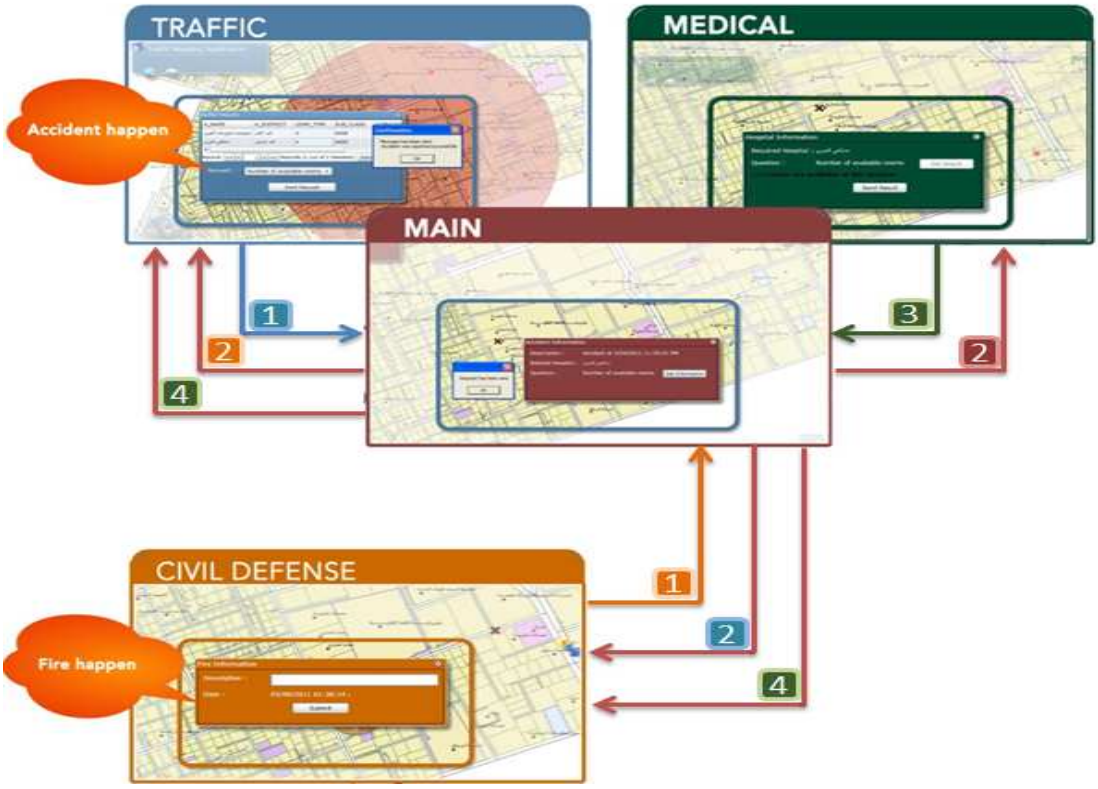

Figure 5 Main, Civil Defense, Traffic and Medical Web Services

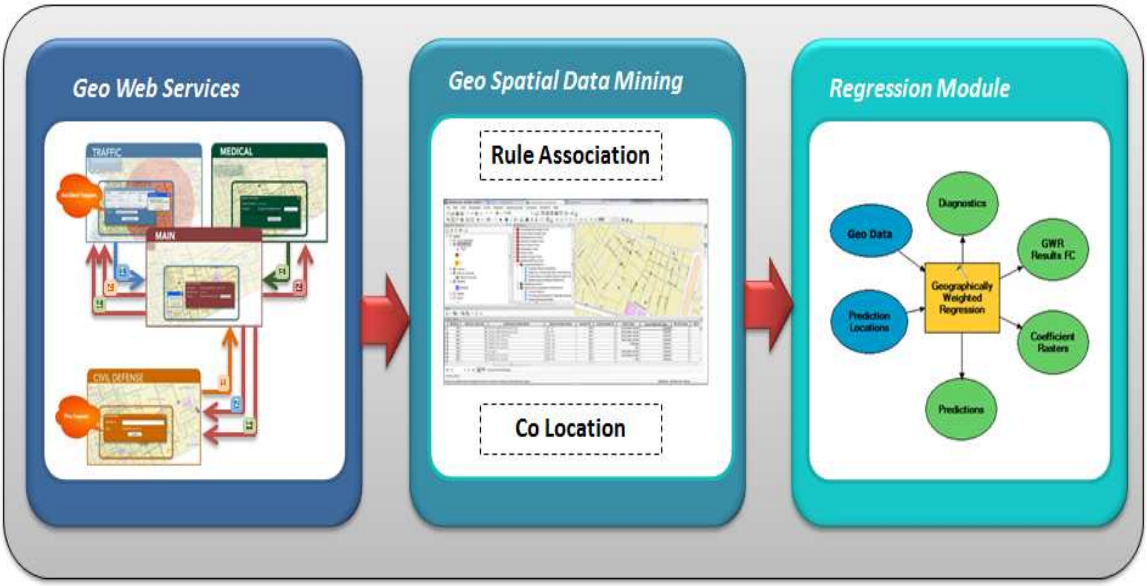

Figure 6 Geo Web Services, Geospatial Data Mining and Regression Module

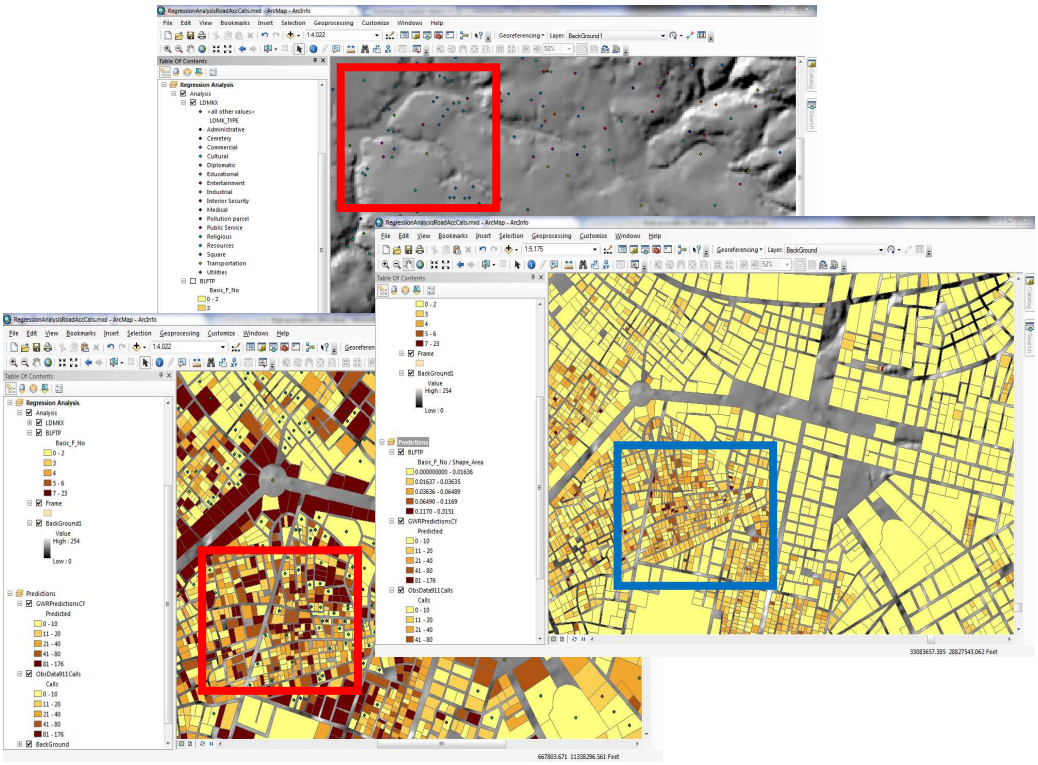

Figure 7 Regression Module Results of Predicted Areas