A Case Study of Geovisual Analytics Approach for Exploring Unknown Patterns over Spatio-Temporal Datasets

Monisha Biswas
Lecturer, Dept. of CSE
Port City International
University
Chittagong, Bangladesh

Md. Monjur-Ul-Hasan
Asst. Professor, Dept. of CSE
Chittagong University of
Engineering & Technology
Chittagong, Bangladesh

Anupia Barua
Lecturer, Dept. of CSE
Port City International
University
Chittagong, Bangladesh

ABSTRACT

An effective web-enabled geovisual analytics approach, that has been developed based on the principles behind visual analytics to interactively visualize and solve spatio-temporal related visualization problems for multivariate data through a visual discovery and analytical reasoning is introduced in this paper. This approach makes extensive use of spatial and temporal filtering, interactive highlighting and color encoding that allows analysts to perform exploratory analysis to make sense of the data A case study is performed illustrating the benefits of using this proposed approach by finding unknown information for effective decision making in real world problem.

General Terms

Visual Analytics.

Keywords

Geovisual analytics (GVA), Spatio-temporal data, Analytical reasoning, Visual variable.

1. INTRODUCTION

Visual analytics(VA)is the science of analytical reasoning which emphasizes the importance of the analyst who applies reasoning techniques and human judgments to gain insight into data. The goals of a visual analytics system are to derive insights from the data by detecting expected and unexpected information and to effectively communicate assessments for action [1]. Geovisual Analytics (GVA) can be considered as a subarea of visual analytics with its specific focus on space and time posing specific research problems. Geovisual analytics is an emerging interdisciplinary field that looks for ways to provide computer support for solving space-related decision problems through enhancing human capabilities to analyze, envision, reason, and deliberate [2]. Geovisual analytics puts the focus on space and time and takes experience from GIS(geographic information system) and geovisualization.

Analyzing time series data with spatial decision support means computerized assistance to people in the development, evaluation and selection of proper policies, plans, scenarios, projects or interventions[3]. Need to analysis and make decisions based on geospatial and temporal data for exploring unknown information or patterns, often in time-critical situations, demands an integrated and automatic approach through visual exploration . In this work, a geovisual analytics system is proposed which is inspired by the visual analytics mantra suggested by Keim et al..[4] which is an extension of Shneiderman's information seeking mantra[5]. The proposed geovisual analytics approach will visualize differences between two data sets of multivariate spatio-temporal data and by interacting with it, analysts will be able to find out interesting patterns or trends between the two datasets.

To address the problem of finding differences between two datasets, a computational method is proposed for automatically extracting difference percentage from two datasets by defining three indexes for a specific area. For each spatial range or point data, which is the unit of this analysis procedure, three index values are calculated based on the difference percentage between two datasets. Then that area is represented on a map using a color encoding procedure. A time graph view is presented where the percentage of each data unit value is shown over temporal range. At a time, detailed visual information of several entities of each data unit is also provided by a bar chart method.

2. RELATED WORK

Visual exploration of spatio-temporal and multivariate data has been the subject of many research papers. N.Andrienko and G. Andrienko illustrated several motivating approaches in their earlier papers [6,7] for the visual exploration of spatial and temporal data. Many research papers and efforts [1,2,8,9] focused on the development of visualization toolkits and systems for different types of data and different purposes. In most recent work, a geo-visual analytics tool (called GTdiff) [8],was developed to enable users to explore the geotemporal differences in the data through an intuitive and interactive interface. This system makes an extensive use of spatial and temporal filtering, data aggregation using spatial and temporal binning, and visualizations. Although there are many tools for spatio-temporal multivariate data, it has still been a challenge and time consuming process for the analysts to gain insight for finding any unknown information by interacting with these tools. Through this research work, an analyst friendly geovisual analytics approach will be proposed for visualizing, exploring and communicating spatio-temporal and multivariate data through web environments in order to find out useful information for effective decision making and analytical reasoning.

3. PROPOSED GVA APPROACH

The system architecture diagram view in order to provide a complete geovisual analytics approach is shown in Fig.1.
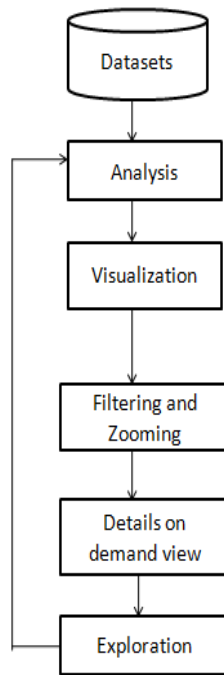
Fig. 1 System architecture of proposed GVA system.

To facilitate various visual analysis tasks, the approach is designed to provide a wide range of visualization methods including information visualization methods and geovisualization methods. Main visualization methods supported in this geovisual analytics approach include:

- a map component with a background map layer (Google map in this case) and a glyph layer (a pie chart glyph layer in this case) for the visualization and analysis of geographic data;
- a bar chart for showing comparison among several attributes ;
- a time graph for the visualization and analysis of time-series data;
- a details on demand view with line chart visualization method.

Multiple linked views is an important feature of visual analytics systems for visual data exploration and analysis processes. To facilitate the development of multiple linked view applications, the system is designed to support a number of linking mechanisms as follows :

- linking based on selection, i.e. if an item is selected in one view, it will be treated as a selected item in all linked views;
- linking based on time animation, i.e. all linked views share the same time animation controller which controls the time period during an animation process; when the time period is changed, all linked views are updated.

4. IMPLEMENTATION

For wide spread use and easy access to internet users, the prototype system is developed as a web-based application. JavaScript is used as programming platform as it is supported by all modern web browsers. Google Map is used for online map services. Datasets are fallen into the map from mysql data source file.

 The data sets used in this prototype were from SSC (Secondary School Certificate)and HSC (Higher Secondary School Certificate) board examination results of some students over the time period of five years where their associated colleges from which they passed their HSC exam are treated as geospatial point data. Each of these college locations is recorded as geographical point data (called visual variable) by their latitude and longitude values. The attributes of each recorded point data that were used within the prototype system are shown in Table 1.

Table 1. List of attributes used for recording each geographical point data

Attributes	Description
LON	Longitude
LAT	Latitude
CID	College ID
CN	College Name

For simplicity and avoiding visual clutter, HSC results of only science group students for each recorded college (geospatial point) are considered in this system. For comparing HSC results of the students with their SSC results, for exploring unknown information between these two public examination results, their corresponding SSC results with same attributes are also considered in the dataset. The attributes which we have considered in SSC and HSC examination result datasets are given in Table 2.

Table 2. List of attributes for SSC and HSC examination result datasets

Attributes	Description
CID	College Id
SID	Student Id
BANG	Obtained GPA in Bengali
ENG	Obtained GPA in English
MATH	Obtained GPA in Math
PHYS	Obtained GPA in Physics
CHEM	Obtained GPA in Chemistry

The proposed system first reads raw data from database, and for each geographical point data i.e , for each college in datasets, a visual variable is marked with a pie chart glyph layer on a background map layer . A computational color encoding method is followed for coloring each visual variable as in Fig. 2.
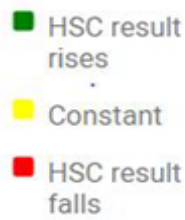
Fig. 2 Color schemes used for each visual variable.

After analysis of data and following computational method for calculating differences between SSC and HSC exam results of each student of each college stored in database, the prototype system initiates the overview visualization with a time graph view and map view. A overall representation of the proposed geovisual analytics system where all multiple visual components are linked together in a single frame is given in Fig. 3.
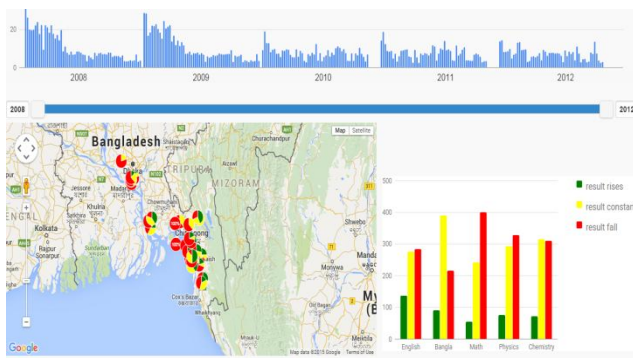

Fig. 3 Overview representation of proposed GVA system.

The visual components of this prototype are: geovisualization (map view), time graph view, bar chart view, a temporal window slider for filtering. Now, the prototype becomes ready for analysts to explore the data sets and a sequence of actions will take place after each of the user interactions.

5. CASE STUDY

In this section, performance of the proposed system in achieving the objectives defined at the beginning of this paper is evaluated. For experimental test, HSC result data from a number of colleges located in Chittagong division are considered for a time period of five years (2008 to 2012). And SSC result data of these students are also recorded in the database system. For avoiding visual clutter, it is not possible to consider data of all colleges located in Chittagong division, but enough of them are considered in the analysis test.

Now from the time graph view of Fig.3, analysts may observe that maximum number of colleges gave a poor performance in 2010 to 2012. So analysts may be interested to change the temporal range in order to observe college performances in the map view within selected temporal range. By applying temporal filtering operation, the map view and bar chart view is updated within the selected temporal range which is shown in Fig.4. Here bar chart view is also updated in order to explore in which subject student's results rise, fail or remain constant according to selected temporal range. For example, at this point analysts may easily find out that most of the student's results (within 2010 to 2012) fall more in Mathematics, than Physics, than Chemistry ,than English and Bengali.

At this stage, analysts may want to remove some visual clutter in the map representation and so zooming operation can be applied on the map view. For example, analysts may be interested to observe the data details of the colleges that are located at Boalkhali region in Chittagong division , as they gave constantly a poor performance. So after zooming into that region, other visual components such as time graph and bar chart view of the prototype interface are also updated as shown in Fig.5. After zooming, it is clearly visible that all of the colleges of that region (five colleges in map view) gave constantly a poor performance on the selected temporal range(2010 to 2012).And in this situation, analysts can explore more such different information of these limited colleges (which are now in the map view) by changing temporal period again.

In order to support analysts' understanding, detail inspection and evaluation of individual colleges are required. This can be done by interactively selecting one of these colleges. One college can be selected by the analysts that are of particular interest . Upon making such selections, a new pop-up window is opened for that selected college where its performance is shown in more detail within the selected temporal range as shown in Fig.6. After zooming into Boalkhali region, analysts may want to inspect each college individually for clear perception. In Fig.6, data details of a selected college (North Satkania Jafar Ahmed College is selected in this case) is shown, where the performance of that college could be analyzed individually through line chart view within selected temporal range.
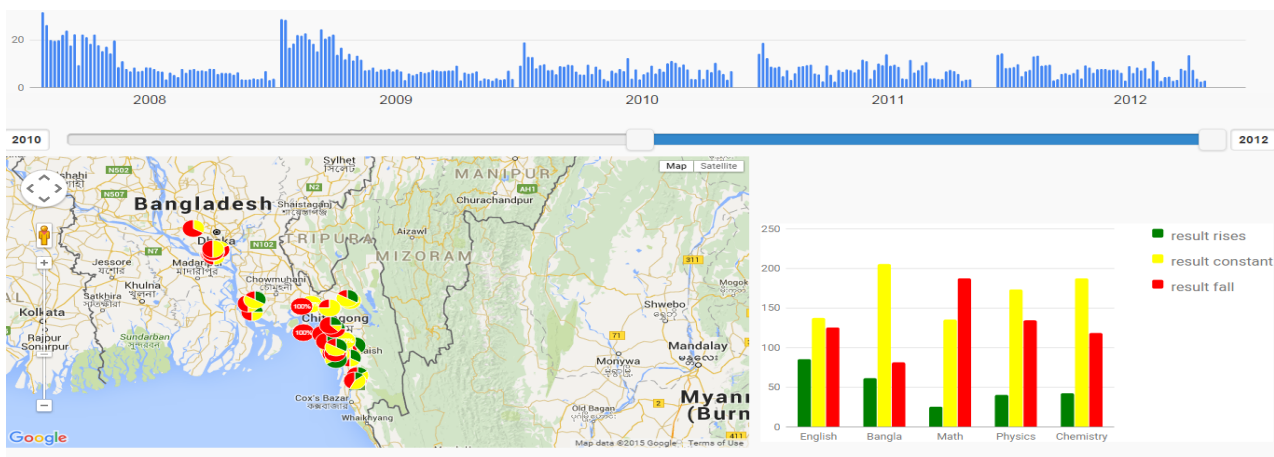

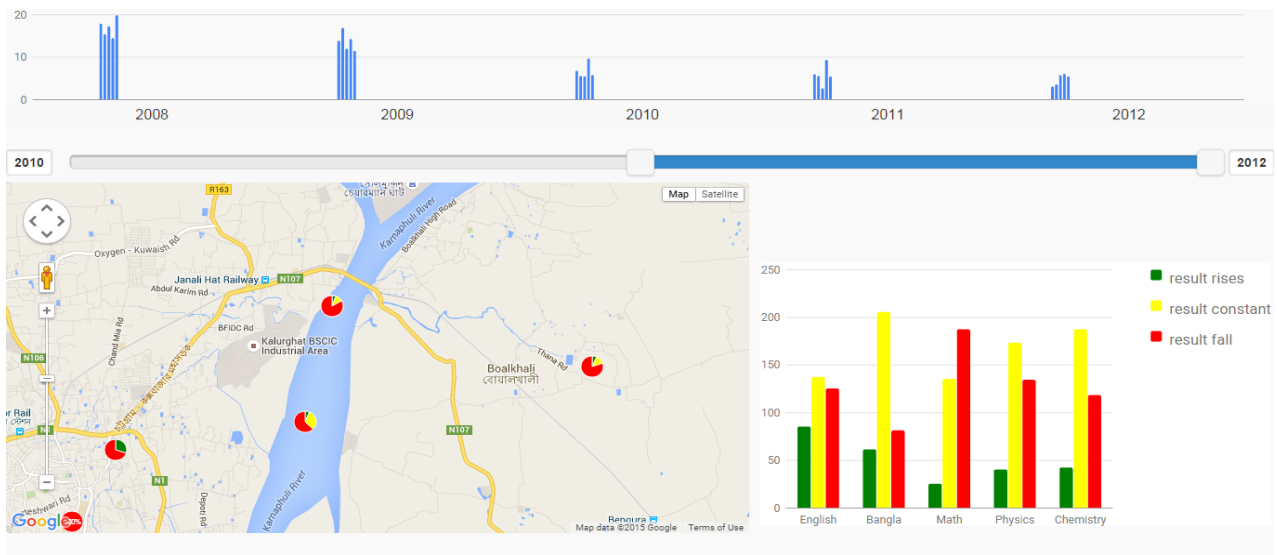**Fig . 4 Updated view (map view and bar chart view) after changing temporal range (2010 to 2012).**

Fig. 5 Data details without any visual clutter after zooming into Boalkhali region
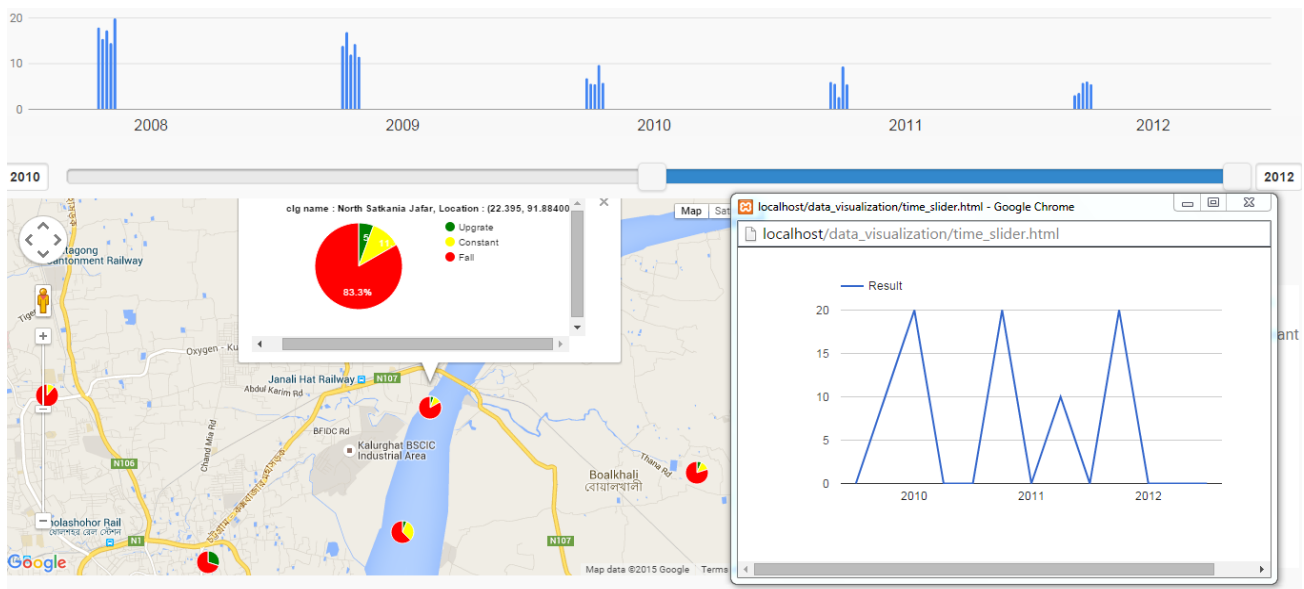

Fig. 6 Result details of a visual variable (college) within a selected temporal range.

This analysis process described above may allow analysts to discover some of the insights and provides evidence for further investigation. Now again the analysts may want to start the analysis process with overview representation or may wish to broaden the temporal range in order to gain another insights. In addition to verifying known facts of the data, new knowledge can also be discovered from the data sets using the geovisual analytics approach presented in this prototype system.

6. CONCLUSION

A web-enabled geovisual analytics approach presented in this thesis has been inspired by the visual analytics mantra proposed by Keim et al..[4] which is an extension of Shneiderman's information seeking mantra[5]. The case study performed with the proposed system shows that the interactive features of this prototype system support exploration of known and unknown information or patterns and allow

analysts to perform analytic reasoning about the underlying behavior of that patterns . The highly interactive features of this system allow analysts to easily focus on the spatial and temporal extents of interest. Although there are many visualizations and interaction techniques supported in this work, users and analysts may want to have more visualizations and interaction techniques to do their jobs better and faster. This is one of the focuses of the future work. As the prototype system is implemented and experimented by considering only small amount of colleges located in Chittagong division, a case study should perform in future after collecting   huge amount of data from education board of the country, in order to make effective decisions in education system. However, there are still many interesting future fields where the introduction of visual analytics would improve the efficiency and performance of daily work.

7. REFERENCES

[1] Hasan,M.M.U. 2015. A Study of Geovisual Analytics for Exploring Event Anomalies Over Multiple Geo-spatial Datasets. Masters thesis, Memorial University of Newfoundland, Newfoundland, Canada.

[2] Ho,Q.V. 2013. Architecture and Applications of a Geovisual Analytics Framework. Linkoping University Electronic Press.

[3] Ware,C. 2004. Information Visualization : Perception for Design. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[4] Keim,D.A., Mansmann, F.,Oelke, D. and Ziegler, H. 2008. Visual Analytics: Combining Automated Discovery with Interactive Visualizations. In Proceedings of the International Conference on Discovery Science, pp.2– 14.

[5] Ahlberg, C. and Shneiderman, B. 1994. Visual Information Seeking : Tight Coupling of Dynamic Query Filters with Starfield Displays.In Proceedings of ACM CHI94 Conference on Human Factors in Computing Systems.

[6] Andrienko , N. and Andrienko , G. 2005. Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[7] Andrienko,G. and Andrienko,N. 2004.Interactive Visual Tools to Explore Spatio Temporal Variation. In Proceedings of the working conference on Advanced visual interfaces, AVI '04, pp. 417-420, New York, NY,USA.

[8] Hoeber,O., Wilson , G. and Harding, S.2011.Exploring Geo-Temporal Differences Using GTdiff .Visualization Symposium,IEEE Pacific, pp.139-146.

[9] Andrienko,G., Andrienko, N. and Wrobel,S.2007. Visual Analytics Tools for Analysis of Movement Data. ACM SIGKDD Explorations Newsletter, 9(2):38–46.