

Spatial Clustering in SOLAP Systems to Enhance Map Visualization

Ricardo Silva, Universidade Nova de Lisboa, Portugal

João Moura-Pires, Universidade Nova de Lisboa, Portugal

Maribel Yasmina Santos, Universidade do Minho, Portugal

ABSTRACT

The emergence of the SOLAP concept supports map visualization for improving data analysis, enhancing the decision making process. However, in this environment, maps can easily become cluttered losing the benefits that triggered the appearance of this concept. In order to overcome this problem, a post-processing model is proposed, which relies on Geovisual Analytics principles. Namely, it takes advantage from the user interaction and the spatial clustering approach in order to reduce the number of elements to be visualized when this number is inadequate to a proper map analysis. Moreover, a novel heuristic to identify the threshold value from which the clusters must be generated was developed. The proposed post-processing model takes into account the query performed, i.e., the number of spatial attributes, the number of spatial dimensions, and the type of spatial objects selected from dimensions. The results obtained so far show: (i) the novel approach to support queries with two spatial attributes from different dimensions allows useful analysis; (ii) the proposed post-processing model is very effective in maintaining a map suitable to the user's cognitive process; and, (iii) the heuristic proposed provide the user participation in the clustering process, in a user-friendly way.

Keywords: Data Visualization, DBSCAN, Geovisual Analytics, SOLAP, Spatial Clustering

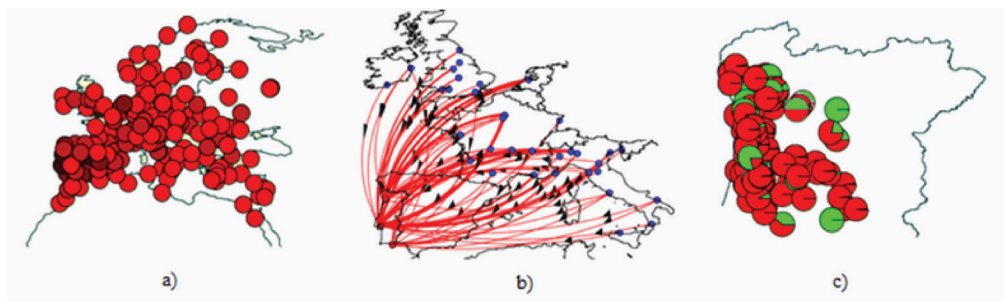
CONTEXT AND MOTIVATION

Most OLAP applications are focused on textual data and numerical measures even though available studies have concluded that 80% of data is associated with spatial information (Bédard, Rivest, & Proulx, 2006). Consequently, the integration of spatial data within the multidimensional model was envisaged.

Rivest, Bédard, and March (2001) defined the Spatial OLAP (SOLAP) concept as “*a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays*”. SOLAP systems allow the integration of spatial data, included either in dimensions (spatial dimensions) or in fact tables (spatial measures), in OLAP ap-

DOI: 10.4018/jdwm.2012040102

Figure 1. Examples of cluttered maps



plications enabling cartographic displays on those applications (Rivest, Bédard, & March, 2001). This way, thematic maps are produced by using the members of spatial dimensions and the numerical measures, combining them with the visual variables (Rivest, Bédard, Proulx, Nadeau, Hubert, & Pastor, 2005).

In this new environment for the analysis of spatial data, several benefits have been mentioned from thematic maps visualization enabled by SOLAP systems. Among them we found better and faster global perception of query results, and the possibility to discover correlations between phenomena, as detailed in Bédard, Rivest, and Proulx (2006).

For the users, it is worth to mention that there is a difference in the results that can be analyzed by an OLAP user compared to a SOLAP user. A typical result of the former involves one to two dozen lines with the aggregated data. However, the latter may involve hundreds of lines if the user is interested in the analysis of data at a lower level of granularity (e.g., customer level). If this is the case, and depending on the geographical distribution and the spatial objects' representation, a thematic map can easily become cluttered and hard to analyze, as illustrated in the examples displayed in Figure 1.

Regarding the example shown in Figure 1a, the points are airports locations and the brightness associated to them is given by the value of a numerical measure. For this map, the user is unable to compare the numerical measure value of different airports as many

markers are hidden due to an overlapping between them. The same happens in Figure 1c, where we have pie charts with information of several facilities in a geographical location. In the Figure 1b, the blue markers correspond to the departure airports; the red markers are the arrival airports; and, the value of the numerical measure is given by the arc's width. Again, the map comprehension/visualization becomes harder due to the high number of arcs and their superposition. Moreover, this last example is a result based on a novel approach to support queries which has involved two spatial attributes from different spatial dimensions (in the query select clause). This novel approach is proposed in this work and further detailed in subsequent sections of this paper.

The presented examples result from the interaction with SOLAP+ system (Jorge, 2009; Silva, 2010), which is afterwards described in this paper. In all of them, the map showed to be inadequate to the user visualization and corresponding analysis.

Thus, in a SOLAP system it is necessary to control the number of results returned to the user in order to maintain the usefulness of maps in the decision making process. To maintain the benefits that come from map visualization, this paper extends our previous work (Silva, Moura-Pires, & Santos, 2011) by proposing a post-processing stage that relies on Geovisual Analytics principles (Andrienko et al., 2007; Keim, Andrienko, Fekete, Görg, Kohlhammer, & Melançon, 2008; Andrienko, Keim, MacEachren, & Wrobel, 2011), combining the

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/spatial-clustering-solap-systems-enhance/65572?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Library Science, Information Studies, and Education, InfoSci-Select, InfoSci-Knowledge Discovery, Information Management, and Storage eJournal Collection, InfoSci-Surveillance, Security, and Defense eJournal Collection, InfoSci-Journal Disciplines Engineering, Natural, and Physical Science, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology, InfoSci-Select. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Partially Supervised Classification: Based on Weighted Unlabeled Samples Support Vector Machine

Zhigang Liu, Wenzhong Shi, Deren Li and Qianqing Qin (2006). *International Journal of Data Warehousing and Mining* (pp. 42-56).

www.igi-global.com/article/partially-supervised-classification/1770?camid=4v1a

Connectionist and Evolutionary Models for Learning, Discovering and Forecasting Software Effort

Parag C. Pendharkar and Girish Subramanian (2003). *Managing Data Mining Technologies in Organizations: Techniques and Applications* (pp. 250-276).

www.igi-global.com/chapter/connectionist-evolutionary-models-learning-discovering/25770?camid=4v1a

Cloud-Based Intelligent DSS Design for Emergency Professionals

Shah J. Miah (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 991-1003).

www.igi-global.com/chapter/cloud-based-intelligent-dss-design/73480?camid=4v1a

Technique for Transformation of Data From RDB to XML Then to RDF

Kaleem Razzaq Malik and Tauqir Ahmad (2017). *Web Semantics for Textual and Visual Information Retrieval* (pp. 70-91).

www.igi-global.com/chapter/technique-for-transformation-of-data-from-rdb-to-xml-then-to-rdf/178367?camid=4v1a