Position statement: Privacy Issues in Geospatial Visual Analytics

Gennady Andrienko and Natalia Andrienko

University of Bonn and Fraunhofer Institute IAIS

Abstract

Visual and interactive techniques can pose specific challenges to personal privacy by enabling a human analyst to link data to context, pre-existing knowledge, and additional information obtained from various sources. Unlike in computational analysis, relevant knowledge and information do not have to be represented in a structured form in order to be used effectively by a human. Furthermore, humans can note such kinds of patterns and relationships that are hard to formalize and detect by computational techniques. The privacy issues related to the use of visual and interactive methods are currently studied neither in the areas of visualization and visual analytics nor in the area of data mining and computational analysis. There is a need to fill this gap, which requires concerted inter-disciplinary efforts.

1 Introduction

Collection and analysis of data about individuals is vital for progress in many areas such as health protection, transportation, security, to name a few. Technologies enabling collection and analysis of various kinds of personal data develop rapidly. A negative side of these developments is the growing threat to the personal privacy. This particularly applies to data containing locations of people. Analysis of such data may conflict with the individual rights to prevent disclosure of the location of one's home, workplace, activities, or trips. A number of researchers in GIScience are seriously concerned with the privacy issues associated with the use of geospatial technologies (e.g. [2][3][6][7][8]). Still, there is no systematic research on geographical privacy in the field of GIScience.

More intensive research on protecting personal privacy in data publishing and analysis is done in the areas of statistics and data mining, which address, among others, the problems of preserving geographical privacy. The recently completed European research project GeoPKDD (Geographic Privacy-aware Knowledge Discovery and Delivery; http://www.geopkdd.eu/) had a particular focus on data about mobility [5] and resulted in creation of new methods for anonymization and privacy-preserving analysis of such data. The ongoing European project MODAP (Mobility, Data Mining, and Privacy; http://www.modap.org/) is a coordination action that continues the efforts of GeoPKDD by coordinating and boosting the research activities in the intersection of mobility, data mining, and privacy preservation.

Being involved in the MODAP project, we represent the visual analytics perspective on the problem of preserving personal privacy in analyzing mobility data. In this position statement, we outline our vision of the possible contribution of the geovisualization and geospatial visual analytics to the research on geographical privacy.

2 Visual Analytics Against Privacy

The essence of Visual Analytics is enabling synergistic work of humans and machines in analyzing large and complex data and solving complex, ill-defined problems. In other words, Visual Analytics is about creating such working conditions in which humans and computers can utilize their inherent capabilities in the best possible ways while complementing and amplifying the capabilities of the other side.

Humans have many unique capabilities that are valuable for analysis and problem solving. Among them, two inherent qualities are especially relevant to the topic of privacy protection:

- the capability to flexibly employ previous knowledge and experience, not only those related to special education and to professional activities but also those related to the everyday life and common sense intelligence;
- the capability to establish various associations among pieces of information.

Since these qualities are precious for analysis, Visual Analytics aims at enabling humans to utilize them in the most effective ways. However, the utilization of these capabilities in data analysis has the potential of increasing the threats to the privacy of individuals whose characteristics or activities are reflected in the data. This applies, in particular, to data about mobility. For example, Andrienko et al. [1] demonstrate the ease of identifying person's home and work places and other frequently visited places by interpreting spatial and temporal patterns of the person's moves and stops from the positions of the human common sense. The interpretation emerged from considering movement data *in spatial and temporal contexts*. The spatial context was provided by visualizing the data in a cartographic map display. The temporal context (in particular, days of the week and hours of the day) was provided by temporal histogram displays.

Researchers on privacy protection in data analysis are typically concerned with the possible threats to privacy arising from computational data processing and from integration of two or more datasets. Little is done on studying the privacy issues arising from the involvement of human analysts empowered with interactive visual tools. Regarding mobility data, it appears necessary to investigate what associations can be established and what inferences can be made by a human, in particular, by considering the data *in context*.

Privacy issues have not been addressed in the visual analytics research so far. There is a need in raising the awareness of the visual analytics community about privacy issues and in providing some guidelines concerning the ways in which visual analytics researchers can contribute to protecting personal privacy.

3 Visual Analytics For Privacy: Research Directions

Visual Analytics can contribute to the privacy protection research in two ways. First, visual analytics researchers can identify what kinds of information can be extracted from various forms of mobility data by means of visually supported analysis and consider potential implications to personal privacy. These findings can be communicated to privacy protection researchers for developing methods to remove or decrease the detected privacy threats. Second, to allow humans to deal with large datasets, visual analytics researchers often employ techniques for data generalization and abstraction. Some of the techniques that are devised for the purposes of visualization can be adapted for protecting personal privacy [9]. Both work directions require close inter-disciplinary collaboration, as exemplified in [9]. The MODAP project aims at promoting such collaborations. The following research directions are suggested for the inter-disciplinary research community.

3.1 Taxonomy of movement context

Sensitive personal information may be uncovered by linking movement data to the context, which includes

- geographical space and inherent properties of different locations and parts of the space (e.g. street vs. park);
- time and inherent properties of different time moments and intervals (e.g. day vs. night);
- various objects existing or occurring in the space and time: static spatial objects (having particular constant positions in space), events (having particular positions in time), and moving objects (changing their spatial positions over time).

Human analysts are very flexible in using various kinds of context information available in various forms, e.g. as structured data, as background knowledge, or as texts or images retrieved from the Web. The research question is: What kinds of general and specific knowledge about context can enable unwanted discoveries of personal information from movement data?

Creation of a taxonomy describing various elements of movement context, their relevant properties, and possible relationships to people's activities and movement may form a basis for a systematic investigation of the potential threats to personal privacy arising from linking movement data to context. The taxonomy should include typologies of spatial locations, paths, spatial objects, time moments/intervals, events, etc. with regard to people's activities and movement. For instance, the typology of locations should contain such notions as home place, work place, shopping place, recreation place, business area, and so on. The typology of paths should include notions of high speed road, main street, minor street, footpath, crossing, etc. The taxonomy of context should also include the possible types of relationships that may occur between moving objects and elements of the context (e.g. spatial proximity, temporal proximity).

3.2 Taxonomy of activities

Movement of people is connected to people's activities. There are examples demonstrating that general knowledge of the possible types of activities and their typical places, and/or typical times, and/or typical durations may allow a human analyst to extract personal information from movement data [1]. An analyst may also use specific knowledge about the kinds of activities that are usually conducted in particular places. A taxonomy of activities and their possible relationships to elements of movement context (places, times, objects, events) would allow researchers to go beyond particular examples and derive general understanding of what can be inferred from movement data by involving general or specific knowledge about people's activities in combination with context information.

3.3 Taxonomy of derivable knowledge/information

This taxonomy should describe the types of knowledge/information that can be extracted from movement data linked to context and activity information. Potentially sensitive types of information should be identified.

A step in this direction is the taxonomy of movement patterns suggested in [4]. This taxonomy, however, is limited to defining possible relationships between movements of two or more objects. With respect to a particular moving object, other moving objects are a part of the movement context. However, other parts of the movement context and activities of moving objects are not considered.

The theoretical work outlined in subsections 3.1-3.3 is useful not only for the research on preserving personal privacy. It may also provide foundations for developing new analysis methods, both in visual analytics and in data mining. In particular, visual analytics researchers may use the taxonomies in the design of the visual interfaces and interactive tools that can effectively support establishing links between movement data and other kinds of information and inferring new information.

3.4 Generalization and abstraction

Methods for data generalization and abstraction are used in visualization and visual analytics for dealing with large amounts of data. A side effect of using these methods is that detailed personal information may be hidden, which is a positive feature from the perspective of preserving personal privacy. Hence, data generalization methods can potentially be adapted to the needs of preserving privacy. This refers, in particular, to methods devised for movement data. We suggest that one of the future research directions should be examination of existing and emerging methods for generalization and abstraction of movement data from the perspective of their possible adaptation for privacy protection. Of course, generalization alone does not necessarily guarantee data anonymity and safety. Hence, like the other research directions, this direction requires cooperation between specialists in visual analytics, data mining, and privacy protection.

4 Conclusion

The problem of preserving personal privacy in publishing and analyzing data is addressed by researchers in statistics, data mining, and, to a limited extent, GIScience. However, the work on privacy protection conducted in these areas focuses on the use of computer technologies and does not consider the specific threats to privacy arising from combining the analytical capabilities of computers and humans. Visual analytics is an appropriate research field to address these issues. We suggest some research directions in which visual analytics could contribute to privacy protection.

References

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

[2] Armstrong, M.P. Geographic Information Technologies and Their Potentially Erosive Effects on Personal Privacy, *Studies in the Social Sciences*, 27(1), 2002, pp. 19-28

[3] Armstrong, M.P., and Ruggles, A.J. Geographic Information Technologies and Personal Privacy, *Cartographica*, 40(4), 2005, pp. 63-73

[4] Dodge, S., Weibel, R., and Lautenschütz, A.-K. Towards a taxonomy of movement patterns, *Information Visualization*, 7(3-4), 2008, pp. 240-252.

[5] Giannotti, F., and Pedreschi, D., eds. *Mobility, Data Mining and Privacy - Geographic Knowledge discovery*. Springer, Berlin, 2007.

[6] Gutmann, M.P., and Stern, P.C., eds. Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data, *Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data*, National Research Council, National Academies Press, Washington, DC, USA, 2007

[7] Kwan, M.-P., Casas, I., and Schmitz, B.C. Protection of Geoprivacy and Accuracy of Spatial Information: How Effective Are Geographical Masks?, *Cartographica*, 39(2), 2004, pp. 15-28

[8] Kwan, M.-P., and Schuurman, N. Introduction: Issues of Privacy Protection and Analysis of Public Health Data, *Cartographica*, 39(2), 2004, pp. 1-4

[9] Monreale, A., Andrienko, G., Andrienko, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., and Wrobel, S. Movement Data Anonymity through Generalization, *Transactions on Data Privacy*, 3(3), 2010, pp. 91-121