THE BOOLEAN
SNAPSHOTS OF DOCTORAL RESEARCH AT UCC

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Geoprivacy Protection of Agricultural Data

Parvaneh Nowbakht[1,2,3,*]

[1] *Department of Geography, University College Cork, Cork, Ireland*

[2] *Teagasc, Crop, Environment and Land Use Programme, Ireland*

[3] *Environmental Research Institute, University College Cork, Ireland*

[*] *Corresponding author: 111222849@umail.ucc.ie*

Abstract

A major challenge of sharing spatially explicit agricultural and agri-environmental data is to identify the trade-off between field parcel confidentiality and spatial pattern preservation. In this study, the main drawback of point-based obfuscation was identified and the polygon-based obfuscation methods were designed and developed to overcome these issues.

Keywords: geoprivacy, geospatial data, obfuscation.

Introduction

Advances in geospatial technology have generated a significant amount of geographical data in many disciplines and industries such as transportation, environment, agriculture, location-based services (LBS) and healthcare. While data sharing and integration with other sources can lead to improve efficiency and generate knowledge, location confidentiality concerns limit the sharing of data between data provider and other researchers as geographical data can be used to identify individual objects and their private information. The most recent and important use of location information and subsequently geoprivacy concerns is in the contact tracing method used for the Covid19 pandemic health crisis.[4]

To prevent object identification and retrieval of private information, geographical data needs to be obfuscated before being released and shared with others. Spatial anonymization, obfuscation and geo-masking methods have been implemented within GIScience to preserve the individual's privacy by a process of degrading the quality of locational information while maintaining the spatial properties of geographical data.[1] Obfuscation methods are widely used in some fields and sectors especially in location-based services, healthcare and criminology.

Geoprivacy for Agricultural data

Digital and smart farming is designed to achieve sustainable agriculture and obtain more production, while simultaneously preserving sustainability and reducing the environmental impact of agriculture activities.[2] Digital farming generates a significant amount of valuable data, much of which is collected from farms that is linked to the Global Positioning System (GPS) location. For example, auto-steering tractors equipped with GPS units reduce the cost, fuel consumption and subsequently protect the environment while increasing crop production by reducing overlapping of passes while planting seeds or applying chemicals and fertilizers.[7] The available spatial agriculture data if combined with climate, environmental and other spatial data can support informed decision-making at farm level, regional or national scale and enables the government to design and develop more sustainable agricultural policies and services in the agriculture sector and society.[3] However, farmers' lack of trust and knowledge about confidentiality risk of data collected on farms directly affects their motivation to adopt digital farming technology.[9] Ultimately this has a negative impact on the availability and accessibility of agricultural data that are valuable for research, innovation and agricultural policymaking.[9] In addition to increased awareness and understanding of the terms and conditions of data usage and data sharing agreements along with General Data Protection Regulation (GDPR), privacy and data protection laws, reducing the risk of identification is another key consideration to increase the farmer's willingness to adopt digital farming. It is challenging to preserve the privacy and confidentiality of agricultural data. Agricultural data often contains spatial information which has the potential for re-identification of the field parcels and disclosure of private information. This information can be leaked through the spatial coordinates, shape, and size of the field. For example, retrieving farm owner information using a farm address which can be extracted using freely available online reverse geocoding map services from geographic coordinates available in agriculture data. Linking personal information with agriculture data, could allow personal or commercially sensitive and private information to be retrieved. Farm information, such as type of crop, size of farm, type and number of animals, type of fertilizer, etc. may be used for interests other than the farmer's interests such as for informing insurance or, farm price or for marketing purposes. Equally, these data can be extracted by government agencies and used for regulatory purposes, for example, to verify compliance with environmental and animal welfare standards, or fiscal obligations. However, to-date few studies have been developed and conducted that specifically examine geoprivacy on agricultural data, despite the unique spatial patterns and privacy concerns associated with such features which can be identified by their location information among other attributes.

Spatial data and Obfuscation methods

A point is the simplest geometry feature or vector data type used to represent a location. Each point represents a location, which consists of a coordinate pair and the individual information associated with that location. Most of the obfuscation methods implemented in GIScience represent points, such as the location of the crime.[10] However, in some cases, the world phenomena are best represented using areas or polygons, such as a field parcel. A polygon is a geometric feature or vector type that represents an area which is defined by a set of connected lines whose start and end points are the same. A polygon divides the space into inside and outside regions, with the interior region representing real-world regions such as lakes, cities, and field parcels. Although, data can be represented in vector format as points or polygons, to-date the predominant obfuscation methods have been derived from point features. To reduce the risk of identification and to protect data privacy for point-based data several obfuscation methods have been developed.[6,8,10] Anonymization and Randomization are two types of point-based techniques that are used widely to protect geoprivacy while preserving statistical analysis. K-anonymity is an anonymization technique that generates an area (neighbourhood) as an obfuscation area that contains at least k-1 other locations to ensure the specific location is unidentifiable among them. Randomization is a two steps producer that first generates an obfuscation area and then transforms the original location into a random location inside the obfuscation area.[6] The generated obfuscation area can be in different shapes, such as buffers, donuts, pinwheels, or grids. The size of the obfuscation area can be constant for all locations in a dataset or can be different depending upon local location density in density-based methods.

Proposed Methods and Results

To explore the performance of point-based obfuscation methods on polygon data, the centroid of a polygon can be considered as point data. Several anonymization and randomization techniques on the centroid of field parcels were developed and tested in.[5] The performance of 27 point-based obfuscation methods were evaluated on the subset of the *Irish Nutrient Management Planning Online (NMP Online)* dataset with high point density and non-uniform distribution. Furthermore, several evaluation methods were tested to measure the ability of each method to satisfy both privacy concerns and spatial pattern preservation. Results highlighted a high percentage of false identification and non-unique obfuscation, indicating the drawback of point-based obfuscation methods when applied to geographic features best represented as polygons. The term ‘False identification’ for point features means that the obfuscated point mistakenly links to another original point. False identification for polygon features can be considered when there was an intersection between the original and obfuscated polygons. This research introduces the concept of ”non-unique obfuscation”, which is important when obfuscating static objects as two or more points might have the same obfuscated location or when the polygon nature of the objects is taken into account, there is an intersection between the obfuscated location and other obfuscated locations. A challenge of this study was to develop and implement a qualitative approach to generate optimal obfuscation area based on k-anonymity satisfaction to minimise the risk of identification and maximise spatial pattern preservation. This has led to the development of polygon-based obfuscation methods, which were designed

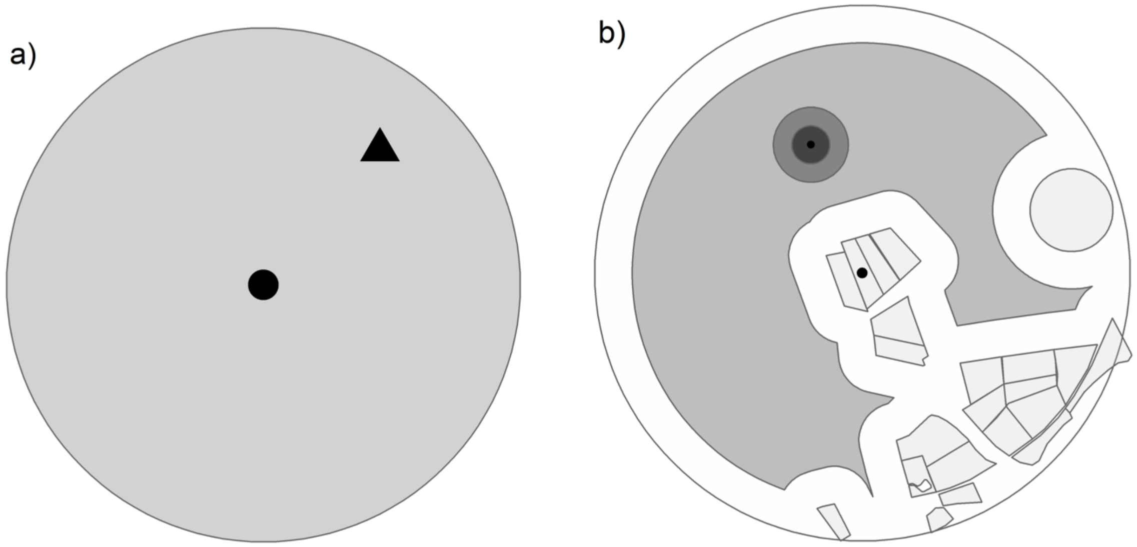

Figure 1: Conceptual diagram of obfuscation methods. a) point-based method: original point (black dot), random location (black triangle) obfuscation area (grey area). b) polygon-based method: original location (polygon with black dot), random location (black circle)

with the consideration of three properties of spatial polygon objects including the spatial coordinates, shape, and size of the polygon that can be used to identify real-world objects. These methods were developed to guarantee that there is no false-identification and non-unique obfuscation which is important for static polygon objects in terms of accuracy and security. The main idea is to eliminate extended sensitive area (field parcels) from the obfuscated area generated by point-based methods. This guaranteed that there was no intersection between the obfuscated location and sensitive area (The results of this chapter are currently under review of Transaction in GIS journal). Figure 1 illustrates the conceptualization of obfuscation methods to compare the point-based obfuscation methods and polygon-based methods.

Conclusion

The results of this PhD research indicate that the density-based methods perform the best to achieve the trade-off between field parcel confidentiality and spatial pattern preservation. Shape and size of the obfuscated polygon are two important factors for map visualization, with further research recommended to improve map visualization quality while reducing the risk of identification. A further study with more focus on the relationship between external environmental and climate data with spatial coordinates and how obfuscation methods maintain this relationship or preserve any possible clustering based on environmental climate data is suggested.

Acknowledgments

Parvaneh Nowbakht is a PhD candidate in the Department of Geography, and Environmental Research Institute, University College Cork, working under the supervision of Dr Paul Holloway and Dr Fiona Cawkwell. She is also a Teagasc Walsh Scholar, working under Teagasc supervisors Dr Lilian O'Sullivan and Dr David Wall. This research has been funded by the Teagasc Walsh Scholarship program [2018034].

References

[1] Marc P Armstrong, Gerard Rushton, and Dale L Zimmerman. Geographically masking health data to preserve confidentiality. *Statistics in medicine*, 18(5):497–525, 1999.

[2] Larisa Hrustek. Sustainability driven by agriculture through digital transformation. *Sustainability*, 12(20):8596, 2020.

[3] Marie-Agnes Jouanjean, Francesca Casalini, Leanne Wiseman, and Emily Gray. Issues around data governance in the digital transformation of agriculture: The farmers' perspective. 2020.

[4] Junghwan Kim and Mei-Po Kwan. An examination of people's privacy concerns, perceptions of social benefits, and acceptance of covid-19 mitigation measures that harness location information: A comparative study of the us and south korea. *ISPRS International Journal of Geo-Information*, 10(1):25, 2021.

[5] Parvaneh Nowbakht, Lilian O'Sullivan, Fiona Cawkwell, David P Wall, and Paul Holloway. A comparison of obfuscation methods used for privacy protection: Exploring the challenges of polygon data in agricultural research. *Transactions in GIS*, 26(2):949–979, 2022.

[6] Dara E Seidl, Piotr Jankowski, and Keith C Clarke. Privacy and false identification risk in geomasking techniques. *Geographical Analysis*, 50(3):280–297, 2018.

[7] Mustafa Topcueri and Muharrem Keskin. Effectiveness of gnss-based tractor auto steering systems in crop spraying. *Mustafa Kemal Üniversitesi Tarım Bilimleri Dergisi*, 24:78–90, 2019.

[8] Jue Wang and Mei-Po Kwan. Daily activity locations k-anonymity for the evaluation of disclosure risk of individual gps datasets. *International Journal of Health Geographics*, 19(1):1–14, 2020.

[9] Leanne Wiseman, Jay Sanderson, Airong Zhang, and Emma Jakku. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS-Wageningen Journal of Life Sciences*, 90:100301, 2019.

[10] Mayra A Zurbarán, Augusto Salazar, Maria Antonia Brovelli, and Pedro M Wightman. An evaluation framework for assessing the impact of location privacy on geospatial analysis. *IEEE Access*, 8:158224–158236, 2020.