Journal of Visual Languages and Computing 25 (2014) 995–1002

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Short Paper

Uncertainty visualization of multi-providers cartographic integration ☆

Bilal Berjawi [a,*], Fabien Duchateau [b], Élisabeth Chesneau [c], Franck Favetta [d], Geoffrey Seccia [e], Claire Cunty [f], Maryvonne Miquel [a], Robert Laurini [a]

[a] LIRIS, UMR 5205, INSA de Lyon, Lyon, France
[b] LIRIS, UMR 5205, UCBL, Lyon, France
[c] EVS, UMR 5600, UJM, Saint Etienne, France
[d] LIRIS, UMR 5205, ENSNP, Lyon, France
[e] EMC, EA 3082, Université Lumière Lyon 2, Lyon, France
[f] EVS, UMR 5600, Université Lumière Lyon 2, Lyon, France

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form
29 October 2014
Accepted 31 October 2014
Available online 11 November 2014

Keywords:
Cartography
GIS
Visual uncertainty
Spatial integration

ABSTRACT

Multiple cartographic providers propose services displaying points of interests (POI) on maps. However, the provided POIs are often incomplete and contradictory from one provider to another. Previous works proposed solutions for detecting correspondences between spatial entities that refer to the same geographic object. Although one can visualize the result of the integration of corresponding entities, users do not have any information about the quality of this integration. In this paper, we propose a solution to visualize the uncertainty inherent to a spatial integration algorithm. We present an integration process that identifies three levels of confidence for spatial and terminological integration results. Based on perceptual tests, we select visual variables to portray these three levels of confidence and we choose a visualization strategy. A prototype has been implemented to present the benefits of our proposal in a use-case scenario. This work has been realized within the framework of UNIMAP[1] project.

1. Introduction

Location-based services (LBS) are daily used in various applications, and cartographic providers play an essential role in displaying points of interest (POI) such as restaurants, hotels, and tourist places. A POI can be defined as a geographic object that has a point geometric shape. A POI has spatial attributes longitude and latitude, and terminological (non-spatial) attributes such as name and type (e.g., restaurant, hotel). Some providers may supply additional terminological attributes such as address, phone number, Website, customers' ratings, etc. A provider usually represents a POI on a map with a specific symbol or icon. Due to lack of completeness, noisy, inaccurate and contradictory data, it is interesting to propose solutions for detecting corresponding entities (i.e., which refer to the same POI) from different providers. This challenge aims at improving the quality and the relevance of information, which has a significant impact in tourist applications.

The integration of spatial information issued from different sources has been studied [9]. Earlier works so called "map conflation" were specifically devoted to vector objects such as roads [22]. In the last decade, the integration problem mainly

☆ This paper has been recommended for acceptance by Shi Kho Chang.

* Corresponding author.
E-mail addresses: bilal.berjawi@liris.cnrs.fr (B. Berjawi), fabien.duchateau@liris.cnrs.fr (F. Duchateau), elisabeth.chesneau@univ-st-etienne.fr (É. Chesneau), franck.favetta@liris.cnrs.fr (F. Favetta), geoffrey.seccia@gmail.com (G. Seccia), claire.cunty@univ-lyon2.fr (C. Cunty), Maryvonne.Miquel@liris.cnrs.fr (M. Miquel), Robert.Laurini@liris.cnrs.fr (R. Laurini).

[1] UNIMAP: http://liris.cnrs.fr/unimap (July 2014).

http://dx.doi.org/10.1016/j.jvlc.2014.10.033
1045-926X/

refers to the "entity matching" research domain, enhanced by a spatial aspect. The discovery of corresponding entities is performed either by exploiting only spatial information [25] or by computing and combining terminological similarities for selected attributes (e.g., name, type) [21]. Machine learning algorithms may be applied for tuning the parameters (e.g., weights) of a matching process [27]. When corresponding entities have been detected, an interesting use case aims at displaying a merged entity, i.e., to use a crafted algorithm to fusion the attributes' values of these corresponding entities. Such merging algorithms are not 100% confident. For instance, two corresponding entities may have a different location and the algorithm needs to determine the correct position. Similarly, the names or the phone numbers of two corresponding entities may differ, and the choice of the correct values relies on the merging algorithm. A merged entity may therefore include at different levels some uncertainties, which have to be presented to end-users [18].

In this paper, we are interested in visualizing the uncertainty resulting from the merging process of spatial entities. Our contributions can be summarized as follows: (i) identifying the "dimensions" which have to be taken into account for uncertainty, i.e., the POI type, the spatial attributes and the terminological (non-spatial) attributes; (ii) measuring the confidence level for each dimension as well as a global confidence score; (iii) proposing many visualizations of a merged entity and its uncertainty and testing them to select the best; (iv) implementing a prototype to demonstrate in a scenario the benefits for end-users.

The next section describes the related work, both in spatial integration and uncertainty visualization. Section 3 provides a detailed explanation of our solutions, tested among different users, to represent and visualize various criteria about a merged entity. In Section 4, we demonstrate the benefits of our approach in a scenario, and we conclude in Section 5.

2. Related work

This section covers the existing works in two domains: the methods for integrating spatial data and the visualization of uncertainty in a spatial context.

2.1. Spatial integration

The same reality is described with a multiplicity of geographical information. This information growth rapidly over the Internet, some may be incomplete, inaccurate or contradictory. Integration of several sources of geographical information is necessary in order to update information that changes daily [12] or to produce a more complete and accurate information [7]. In [32], authors define three categories of imperfection that occurs when integrating several spatial data sources, namely (i) inaccuracy, which concerns wrong spatial information that do not correspond to reality, (ii) imprecision, which deals with spatial information that corresponds to reality but is not sufficiently precise and (iii) vagueness, which is about ambiguity of spatial information (e.g., boundaries heterogeneity). Geospatial integration has been widely studied under the term "map conflation" where two whole maps are integrated. Integration of maps consists in identifying the corresponding entities and to fuse them [5]. In [22], authors describe existing works in map conflation regarding their formats (raster and vector) and their criteria such as spatial data, terminological data and topological relationships between entities. Some works have been proposed in map conflation using points [23,6,30], lines [24,10,31] and polygons [1,11,19].

In [2,25], the authors use only the spatial information (location) to detect the corresponding entities with a similarity measure based on probabilistic consideration. The probability that two entities are corresponding is estimated using the Euclidean distance between them. In order to improve the quality of integration, some works propose to combine similarity measures that use spatial information with similarity measures that use terminological information to identify correspondences. In [26], three algorithms were proposed using a first similarity measure to filter the entities and a second to detect the corresponding entities. For example, a string similarity measure can be applied on the name of the POI, then for each pair of entities that are not considered as corresponding, the distance between them is increased. The final step is to apply a similarity measure on spatial information with the new distances. Note that increasing the distance between two entities lowers the probability that they will be considered as corresponding entities when we apply a similarity measure on spatial information. A variety of learning-based methods including logistic regression, support vector machines and voted perceptron has been proposed to find out how to combine and tune several similarity measures in order to identify the corresponding entities [27].

The "Theory of Evidence", also called "Dempster–Shafer theory" [28], combines an evidence measure of different sources and finds a degree of belief that takes into account all the available evidence. That is, a belief mass represented by a belief function, is associated to each evidence, then Dempster's rule is used to combine them. The "Theory of Evidence" is proposed to integrate geospatial databases [21] and to match geospatial entities of several LBS providers [14].

Kang et al. propose a visual interface to detect the corresponding geospatial entities based on a neighborhood similarity [13]. It takes two sources of entities as input, and then the user chooses a similarity measure to apply on terminological information or on spatial information. Detected entities are considered as potentially corresponding. Then each pair of entities is visualized on the screen. Their shared neighborhood of entities are placed between them and non-shared neighbors on the sides. Finally, the user has to make a decision for each pair to be considered as corresponding or not.

2.2. Spatial uncertainty visualization

Thomson et al. [29] and MacEachren et al. [17] define nine categories of uncertainty paired with three components of geographic information: space, time and attribute (terminological). On this basis, Thomson et al. [29] make an empirical study to characterize the kind of visual signification that is appropriate for representing those different categories of uncertainty. The authors use a set of visual