

Reputation as tool to ensure validity of VGI

Patrick Maué

Institute for Geoinformatics / University of Münster

pajoma@uni-muenster.de

The term “volunteered geographic information” (VGI) refers GI which is created in collaboration by users who usually don't have special skills in handling spatial data. This makes it difficult to incorporate such GI in applications depending on high quality data. An example is local decision-making (Cara et al., 2007) within participatory GIS (PGIS). But relying only on GIS experts neglects the fact that involving interested users is an important step towards an open and democratic approach for PGIS (Rattray, 2006; Tulloch, 2007) . And the local population has usually better knowledge about the area of interest, which is crucial for accurate decisions. Our research is focusing on an integral part of web-based PGIS: a catalog which allows for publication and discovery of VGI. Users register VGI in form of feature collections, others use spatial, keyword-based, or even semantic-enabled queries to locate and select the published GI. This is efficient as long as the GI (e.g. a map providing hiking trails) is only downloaded for personal use (e.g. a person which plans a hiking trip). But if one of the actors (a local tourist office which wants to compile a guide for popular hiking routes) is liable for the validity of the information, having a possibility of evaluating VGI's reliability and credibility becomes crucial. The concept of information asymmetry¹ plays a significant role for the selection of suitable VGI. The author has probably walked the trails by himself and might have put much effort into digitizing the route. The tourist office doesn't have this information. But they know that, in many cases, VGI lacks the quality needed for compiling the planned guide. At the end they therefore decide to buy the information and to entrust a company to gather the required GI. Perhaps the tourist office acts different next time, if we show them a way to evaluate the validity of VGI and distinguish between “good” GI created by reliable users and GI created by users having a *reputation* of being usually less careful in digitizing.

Reputation is build upon the history of past interactions happening between members of one community. The "other party's abilities and disposition" (Resnick et al., 2000) are the features reputation is based on. Reputation is used to estimate the risk of future interactions. EBay's rating system², used to assess the credibility of auctioneers, is a well-known example for a reputation system. Within the catalog, multiple actions can be analyzed to infer the reputation change of the participating actors. Metadata can be created, extended and modified. The described feature collections can be rated, tagged, discussed, annotated, and more. Some actions, like tagging, are explicit. Relevance feedback on the other hand is an implicit action. Both affect a user's reputation value, which is used for the following tasks:

1. **Assess reliability:** VGI by a user with high reputation in creating maps (which means that most of his creations have good ratings and are frequently used) is usually trustworthy. This should of course only be a part of the evaluation of the suitability and usability of VGI. Other characteristics to consider are completeness, level of detail, accuracy (if this is part of the data), popularity, and more.
2. **Infer local knowledge:** If a user is known for creating reliable GI of a particular region, we can infer that he might have local knowledge. If such a user suddenly creates GI of a completely different area, we can not simply assume that this information has the same reliability. Reputation is not a single value which can be

¹ Wikipedia is good source for explaining the concepts of Information Asymmetry and Moral Hazards

² More information here: <http://pages.ebay.com/services/forum/feedback.html>

applied to every setting. It is the history of past actions, and depending on the scenario only parts of the history are of interest and need to be extracted.

3. **Assess Skill:** The quantity of a user's past interactions show how familiar he is with the catalog's functionality, and if he can be trusted with more complex, but also more effective methods to enhance metadata. A user just registered to the catalog should not even be able to tag specific feature collections. Users with high reputation on the other hand will be, for example, allowed to directly modify the ontologies used for the discovery of the metadata. This is an incentive for users to achieve a high reputation, which potentially ensures more user contributions and at the end more elaborated metadata.
4. **Avoid Moral Hazards:** Reputation can act as sanctioning device to avoid moral hazards (Dellarocas, 2006). Detected incorrect modifications by a user have a negative impact on his reputation, and restrict, as consequence, his access to less important operations. Reputation is a dynamic property, which can decrease due to misbehavior like deliberately wrong tags (e.g. spam) or imprecise positions due to laziness.

An existing real-life relationship between registered users has an impact on some contributions. Users will usually rate a friend's feature collection higher than (potentially better) VGI of an unknown user. Our current research is focused on a model which is able to capture explicit and implicit actions and their effects on the reputation. Moreover, the model has to incorporate the social network reflecting the existing relations between users of the community as well as the reputation (modeled as history of past actions) of a single user. A catalog providing a set of basic user feedback techniques like relevance feedback or tagging will be implemented to test the model. We believe that gaining reputation to get access to higher-level operations (like semantic annotations) is a sufficient motivation for users to contribute to the existing metadata records. And having many users actively contributing is expected to result in elaborated metadata which makes the described VGI, at the end, more useful (in terms of validity and findability) for critical applications like PGIS.

References

- S. Cara, G. Carcangiu, G. Padalino, M. Palomba, M. Tamanini, R. Kingston, Carver, and I. Turton (2000). Web-based public participation geographical information systems: an aid to local environmental decision-making. *Computers, Environment and Urban Systems* 24(2):109–125.
- C. Dellarocas (2006). *Handbook on Information Systems and Economics*, chapter Reputation Mechanisms.
- N. Rattray (2006). A User-Centered Model for Community-Based Web-GIS. *URISA Journal*, 18(2):25–34
- P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman (2000). Reputation systems. *Commun. ACM*, 43(12):45–48.
- D. Tulloch (2007). Many, many maps: Empowerment and online participatory mapping. *First Monday*, 12(2).