Short Term Scientific Mission

COST Action IC0903

"Mining Spatio-temporal Change-patterns from VGI History"

Reference Number: ECOST-STSM-IC0903-070512-016875

STSM Grant Holder: Khatereh Polous

Cartography Department

Technical University of Munich (TUM)

Munich, Germany

Email: polous@bv.tum.de

STSM Host: Dr. Peter Mooney

Geotechnologies Research, Computer Science Department,

National University of Ireland Maynooth (NUIM),

Maynooth, Ireland

Email: peter.mooney@nuim.ie

STSM Date: 7.05.2012 – 18.05.2012

Background

With the rapid spreading of the concept that uses the web as "participatory platform", increasing amount of information is uploaded on the internet. This "participatory platform" provides colloquial read-and-write functionality for individual users to collaboratively generate new contents or modify the existing ones. Texting, social networks, photos, videos, blog entries are the most popular forms of user generated contents. In parallel, technological advancements in location-aware devices, web mapping technologies and mobile cartography facilitated the contribution of none professional volunteers in providing and distributing geographic information (Goodchild, 2007). Many web sites such as OpenStreetMap (OSM) and Google MyMaps have been launched during recent years to encourage the actions of volunteers. With an increase in the number of contributors for creating and assembling of data, open source database products have drawn the attention of decision makers for facility management and service planning. From the scientific point of view, the availability of accurate and/or up-to-date mass data stimulates the development of innovative approaches for the assessment of spatio-temporal processes and detailed change detection. In fact, by developing these facilities, assessing spatio-temporal processes and detailed change mapping, which highly demand accurate and up-to-date data, has become more affordable. Many studies have been already conducted concerning the detection, monitoring and visualization of changes from time series data (Coppin et al 2004, Liu et al 2010, Tanathon 2009). But open questions on how the detected changes should be decomposed and formulated to reveal an event still remain.

In order to answer why the study of event is a vital, one can refer to; the events and their behavior patterns represent a higher level of knowledge than changes, and thus more valuable for decision makers. Indeed, to explore the mechanism of the changes, one must investigate the mechanism of events. Events underlie changes (Chrisman 1998, Worboys and Hornsby 2004). Highly complex and irregular patterns still can be represented visually for further analysis or exploration. In addition, for the efficient planning and management of a complex system composed of engineering, natural and social components, it is necessary to consider the relationships between an event and the reactive behavior of different components.

In the context of VGIs, the term event refers to anomalous user activities, number of contributors plus amount of contributions, which happened at a time point or within a specific period of time. In this mission, we focused on the events which have happened and events for which we have prior knowledge. As event, one can refers to a broad category of significant changes which was caused by a social or natural factor. By social events we refer mostly to the events that are planned by people, attended by people and that people will use social media to capture, document, or discuss about them. The implemented methods can be used to mine any relevant knowledge from VGI history for assessing motion patterns of moving objects. The normal or abnormal patterns of movement/distribution for human activities, vehicles, disease and animal/ plant spices (such as eBird community) can be extracted and visualized.

The aim of this study was to gain a deeper insight into the structure of OSM data for extracting change-pattern of dynamic and static objects and evaluating different possibilities for event detection. The main scope was to detect retrospective events from OpenStreetMap

(OSM) as one of the most prominent examples of VGIs for the specific sites in Vechta, Munich, Los Angeles, and Sendai. Indeed, this work concentrates on detection of different types of events, which are bound to a specific time and place from delivered information by internet users. In order to examine the ability of the OSM history for event detection, two annual social events -“Stoppelmarkt” in Vectha and “Okroberfest” in Munich- one temporary event -freeway closure in San Diego (I-405) for two days- and one natural event -Tsunami in Sendai, Japan- were studied. The reason behind selecting two different festivals like “Stoppelmarkt” and “Oktoberfest” is to evaluate the effect of event popularity and the number of participants on the possibility of event detection from OSM history. These two locations were chosen for the study, because Germany is considered as one of the most active OSM communities in Europe (Mooney and Corcoran, 2011). Subsequently it might provide a higher possibility to detect these kinds of events.

Purpose of the visit

Short Term Scientific Mission (STSM) was performed in order to

- Accessing historical information from OpenStreetMap (OSM)
- Exploring the structure of OSM history database
- Evaluating different possibilities to track significant changes (events) in OSM databases

Description of the work carried out during the visit

For the purposes of retrospective event detection, it is necessary to analyse the history of OSM for the area of the event. It might be very difficult to detect any traces of events, if one limits analysis to the current available snapshot of OSM from outlets. Because the current snapshot shows the most up-to-date version of every geographical object in the OSM database. The OSM History is a very rich spatial database as every edit for the entire planet (where an edit is considered to refer to the creation of objects and their subsequent update) is recorded in this database.

The OSM historical data for the entire world in XML format were downloaded from planet OSM history files. Considering the size of entire history and required time for processing, it is necessary to extract the area of interest. To achieve the objective, the entire OSM history of Vechta, Munich, Los Angeles, and Sendai, around the area where events happened was studied to examine the ability of OSM for event detection. In addition to handle the process efficiently, the history file was divided into two separate parts of Nodes and Ways. The extracted OSM-XML was then stored and processed in a Postgresql PostGIS database for further analysis. For each element (node, way), geometry and corresponding semantic information were extracted for the entire time of monitoring. We hypothesize that there would be an increase in the number of contribution through the increase in the number of contributors in OSM during the festival each year or the time when the event happened. It is assumed that the hottest point/ period time are the most likely points/ period for detecting the occurred events. Indeed they are the point in which, an anomalous user activity appeared due to the happening of an event in the reality. In order to achieve the main objective of the study, aforementioned steps were applied for all selected sites. Only in the case of I-405 in L.A the GPX traces within the area, where the freeway was closed temporarily, were downloaded additionally as auxiliary data from OSM history. OSM provides access to all of the GPX traces uploaded by contributors over time to the OSM database. Unfortunately the GPS traces are not organized in the same structural as the OSM history. One must execute calls to the OSM API service to download these GPS traces. A Python script which was given a bounding rectangle area coordinates was developed to handle this step automatically. The script downloaded all of the GPS traces, which were inside or intersected this bounding rectangle. As all of the GPX traces are not time stamped, the script discarded these from analysis, because it is impossible to ascertain their temporal relevance. The downloaded GPX format then was imported directly into QGIS for further exploration and visualization.

Description of the main obtained results

In this study, we extracted knowledge of different type of events which have happened and events for which we had knowledge (spatial size, impact, type of event). OSM like the other sharing platforms have the possibility to provide the semantic information for the mapped objects by tagging them. Tags do not follow any preset rules other than the ones chosen at the moment of their creation by the contributors themselves. As there is no formal norm for generating tags (OSM has a guide but these are not strictly enforced rules), users usually tap in their personal thesaurus clues that can be inspired by their state of knowledge or their representations or their imagination. In addition, changes to OSM can be effectively rendered on the globally visible OSM maps in a few hours. These functionalities can make OSM a powerful community for collecting both geometry and semantic information of physical mapped objects. The key motivation of this work was to mine hidden pattern of social activity and their interest for sharing event related knowledge within OSM community.

The obtained results revealed that OSM is not a tool which citizens would naturally turn to mark an event or even provide the attributes for involving spatial objects. The contributors do not treat this community in the same way that they contribute to user-friendly electronic exchange platforms such as Twitter, Facebook, or Flickr. According to the results, in the cases of Munich and Vechta, the activity of contributors (Number of contributors plus Number of edits) for a specific period of time cannot be considered as an appropriate indicator for detecting social/temporary events in OSM community. There was no detectible indication of unexpected increase in the amount of contributions and the number of contributors during event time. Mining the tags for both events showed that the contributors have mapped the temporary objects and reported the events in their tags and comments but unfortunately the information are only available in the advanced editing modus of the map for the registered contributors. This implies that in the context of OSM, a new definition for event must be considered.

Regarding pattern recognition; the mapped objects of “Vechta” festival are constantly displayed on the map while they are only available during the festival time. Indeed, no repetitive pattern was recognized from constant existence of the involving objects. Based on the history file, the mapped objects are static objects which have been created after a sudden event (not repetitive). In the case of Oktoberfest, the tents are not visible in the view modus of the map as the volunteers have identified the start, and end dates of the event, and have hidden the temporal tents in the Oktoberfest place. In fact, pattern of the event can be identified as the contributors regularly have updated the status of the tents for being hidden and unhidden regarding the event time.

It can be a good hint to manage event related knowledge in OSM but this attempt has many deficits. The biggest deficits of this attempt are as following:

- Unregistered users do not have any access to the embedded information (e.g. start and end date of events) as they are only visible in the editing modus.
- Continuous maintenance is needed to update the map for each period of time. For the Oktoberfest and Stoppelmarkt, the maintenance interval is yearly but for ir/regular daily events, the maintenance is almost impossible.
- Mapping of temporary objects for repetitive periodical events affect the quality of OSM for the users who are not frequently updating their downloaded maps.
- Many events are not manageable to be presented through hide and unhide mechanisms, such as a meeting in a hall. In these cases event and objects are connected unidirectional with each other. By unidirectional connection we refer to an event which needs a building to be held but the building can exist without any event.

The closure of freeway I-405, 16-17 July 2011, in Los Angeles was the theme of our third case study. Changes in the road infrastructure could imply changes in motion patterns of moving objects and vice versa. For I-405, due to the nature of the event and the place where event happened, we did not expect any anomalous changes in the activity level of OSM community. The contributors are drivers on the freeway not the people who want to attend a specific event. Tag mining of OSM history for the area where the freeway was closed showed that no one reported anything to update the status of the freeway for the time when event happened. Only two contributions were recorded during July 2011, which were not referred to the event. In addition, uploaded GPX traces into OSM within the area of event were downloaded as auxiliary data. OSM provides access to all GPX traces uploaded by contributors over time to the OSM database. While there are disadvantages to the use of GPS traces but the quantity of GPS traces in a given area could be used to indicate some type of "interest" in an area which could possibly be related to the occurrence of some event. In our case, decrease in the number of GPX traces might be a sign of this interstate interruption. According to the result, only one GPX trace was uploaded for the given bounding box which could not present any trace of event.

For the case of tsunami in Sendai, a natural event in which the geometry of objects is affected, post-disaster, the structural and environmental damages (demolished buildings, road infrastructure changes, etc.) was detectable from OSM platform. It can be concluded that for the natural events the "activity" is considered as the first pre-processing step towards event detection. While the tags mining is used for further exploration such as the type of event, the required facilities for understanding of the disaster. In this case, an obvious increase in the number of contributors and amount of contributions over a very short space of time show that "something" had caused these people to begin contribution to OSM on a large scale. Of course OSM activity levels (number of edits, number of editors) in a given area over a given length of time can indicate that there is “increased OSM activity” in the area.

Overall there appears to be a lack of event detection and pattern recognition possibility for OSM community especially for the social and temporary events. This disability can be driven from two reasons; the first is the way people treat OSM for creating spatial data and sharing corresponding information which are bound to the objects at specific time and space. Indeed, there are many physical objects that could be mapped in OSM if contributors are willing to map them. The Second reason refers to the ability of OSM itself for collecting this knowledge and providing them to the users. One might be interested in mapping the temporary objects and reporting the relevant events, but still OSM cannot be able to provide the information properly for the users. In this case the temporary objects either are displayed constantly on the map (Vechta case) or have to be continuously manipulated by contributors before and after the events (Oktoberfest case).

However, to make best use of this huge amount of information within these databases, new information retrieval systems with the capability of reasoning and representing latent knowledge is of vital importance.

An extensive analysis of the obtained results can be found in the paper entitled “Mining Event Related Knowledge from OpenStreetMap”.

Future collaboration with host institution

Both institutes are interested in future collaboration to exchange the methods and experiences in the field of event detection from different VGI sources.

Projected publications/articles resulting or to result from the STSM

The obtained results were documented as a full paper entitled “Mining Event Related Knowledge from OpenStreetMap” and submitted to the 9th International Symposium on Location Based Services (LBS) in Munich, 16-18 October 2012. It has been accepted to be published as a peer-reviewed paper in a series of a book in the series Springer’s Lecture Notes in Geoinformation and Cartography.

Confirmation by the host institute of the successful execution of the mission

See annex

Acknowledgements

I gratefully acknowledge the financial support by European Cooperation in Science and Technology under MOVE-COST program for the COST Action IC0903. I also would like to thank Dr. Peter Mooney for his valuable help and guidance and willingness to share his knowledge. Due to his intensive knowledge and various experiences in processing and

analyzing OSM history database, the visit became very fruitful and invaluable for my PhD project.

References

COPPIN, P., JONCKHEERE, I., NACKAERTS, K., & MUYS, B., 2004: Digital change detection methods in ecosystem monitoring: a review. In International Journal of Remote Sensing, Taylor & Francis Ltd, vol. 25, no. 9, pp. 1565-1596.

CHRISMAN, N., 1998: Beyond the snapshot: changing the approach to change, error, and process. - in: M. J. EGENHOFER & R. G. GOLLEDGE (Eds.), Spatial and Temporal Reasoning in Geographic Information Systems. - Spatial Information Systems, Oxford University Press, New York, NY, pp. 85-93.

GOODCHILD, M. F., 2007: Citizens as sensors: The world of volunteered geography. GeoJournal, 69, 211–221.

LIU, Y., MCGRATH, R. E., WANG, S., PIETROWICZ, M., FUTRELLE, J., & MYERS, J. D., 2008: Towards A Spatiotemporal Event-Oriented Ontology. - Microsoft eScience Workshop, December 7-9, 2008, Indianapolis, IN, USA.

MOONEY, P. & CORCORAN, P., 2011: Accessing the history of objects in OpenStreetMap. Proceedings of the 14th AGILE International Conference on Geographic Information Science, Utrecht, The Netherlands, Eds: Stan Geertman, Wolfgang Reinhardt and Fred Toppen p 141.

TANATHO, S., RUDAHL, K. T., & GOLDIN, S. E., 2009: Object oriented change detection of buildings after a disaster. In ASPRS, Annual Conference, Baltimore, Maryland, March 9-13.

WORBOYS, M., & HORNSBY, K., 2004: From objects to events: GEM, the geospatial event model. - Lect. Notes Comput. Sci., 3234, pp. 327-343.