Summarizing geo-annotations from social media: a promising approach for smart cities and communities

This is a pre print version of the following article:

Original Citation:
Summarizing geo-annotations from social media: a promising approach for smart cities and communities / Meo, Rosa; Ruggero G. Pensa, ; Blasi, Gianpiero Di; Bertorello, Mattia. - (2015), pp. 1-19. ((Intervento presentato al convegno CINI Annual Workshop on ICT for Smart Cities & Communities tenutosi a Palermo nel 29-30 Ottobre 2015.

Availability:
This version is available http://hdl.handle.net/2318/1563591 since 2016-06-28T17:36:02Z

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
Summarizing geo-annotations from social media: a promising approach for smart cities and communities

Rosa Meo, Ruggero G. Pensa, Mattia Bertorello and Gianpiero Di Blasi
Dipartimento di Informatica, University of Torino

Abstract—This paper describes a geosummary, an artifact made by geo-referenced annotations provided by volunteers of social media applications. The obtained geosummary constitutes a valid support for all the applications of the smart cities and communities: environment monitoring, urban services planning, city virtual exploration, for tourism and culture sharing, urban security planning and protection.

I. INTRODUCTION

Social media are frequently used by people in almost every activity of their lives. When they travel, when they work to gather socially related information, to share their experiences, to organize their lives or their free time, etc. Almost all this information is fresh and continuously updated and often it is geo-referenced. New social media applications are launched everyday (such as Flickr [1], Instagram [2], Foursquare [3]) and provide a variety of information related to almost every issue of our lives and knowledge heritage. In fact often social media applications are equipped by a semantic equipment under the form of ontologies, query languages and knowledge representation graphs that are needed to represent and deploy the users knowledge. In addition, the social media applications provide API and adopt open standards with the purpose to share and represent in the most interoperable way the accumulated knowledge heritage. Finally crowd-sourcing applications [4], [5] and collaborative knowledge sharing open projects, such as Wikipedia, OpenStreetMap [6], Onki [7] and most of the projects from the LinkedOpenData cloud [8] demonstrated that the open collaboration in knowledge representation and sharing is able to provide verified and reliable information. This occurs thanks to the amount of independent people that provide, use and verify the information and the mechanisms the exploit the redundancy in the information to measure and test its reliability [9], [10].

Overall these social media applications constitute a rich repository of information that can be used to create up-to-date maps and can be shared to improve our lives organization and our neighborhood. The maps represent a variety of issues as a consequence of what is represented in the knowledge repository of the social media and open content projects: the services available in our cities, the business organization, the events and venues offered for our free time, contents from the cultural heritage and the tourists attractions, etc. These are examples of the multifaceted semantics of the annotations left by the users of the social media and collaborating human sensing applications that we can exploit and represent in the maps.

II. GEOSUMMARY

In the context of SMAT-F2 [11] we initially designed Geosummary [12], [13]. SMAT-F2 is an environment monitoring project by means of unmanned aircraft vehicles, equipped with payload sensors for the high-distance monitoring of the territory in multiple weather contexts. Geosummary is a web application that we originally thought as a way to support mission planning or during mission post-processing for the interpretation of the images recorded by sensors by superimposition of the images on the annotated cartography.

Soon, we realized that the geosummaries were useful even in broader contexts: for citizens, communities, scientific projects in environment monitoring and planning, for e-government institutions (e.g., public administrations), media (with initiatives such as story-telling and e-journalisms). All these actors have in common the necessity of making sense from the big amount of information coming from the web on the territory, in support to the smart cities and communities.

We created a geosummary as an artifact aggregating the social media annotations. The aggregation occurs by means of clustering of the many data-points (geo-referenced annotations left by the users of the social tagging projects) after a first phase of data preparation. Clustering was made difficult by the sparsity of the data and by the multitude of the possible annotation categories. We implemented a variant of Subclu [17], an algorithm for clustering in multiple dimensions.

A geosummary acts as a sort of thematic map that represents in a colored and interactive way the areas of the territory in which some geographic features or certain categories of annotations on the human activities are present with a particular density. Geosummary in particular is able to automatically detect the statistically significance of the density threshold. It is able to consider even the zones in which the combinations of annotation categories occur with a frequency higher than the threshold given if the categories were independent. A geosummary makes evident the interesting issues of the geography by means of the visualization of the phenomena related to the territory (such as the human activities) and allows an easier comprehension by visualization of meaningful patterns.

A geosummary is composed of fingerprints, colored polygons annotated by the category of the events, points of interest
or venues that occur in the covered area. Each fingerprint is then described by the observed statistics (surface area, density, heterogeneity, etc). Figure 1 provides a geosummary produced by Geosummly on Milano with annotations extracted from Foursquare in February 2014.

III. SOCIAL MEDIA FUSION AND RELIABILITY VERIFICATION

Soon we realized of the necessity of addressing two issues: (1) integrating the available data sources and (2) verifying the reliability of the users provided data, thanks to the presence of independent sources and exploiting the existence of some redundancy in the annotations. We approached to the integration of social media by matching the ontologies of the knowledge repositories into a reference ontology. We chose the ontology of OpenStreetMap [14] as a referential one given its coverage of different issues related to territory annotation and feature representation. When necessary, and in particular when the categories of the ontologies were not in a one-to-one relationship, we approached by single instance classification into the categories of the referential ontology. To this purpose we developed machine learning classifiers (with RandomForests models [15]) and trained them on the data coming from the entire territory of Italy in the month of June 2015, from both Foursquare and OpenStreetMap. We obtained classification accuracies ranging from 70% to 98% depending on the categories [16]. We approached the second issue by application of some heuristics based on the co-occurrence of independent annotations on similar named entities [16], and computing the precision of the annotations coming from independent users in crowdsourcing applications.

IV. DEPLOYMENT OF A GEOSUMMARY IN THE SMART CITIES AND COMMUNITIES INITIATIVES

Geosummaries can be used for the goals of virtual exploration, monitoring and planning of the territory. For surveilling human activities for environment protection and disaster prevention, early intervention in case of emergencies and for the projects of support to public safety and urban security.

They can be used by private people and tourism agencies for planning tourism activities in a foreign city in order to explore the tourism attractions, the areas of food offer, for hotel/travel reservation as well as for visits planning to cultural and artistic attractions.

Geosummaries can be used by the metropolitan agencies to present the historic and cultural heritage, for their protection and the improvement of the city security. New web applications (similar to TripAdvisor) can be offered to citizens and tourists such that they can give useful feedbacks. In turn people feedbacks can be aggregated in new geosummaries, published and shared with the goal to improve the transparency of the offer, the city attraction and its security.

Well-known disaster management situations and early intervention projects (such as for Haiti or for Fukushima earthquakes) deployed social media, the web and the users feedbacks to plan the intervention and monitor the situation. Also in these cases the production of geosummaries for specific annotation categories (emergency organisms, hospitals, schools, railways, roads, power towers, etc) can constitute essential tools to summarize important issues in a short time and rank the priority of the activities.

V. CONCLUSIONS

This paper introduced a geosummary as an aggregation of the multifaceted issues occurring on the territory. We first overviewed the series of activities underlying the computation of a geosummary. Then we listed the possibilities for the deployment of geosummaries in the context of Smart Cities and Communities.

REFERENCES


Fig. 1. Example of a geosummary