Applied Geography 43 (2013) 36–44

Contents lists available at SciVerse ScienceDirect

Applied Geography

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

Automated geographic context analysis for volunteered information

Laura Spinsanti [1], Frank Ostermann*

European Commission Joint Research Centre, Institute for the Environment and Sustainability, Digital Earth and Reference Data Unit, Via E. Fermi 2749, 21027 Ispra, Va, Italy

Keywords:
Volunteered geographic information
Social media
Spatial data infrastructures
Spatio-temporal clustering
Geographic context analysis

A B S T R A C T

Several studies show the impacts of (geo)social media and Volunteered Geographic Information (VGI) during crisis events, and have found intrinsic value for rescue teams, relief workers and humanitarian assistance coordinators, as well as the affected population. The main challenge is how emergency management and the public can capitalize on the abundance of this new source of information by reducing the volume to credible and relevant content.

In this paper, we present the GeoCONAVI (Geographic CONtext Analysis for Volunteered Information) approach and a prototype system, designed to retrieve, process, analyze and evaluate social media content on forest fires, producing relevant, credible and actionable VGI usable for crisis events. The novelty of the approach lies in the enrichment of the content with additional geographic context information, and use of spatio-temporal clustering to support scoring and validation. Thus, the system is focusing on integrating authoritative data sources with VGI. Evaluation in case studies shows that the prototype system can handle large amounts of data with common-off-the-shelf hardware, produces valuable results, and is adaptable to other types of crisis events.

Introduction and related work

In this paper, we present a proof-of-concept system to extract, to process and to analyze volunteered information on forest fires from social media networks. The novelty of the approach lies in its focus on using location to filter and verify the information circulating in social media networks. By first geo-referencing the retrieved content and turning it into volunteered geographic information (VGI), it can subsequently enrich it with additional geographic context information from authoritative data sources, and cluster it spatio-temporally to support filtering and verification.

We have argued elsewhere for the importance of clearly distinguishing terms such as "crowd-sourced geo-information", "citizen science", and VGI (Craglia, Ostermann, & Spinsanti, 2012). The GeoCONAVI system currently consumes all types of user-generated content: Volunteered as well as contributed (Harvey, 2012), on geographic locations or not. We have decided to use the established term VGI throughout the paper, because the output of the GeoCONAVI system is just that: Information on geographic features and processes related to a crisis event, which has been volunteered with the purpose to be consumed and acted upon by others who are affected by that same event.

Our motivations are the opportunities provided by changing ways in which environmental information is collected, distributed and used. Public authorities traditionally generate, manage, update and distribute information in accordance to established rules and procedures in closed systems to ensure reliability and trustworthiness. Information flowed from "top" to "bottom": Public authorities informed the public on notable risks or events through traditional broadcasting media (e.g. newspapers, radio, and television). The limited reach of available horizontal media (e.g. word-of-mouth, letters, and telephones) restricted any peer-to-peer communication. In the past decade, new information and communication technologies have greatly increased opportunities for collaboration and participation: Nowadays, many citizens use mobile device with wireless internet access to share freely various media through social networks[2] or more focused platforms for text messages,[3] images,[4] videos,[5] and maps.[6] Thereby, citizens have become providers of environmental information during crisis

* Corresponding author. Tel.: +39 033278 9242.
E-mail addresses: laura.spinsanti@jrc.ec.europa.eu, jrc.vgi.ff@gmail.com (L. Spinsanti), frank.ostermann@jrc.ec.europa.eu (F. Ostermann).
[1] Tel.: +39 033278 5544.

[2] e.g. Facebook, Google+.
[3] e.g. Twitter, Blogspot, Wordpress.
[4] e.g. Flickr, Picasa, Panoramio.
[5] e.g. YouTube, Vimeo.
[6] e.g. GoogleMaps, GeoCommons, MapBox.

0143-6228/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apgeog.2013.05.005

events, as several studies have observed (De Longueville, Annoni, Schade, Ostlaender, & Whitmore, 2010; Palen & Liu, 2007; Puras & Iglesias, 2009; Roche, Propeck-Zimmermann, & Mericskay, 2011).

Other studies (Al-Khudhairy, 2010; De Longueville, Smith, & Luraschi, 2009; Hughes & Palen, 2009; Liu & Palen, 2010; Schade et al., 2013) indicate when and how VGI can be most useful in a crisis context: During the response phase, affected citizens are on the ground before the first emergency response teams arrive. Instead of remaining passive victims, they can become now active collectors and distributors of information, thus acting as in-situ sensors (Goodchild, 2007). Their local knowledge could be especially valuable in providing near real-time localized and accurate updates, increasing situational awareness for emergency managers and peers searching for information. During the damage assessment in the recovery phase of a disaster, photographs and reports add to information from remotely sensed images and professional surveyors.

From the perspective of public authorities, the main challenge to using VGI is the lack of managerial control over the lineage of the information, and thereby an unknown reliability and trustworthiness (Jennex, 2010). As a result, most public agencies have been more reluctant in adopting social media information than non-governmental or volunteer organizations. The latter, however, face the issues of sustainability and scalability: There is no guarantee that for a given event there is a sufficient volunteer force.

Therefore, we propose a novel approach to filter VGI, evaluate its quality, and thereby improve its utility. It focuses less on the analysis of the source and content of an individual piece of information. Instead, it relies on geographic location and geographic context information to emulate two heuristics that humans use to deal with new information (Flanagin & Metzger, 2008): Expectancies ("What do I already know?") and social confirmation ("What do others say?"). To this end, the GeoCONAVI system queries authoritative geographic context information and clusters VGI in space and time. The system is implemented as a fully operational prototype based on prior research (Ostermann & Spinsanti, 2011).

This paper investigates the following main research questions:

1. How can geographic location improve the filtering and quality assessment of social media content?
2. What are the specific design requirements to utilize geographic context information?
3. What is the performance of the prototype system, and how well is adaptable to other crisis events?

The objective of the paper is to show how the GeoCONAVI approach works and that it produces valid and actionable output. We then discuss the requirements that need to be met in order for GeoCONAVI to work, showing that the approach is feasible for many types of applications and environments. Finally, the paper discusses the performance of our specific prototype and its adaptability to other types of crisis events. The paper structure is as follows: In the next section, we describe the overall system, the modules and the methods used. In the third section, we briefly present results and evaluation of the case studies, while the fourth section discusses the system evaluation and we answer the research questions. The paper concludes with a synthesis of the results and knowledge gained.

Methods and system architecture

This section presents the modules that make up the GeoCONAVI workflow and the corresponding software implementation. The objective was to implement a proof-of-concept prototype that runs autonomously for a substantial amount of time and allows the evaluation of the approach. We decided to investigate forest fires because of their increasing importance in environmental domain conservation, their seasonality and the availability of the European Forest Fire Information Service (EFFIS). We limited the geographic scope to four countries (Italy, France, Portugal and Spain) because they have frequent seasonal forest fires, the authors are able to read and understand content in these languages, and to reduce the challenges of character encoding. The following Fig. 1 shows the main processing phases for the social media information in the left column, the corresponding state of the VGI in the middle column, and the respective GeoCONAVI modules in the right column.

Fig. 2 below gives a more detailed overview of the system architecture, detailing modules and sub modules. The first three phases of Fig. 1 aligns from left to right instead from top to bottom. The top layer in Fig. 2 shows external data sources, while the implemented GeoCONAVI (sub) modules form the middle layer, and the lower layer is the data storage, implemented in an Oracle DBMS. The Disseminator has been implemented as a web-mapping interface (upper right).

The following sections describe each module in more detail.

Sensor

The GeoCONAVI Sensor is an opportunistic sensor. This means that it listens to a specific "frequency" of broadcasted information by citizens, but it does not have a particular interface where someone could provide information directly. This distinguishes it from portals where citizens can provide information for a specific purpose in a participative effort. The rationale for this is that persons in distress are unlikely to use an interface or infrastructure that they are not used to, instead relying on known services and social networks. There is no reason why some stakeholder could

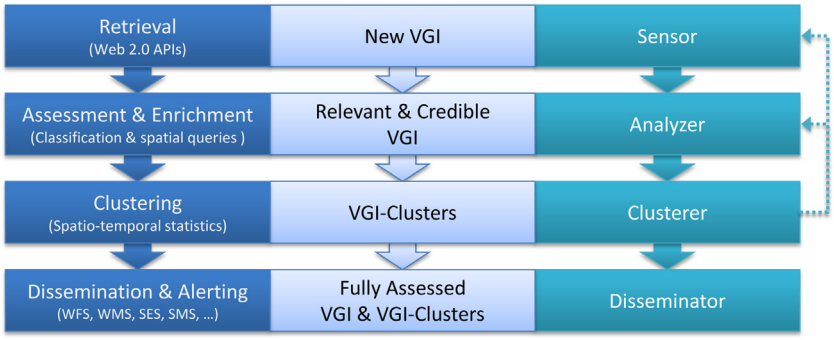

Fig. 1. Overview of Processing Steps.