Chapter 27
Classification of Traffic Events Notified in Social Networks' Texts

Ana Maria Magdalena Saldana-Perez
Instituto Politecnico Nacional, Mexico

Marco Antonio Moreno-Ibarra
Instituto Politécnico Nacional, Mexico

Miguel Jesus Torres-Ruiz
Instituto Politécnico Nacional, Mexico

ABSTRACT

It is interesting to exploit the user-generated content (UGC) and to use it with a view to infer new data; volunteered geographic information (VGI) is a concept derived from UGC, whose main importance lies in its continuously updated data. The present approach tries to explode the use of VGI by collecting data from a social network and a RSS service; the short texts collected from the social network are written in Spanish language; text mining and a recovery information processes are applied over the data in order to remove special characters on text and to extract relevant information about the traffic events on the study area; then data are geocoded. The texts are classified by using a machine learning algorithm into five classes, each of them represents a specific traffic event or situation.

INTRODUCTION

Many researching works implement procedures to use web data and users generated content on critical situations management, or to notify relevant events (Wakefield, 2013). Every day millions of publications, where people describe their environment, and ideas about many interesting topics are done at social networks. From such publications, it is possible to acquire data related to the urban factors that concern most citizens, in order to improve the people's activities and conditions, and with a view to structure smart cities. Web publications that have coordinates or provide geographic information, can be considered as VGI.

The main purpose of this work, is to obtain useful information for a further traffic prediction procedure, by integrating and processing VGI and authoritative data. The VGI data is collected from a social network, where participants are considered as intelligent sensors, which constantly communicate the environment changes they perceive; such reports are influenced by the people context.

Two of the main advantages of VGI, are its continuously updated data, and the easily data consultation. Such kind of source has not been deeply studied, most of the time its social characteristics have been analyzed over its technological usages.

The present approach makes use of VGI and authoritative data sources, to extract, analyze, classify and geocode data related to the traffic in the study area. The case of study is Mexico City, the largest metropolitan area in America, and the largest Spanish-speaking city in the world (Zimmerman, 2015).

The social network consulted is Twitter, also a RSS service is considered. In the analysis stage, a text mining and a recovery information processes are applied over the short texts, extracted from the social network and from the rss service, with a view to identify the traffic events described on the posts; on the classification stage, a Naïve Bayes algorithm is trained to classify the traffic events. The geocoding process is done through the use of the methodology proposed by Salazar et al. (2016). The events and the authoritative data, are related with the purpose to establish links, between the streets and the business located on the area where the traffic is presented; also a probabilistic analysis is applied. Finally, the events and the authoritative data are mapped to visualize the traffic distribution.

The paper is organized as follows: first the background of the work is shown, below the data treatment is described; then, solutions and recommendations are done; after the future research directions and conclusions are presented; finally, the additional readings are mentioned.

BACKGROUND

User generated content (UGC) is any publication on internet, done by users of web services such as blogs, wikis, forums, social networks, podcast and chats. It is used in many applications, including: researches, information and news spread, problems processing, disaster management, and collaborative mapping. UGC has originated some other concepts, volunteered geographic information (VGI) and crowdsourcing are two of them (Chard, 2015).

Crowdsourcing is the process of getting ideas, information, or work done, from a group of interested people; it has been a recurrent data and services source for some businesses and researches (Chard, 2015). Mobile devices as smartphones, mobile GPS, cartographic applications and social networks, make crowdsourcing possible.

According to Wen Lin (2013), VGI is composed of volunteered information generated by users, who have not a geographic specialized knowledge, but are interested on provide data with geographic characteristics; such data are employed on many web services as Open Street Map (OSM), WikiMapia, Google Maps, among others. This association between VGI and web services are the GeoWeb basis.

VGI makes possible the generation of new spatial information usages, such as collaborative mapping, georeferenced content, network vectorization, and collective ranking of places (Gouveia & Fonseca, 2008; Elwood, 2013; Sui, 2012; Goodchild, 2007). There exist web services that invite people to provide personal information, their location, or some other geospatial information, such as landmarks, points of interest (POI), and street segments; with the purpose to increment their data repositories, and

Related Content

An Examination of Verbal Aggression in Politically-Motivated Digital Discourse
Jean Mathieu Tsoumou (2021). *International Journal of Social Media and Online Communities (pp. 22-43).*
www.irma-international.org/article/examination-verbal-aggression-politically-motivated/298609

Gender Issues in the Diversity and Practice of Public Relations in the UAE Case Study of P.R. Male Managers and Female P.R. Practitioners
Badreya Al-Jenaibi (2011). *International Journal of E-Politics (pp. 35-56).*
www.irma-international.org/article/gender-issues-diversity-practice-public/55818

"I've Got a Situation and Would Appreciate Your Experience": An Extra-Organizational Virtual Community of Practice for Independent Professionals
Enrique Murillo (2012). *International Journal of Virtual Communities and Social Networking (pp. 52-80).*
www.irma-international.org/article/got-situation-would-appreciate-your/75779

Election Campaigns on the Internet: How are Voters Affected?
Jens Hoff (2010). *International Journal of E-Politics (pp. 22-40).*
www.irma-international.org/article/election-campaigns-internet/38967

Social Media Applications Promote Constituent Involvement in Government Management
Gerald A. Merwin Jr., J. Scott McDonald, John R. Bennett Jr.and Keith A. Merwin (2016). *Strategic Integration of Social Media into Project Management Practice (pp. 272-291).*
www.irma-international.org/chapter/social-media-applications-promote-constituent-involvement-in-government-management/145682