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

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 make them available for further analysis. OSM, Foursquare, and Twitter are some of those web services (Kunze, 2015).

Kunze (2015), proposes a VGI integration process, to generate estimation models of dwelling occupancy. In his approach, OSM and topographic data from Dresden city are used. When working with data from many spatial data sources, it is needed to correctly integrate such data on a GIS environment, and to consider some specific features, as the data precision, and the reference system applied over them (Flowerdew, 1991).

Most of the internet users, are part of the social networks; according with previous investigations 70% of them, access to internet by using mobile devices; and 73.4% do it through a computer (Castells, 2010).

Nowadays, internet users are interested on describe their environment, or give information about the events they perceive; on their publications people describe physical aspects of their surroundings, as the traffic or pollution situation, the security degree they consider appropriate for their neighborhood, socio-politic events, and some other topics of interest; also, they provide metadata such as coordinates, URL's, among others (Fleming, 2014).

User generated content published on social networks and VGI sources, has an explicit semantic value, since the content has been produced by people interested on provide their data for further analysis, unlike the other web content generated by people without a specific purpose (Resch, 2015).

Considering people as sensors, gives a wide variety of heterogeneous data, to researches that study the human behavior or the city dynamics, through the analysis of VGI and UGC data. The people's point of view is different from one person to another, and it is influenced by the person's living conditions, activities, education, ideas, and the geographic area where he or she lives (Resch, 2015).

DATA TREATMENT TO DETECT TRAFFIC EVENTS ON TEXTS

The present approach has been designed to obtain information about the traffic situation in the study area, by analyzing and processing VGI data. The proposed methodology, consist on 5 stages: the first, is the data acquisition; the second, are

Related Content

Towards Higher Software Quality in Very Small Entities: ISO/IEC 29110 Software Basic Profile Mapping to Testing Standards
Alena Buchalcevova (2021). *International Journal of Information Technologies and Systems Approach (pp. 79-96).*
www.irma-international.org/article/towards-higher-software-quality-in-very-small-entities/272760

Towards Higher Software Quality in Very Small Entities: ISO/IEC 29110 Software Basic Profile Mapping to Testing Standards
Alena Buchalcevova (2021). *International Journal of Information Technologies and Systems Approach (pp. 79-96).*
www.irma-international.org/article/towards-higher-software-quality-in-very-small-entities/272760

Communities of Practice as a Source of Open Innovation
Diane-Gabrielle Tremblay (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5027-5035).*
www.irma-international.org/chapter/communities-of-practice-as-a-source-of-open-innovation/184205

Methodologies for Modeling Gene Regulatory Networks
Shruti Mishraand Debahuti Mishra (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 426-436).*
www.irma-international.org/chapter/methodologies-for-modeling-gene-regulatory-networks/112353

Design of Graphic Design Assistant System Based on Artificial Intelligence
Yanqi Liu (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-13).*
www.irma-international.org/article/design-of-graphic-design-assistant-system-based-on-artificial-intelligence/324761