

Chapter 5

A Route Recommender System Based on Current and Historical Crowdsourcing

Marlene Goncalves

Universidad Simón Bolívar, Venezuela

Daniela Andreina Rodríguez

Universidad Simón Bolívar, Venezuela

Patrick Rengifo

Universidad Simón Bolívar, Venezuela

Ivette C. Martínez

Universidad Simón Bolívar, Venezuela

ABSTRACT

Due to the rise of the social networks it's possible to use techniques based on crowdsourcing to easily gather real-time information directly from citizens in order to create recommendation systems capable to employ knowledge that is shared from the crowd. Particularly, in Twitter, the users publish a big amount of short messages; however, to automatically extract useful information from Twitter is a complex task. In order to provide an informed recommendation of the current best route between two city points, this chapter introduces a workflow that integrates natural language techniques to build an vector of features for training two linear classifiers which obtain current information from Twitter, and integrates that information with historical information about possible routes using exponential smoothing; current and historical data to feed a route selection algorithm based on Dijkstra. The effectiveness of the proposed workflow is shown with routes between two interest points in Caracas (Venezuela).

INTRODUCTION

Over the last years, new ways of sharing information have been created with the purpose of the users be informed, and thereby provide users with more capacity to make a decision. Today, the Web allows to share any type of information about different subjects and events. Specifically, the social networks are utilized to share on real time events of great importance locally or globally.

The age of social networking has made possible to use techniques based on crowdsourcing in order to support recommendation systems able to utilize the knowledge that is shared on these social networks. In particular, the crowdsourcing consists in obtaining the collaboration of a great amount of people with

DOI: 10.4018/978-1-5225-0648-5.ch005

the objective of performing several tasks with their help. An example is reCAPTCHA (Von Ahn, Maurer, McMillen, Abraham, & Blum, 2008), which seeks to improve the optic system in charge of digitizing old books. In this system, the users help to recognize the words that are not correctly detected.

On the other hand, social networks represent a rich source of information to evaluate situations such as protests, traffic or natural disasters, because they are widely used to report events in real time. However, the automatic interpretation of this information is a difficult task. On the one hand, the information is expressed in natural language and on the other hand, the first step is to decide whether this information is relevant to the situation to be evaluated, and the second step is to determine what is actually saying about the situation.

In the Twitter social network, the users publish a good deal of tweets or short messages of 140 characters written in natural language. For this reason, automatically extract information from this social network is complex. First, the grammatical structure of the tweets is affected by the character restriction, which leads the user to use abbreviations and a particular lingo of the social network. On the other hand, the amount of tweets published daily makes difficult to detect if a tweet is relevant to a subject on which information is being sought. Finally, a tweet processing can be complicated, depending on what you want to measure. For example, in the case of traffic, a user may want to know, first, if there is traffic in a route and, then, quantify it.

Recently, several works have been realized about classifying text extracted from Twitter. In (Sriram, Fuhry, Demir, Ferhatosmanoglu & Demirbas, 2010), the tweets are classified in preselected categories, like News, Opinion and Events with the intention to simplify the information flow. Similarly, in (Lee, Palsetia, Narayanan, Patwary, Agrawal, & Choudhary, 2011), the tweets belonging to the Trending Topics are classified in their possible categories as these are hard to understand without previous knowledge of the subjects. Both attack the classification problem, but nevertheless, they not see the potential to use this classified information to make analysis and make predictions or recommendations.

The general objective of this chapter is to describe a recommendation system based on crowdsourcing with the purpose to collect documents from the Internet about some topic, process them and be capable to give recommendations on the subject in question. More specific, traffic is the selected topic, Twitter is the information source and the best route from a point to another is the recommendation.

Additionally, this work presents the traffic in Caracas as a case study. Caracas (Caracas, n.d.) is the capital of Venezuela and has more than 5,000,000 habitants on 1,820.5 square miles. Over the years, the city of Caracas has experienced significant population growth. This growth coupled with insufficient planning of expansion and maintenance of arterial roads and frequent accidents that occur have hampered the mobility of inhabitants and visitors who must drive through the city every day (Lizarraga, 2012). Particularly, the Simon Bolivar University is located on the periphery of the city and is isolated from the urban disturbance of Caracas. However, its access is limited due to the remoteness from the rest of the city and the routes to/from this university are not designed for vehicular traffic which owns the area. In addition, the travel to the university (or from the university) also includes the delays that can be found within the city, as well as problems from nearby roads. If to this we add unexpected variables such as protests which limit the free movement of the streets and avenues, then the arrival to the campus can be complicated such as we experienced in 2014 due to protests over the whole city (Buxton, 2014).

Currently, there are several mechanisms to know the traffic conditions in the city. One is Waze, a mobile application, which finds the quickest route from the current user's position to the desired destination. In addition to this data, in order to give directions to the chosen location, Waze utilizes user's GPS (Global Positioning System) data and reports. Nevertheless, Waze may not have enough users to report

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-route-recommender-system-based-on-current-and-historical-crowdsourcing/161962

Related Content

Creating Strategic Information for Organizations with Structured Text

Roberto Penteado and Eric Boutin (2008). *Emerging Technologies of Text Mining: Techniques and Applications* (pp. 34-53).

www.irma-international.org/chapter/creating-strategic-information-organizations-structured/10175

Logical and Physical Design of Spatial Non-Strict Hierarchies in Relational Spatial Data Warehouse

Ferrahi Ibtisam Ibtisam, Sandro Bimonte and Kamel Boukhalfa (2019). *International Journal of Data Warehousing and Mining* (pp. 1-18).

www.irma-international.org/article/logical-and-physical-design-of-spatial-non-strict-hierarchies-in-relational-spatial-data-warehouse/223134

Use of Risk Adjustment Models for Provider Reimbursements

Patricia Cerrito (2010). *Text Mining Techniques for Healthcare Provider Quality Determination: Methods for Rank Comparisons* (pp. 357-385).

www.irma-international.org/chapter/use-risk-adjustment-models-provider/36641

Mining Association Rules in Data Warehouses

Haorianto Cokrowijoyo Tjioe and David Taniar (2005). *International Journal of Data Warehousing and Mining* (pp. 28-62).

www.irma-international.org/article/mining-association-rules-data-warehouses/1755

Big Data Literacy: A New Dimension of Digital Divide, Barriers in Learning via Exploring "Big Data"

Dimitar Christozov and Stefka Toleva-Stoimenova (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 2300-2315).

www.irma-international.org/chapter/big-data-literacy/150266