Information Systems 54 (2015) 147–155

Contents lists available at ScienceDirect

Information Systems

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

Mining citizen emotions to estimate the urgency of urban issues

Christian Masdeval [a,*], Adriano Veloso [b]

[a] Dataprev, Brazil
[b] Universidade Federal de Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form
30 May 2015
Accepted 22 June 2015
Available online 26 June 2015

Keywords:
Emotion classification
Networked citizenship
Crowdsourcing

ABSTRACT

Crowdsourcing technology offers exciting possibilities for local governments. Specifically, citizens are increasingly taking part in reporting and discussing issues related to their neighborhood and problems they encounter on a daily basis, such as overflowing trash-bins, broken footpaths and lifts, illegal graffiti, and potholes. Pervasive citizen participation enables local governments to respond more efficiently to these urban issues. This interaction between citizens and municipalities is largely promoted by civic engagement platforms, such as See-Click-Fix, FixMyStreet, CitySourced, and OpenIDEO, which allow citizens to report urban issues by entering free text describing what needs to be done, fixed or changed. In order to develop appropriate action plans and priorities, government officials need to figure out how urgent are the reported issues. In this paper we propose to estimate the urgency of urban issues by mining different emotions that are implicit in the text describing the issue. More specifically, a reported issue is first categorized according to the emotions expressed in it, and then the corresponding emotion scores are combined in order to produce a final urgency level for the reported issue. Our experiments use the SeeClickFix hackathon data and diverse emotion classification algorithms. They indicate that (i) emotions can be categorized efficiently with supervised learning algorithms, and (ii) the use of citizen emotions leads to accurate urgency estimates. Further, using additional features such as the type of issue or its author leads to no further accuracy gains.

1. Introduction

City maintenance is extremely expensive, since it involves monitoring and fixing a variety of complex issues related to public safety, environmental problems, and quality of life. In particular, monitoring urban issues (e.g., potholes, damaged street signs, graffiti, street light issues, damaged trees, park maintenance) usually requires a large number of employees working on a permanent basis. Alternatively, crowdsourcing can promote a participatory way of monitoring urban environments, and thus municipality's resources can be redirected to effectively fixing the reported issues. As citizens are increasingly equipped with smartphones, they are also increasingly able to perform pervasive crowdsourcing at urban level, that is, to report urban issues with GPS location directly to the city's appropriate department. This model of crowdsourced city enables not only better resource allocation, but also fine-grained monitoring capabilities.

Early examples of crowdsourcing platforms for monitoring urban issues include See-Click-Fix,[1] FixMyStreet,[2]

* Corresponding author.
E-mail addresses: christiancleber@dataprev.gov.br (C. Masdeval), adrianov@dcc.ufmg.br (A. Veloso).

[1] seeclickfix.com
[2] www.fixmystreet.com

http://dx.doi.org/10.1016/j.is.2015.06.008
0306-4379/

CitySourced,[3] and OpenIDEO.[4] These platforms enables citizens to report issues by entering a description of what needs to be done, fixed or changed. As an example, Fig. 1 shows a report extracted from seeclickfix.com/issues/1273509-turns-on-and-off-all-night. Citizens may also vote for specific issues, thus endorsing the wish that the problem is solved. While the number of votes an issue receives ultimately reflects its urgency to be solved, acquiring a significant number of votes may take several days or weeks, leading to ineffective and late responses. In order to become more responsive, government officials must be able to prioritize more urgent issues, and thus estimating the number of votes an issue will receive may help officials to better meet the needs and concerns of the citizens.

We here are interested in estimating the urgency of a reported issue by the number of votes it receives, and we investigate the extent to which the textual description by itself determines the urgency of the reported issue. We observed that issues are usually described in an opinative way, but with different viewpoints and inclinations. Thus, we hypothesize that emotions[5] that are implicit in the textual description of an issue (i.e., fear, distress, shame) may be a good evidence of its urgency. We propose algorithms for categorizing the reported issues according to the emotions expressed in their textual description, and then we exploit the corresponding emotion scores in order to estimate the number of votes the reported issue will receive. Since such estimates are made based solely on the text used to describe the issue, government officials may have an immediate view of the urgency of the issues, enabling them to prioritize solving problems that are more urgent according to the citizens.

Experiments using real data obtained from the SeeClickFix hackathon (www.meetup.com/software/events/138126482) demonstrated the effectiveness of exploiting citizen emotions for the sake of estimating the urgency of urban issues. Specifically, we evaluated a set of emotion classification algorithms and conclude that estimates based on emotions implicit in the textual description are significantly more accurate than estimates obtained directly from the text. Also, using additional features related to the author of the issue and to the type of the issue leads to no further significant gains.

2. Related work

2.1. Participatory sensing and crowdsourcing in urban spaces

The revolution in communication and crowdsourcing [4,5] technologies has been changing not only the daily lives of people but also the interactions between governments and citizens. In recent years, many efforts have been made in order to understand these interactions. Kanhere [13] provides a comprehensive overview of these efforts.

For Chun et al. [8], the final stage of transformation of "open government", the so-called Government 2.0, implies that information should flow not only from the government to the citizens but also from citizens to the government and among citizens. It will require principles, functions and technological enablers to lead to a transformative and participatory model. Participation encourages the public engagement by increasing opportunities for the public to participate in policy making. They also define Web 2.0 as a collection of social media through which individuals are active participants in creating, organizing, editing, combining, sharing, commenting, and rating Web content as well as forming a social network through interacting and linking to each other and state that the required functions of Government 2.0 can be achieved by adopting the Web 2.0 technologies.

Kavanaugh et al. [14] make an exploratory study of using traditional social media content as Twitter, Facebook, Flickr and YouTube, to detect, in real time, spikes in activity related to issues concerning public safety. Analyzing information from multiple social media sources should be possible to identify convergence situations (meaningful patterns and trends) and it will help, among other things, in treat crisis situations, from the routine (e.g., traffic, weather crises) to the critical (e.g., earthquakes, floods). This is difficult once a lot of noise should be filtered to make the information useful and reliable. To accomplish its goal, they employed mining techniques covering multiple media types (i.e., text, audio, image, and video). Also, they developed tools to recognize events and to help the visualization of the "big picture" of social media activity and content, and changes in both over time.

Handte et al. [10] proposed methods for crowd density estimation for improving public transportation. Specifically, the authors proposed approaches to estimate the number of passengers in a vehicle. Artikis et al. [1] presented a system for heterogeneous stream processing and crowdsourcing supporting intelligent urban traffic management, and Schnitzler et al. [21] provide an overview of an intelligent urban traffic management system, including approaches for dealing with complex events such as congestions. Litou et al. [15] proposed an approach for emergency notification using online social networks. The proposed approach selects the most efficient routes to maximize the information reach.

2.2. Sentiment analysis and emotion classification

Sentiment analysis methods are typically divided into two broad categories: those that are based on lexical approaches and those that are based on machine learning algorithms. One advantage of learning-based methods is their ability to adapt and create trained models for specific purposes and contexts. In contrast, lexical-based methods make use of a predefined list of words, where each word is associated with a specific sentiment. Next we will focus on learning-based methods, since this is the approach we will follow in this paper.

[3] www.citysourced.com
[4] openideo.com
[5] Although there is not an exact definition for the concept of emotion, most agree that emotions are reactions to events deemed relevant to the needs, goals, or concerns of an individual.