Online Social Networks and Media 5 (2018) 51–60

Contents lists available at ScienceDirect

Online Social Networks and Media

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

Can we monitor the natural environment analyzing online social network posts? A literature review

Philip Lei[a], Gustavo Marfia[b,*], Giovanni Pau[b,c,d], Rita Tse[a]

[a] Macao Polytechnic Institute, Macao
[b] Alma Mater Studiorum, University of Bologna, Italy
[c] University Pierre and Marie Curie, Paris, France
[d] University of California, Los Angeles, US

ARTICLE INFO

Article history:
Received 3 November 2017
Revised 7 December 2017
Accepted 14 December 2017
Available online 28 February 2018

Keywords:
Environmental sensing
Online social networks
Human perception
Wellbeing

ABSTRACT

A number of works have addressed the question of assessing the status and the $\frac{2}{4}$ quality of the environment through the lens of Online Social Networks (OSNs). These contributions fall in the area of human-centric sensing, area specialized in using what people spontaneously say on social media to detect the occurrence of given events. Research in this area has exhibited interesting results, regardless of the accuracy of sensing operations. In fact, in some cases it is possible to corroborate the information extracted from OSN posts with the ground truth obtained from specialized hardware sensors. In others, the information extracted from OSNs does not reveal true environmental conditions. Nevertheless, OSNs may help shed light on the sensitivity of human beings to a wide variety of environmental phenomena. We here review the work that has been published to this date. In particular, we provide a survey that may benefit both environmental and computer scientists, as this work aims to show where we stand in the understanding of the complex relationship between human beings and the natural environment, when this is mediated by OSNs.

1. Introduction

Understanding the dynamics and the state of health of the natural environment has always been one of the main interests of people, in all historical times. Almost 2000 years ago, Pliny the Younger wrote a description of the eruption of the Vesuvius which is still nowadays studied by millions of students from all around the world [1]. The interest for the study and analysis of environmental dynamics is as live as ever, as new awareness for its state of health is strong, being an important component of human wellbeing [2].

Nevertheless, such type of interest roots in the one normally exhibited by people for weather forecasts, as well as in the fear which instead natural disasters or pollution hazards trigger. In fact, regardless of the specific phenomenon and where and when people discuss it, assessing the status and predicting the dynamics of the natural world, in any of its aspects, is an age-old problem of statistical inference. Even simply knowing whether it will rain or not on a short notice may attract much attention: harvesting, warfare, trips and outdoor sporting events often depend on it [3]. Before the Grand Prix, one of Formula One pilots' most-discussed arguments is the weather, as reliable forecasts are key to winning a race. Such type of interest is exacerbated when considering natural hazards and pollution spillovers whose effects can harm properties and human lives [4]. For this reason we have seen in the past decades a proliferation of different sensors aiming at quantifying specific physical phenomena [5–7]: temperature, wind speed, humidity, ozone layer thickness, chemicals' concentrations, etc. Many of such quantities are then reported by the news or specific websites, providing the population with a representation of environmental conditions and of how they may be evolving.

The reporting and discussion of any of the aforementioned phenomena has however changed since the introduction of Web 2.0 paradigms, in particular with the widespread use of Online Social Networks (OSNs) [8]. OSNs have fostered, since their birth, new lifestyles, habits and ways of communicating. Personal communications now follow a one-to-all information flow, allowing posts and comments to be read, answered and reposted by a multitude of users. In addition, posts can in principle touch upon any topics, as users can spontaneously write anything with no censorship, often acting as a human-based sensor [9,10]. All this has been exploited by many works that have appeared in the scientific literature on human-centric computing to assess the following

* Corresponding author.
E-mail address: gustavo.marfia@unibo.it (G. Marfia).

https://doi.org/10.1016/j.osnem.2017.12.001
2468-6964/

idea: the big data of spontaneously shared posts, that typically exposes what a person thinks and how s/he feels and behaves, could be an interesting source of environmental related information. In essence, a person, when posting on an OSN, may reveal some information pertaining the surrounding environment, thus acting in a way that resembles how a sensor works.

We should not be mistaken thinking of OSNs as a new sensing technology, which may seamlessly be integrated with legacy ones. In fact, whenever people post information pertaining the natural environment, they throw a representation of what they perceive, rather than a measurement quantifying the magnitude of a physical phenomenon. Different people may perceive the same physical phenomenon in different ways, for a number of different reasons. A thoroughly studied example is the one related to temperature: women are more sensitive to temperature (mainly cool) and less to humidity than men [11–13]. In many cases, hence, the many existing different ways of perceiving the world may represent a source of uncertainty which is difficult to factor out.

On the other hand, it may be possible to find natural events whose perception is still strongly subjective, but also affected by how much a person feels threatened, i.e., close to given critical ecological thresholds [14]. Non-cognitive environmental aesthetics, for example, a branch of aesthetic philosophical studies, explains human reaction to these events in terms of primordial, perceptive and emotional states. In [15] the authors argue: *those...who have at heart the welfare of humans or nonhumans react to environmental degradation with dismay*, stating the existence of a shared attitude towards given classes of events. Contrasting conclusions arrive from other areas of research, however. Recent research on mass media, for instance, has observed that the media may mistakenly represent the severity or frequency of natural hazards, thus altering the way in which those events are perceived and understood [16]. In [17], the authors analyze the psychological dimensions of air pollution, revealing how its perception may be independently influenced by the 'social class' of a person and by the fact that people are reluctant to acknowledge the existence of potential risks in the neighborhoods where they live.

Nevertheless, an observation of what people spontaneously say regarding a given phenomenon, regardless of its correspondence to the ground truth measured by a sensor, can provide the research community with important insights concerning human wellbeing [2]. For this reason, unlike other surveys which take different or broader perspectives on the study of the existing body of work concerning OSN-based sensing [18–27], we here concentrate on those works that seek to assess the status or the quality of the environment. Additionally, we only look at those works that have based their findings on the analysis of spontaneous posts, as opposed to those participatory or citizen sensing initiatives which require some degree of user involvement [28,29]. Compared to the corpus of works that, instead, study the use of OSNs to enhance emergency situation awareness, we take a different perspective, as such works only focus on critical events, which may be of any kind, ranging for example from nuclear incidents, to war events, to natural disasters, in order to aid authorities during search and rescue operations [30–36]. In essence, we here provide an updated critical map and overview of what is available to this date, as well as suggestions of what may still be missing, for the benefit of data engineers and scientists, as well as for environmental scientists, psychologists, philosophers and for all those interested to understand how human beings perceive and report their perception of the environment on OSNs.

This survey is organized as follows. In Section 2 we explain how we have chosen the works that are here discussed, in addition to how we here analyze their contributions. Section 3 provides a high level overview of the analyzed body of literature, for the benefit of both technical and non-technical audiences. Technical details are analyzed in Section 4. Finally, a critical discussion is provided in Section 5, where also possible future directions of work are delineated.

2. Methods

We here describe the criteria applied to select the scientific contributions of interest and then move on to explain how such works have been grouped and compared. We also technically discuss the data science approaches employed in these works, exhibiting which have been the paths taken to detect environmental events or to measure its variables.

2.1. Literature selection

The scientific literature that has been selected and analyzed in this work fits three different requirements.

The first is the interest for some quantity related to the natural environment: either concentrating on the natural phenomena and metrics that characterize its state (e.g., storm, aurora, temperature, humidity, etc.) or on those events and variables that denote its health status (e.g., air/water/land pollution). In particular, we considered geological (e.g., earthquakes, volcanic activities, landslides, etc.), oceanographic (e.g., breaking waves, tsunamis, etc.) and meteorological (e.g., storm, aurora, etc.) events [37], as well as the quantitive variables measured within the United States Department of Agriculture Water Erosion Prediction Project [38]: solar radiation, temperature, wind speed, average dew point temperature and precipitation. In addition to the aforementioned variables, we also selected pollution-related ones.

The second is that of basing the proposed study on posts that have been spontaneously written on OSNs.

Finally, we only consider those works which compare the results obtained from OSNs to the ground truth, defined in terms of objective physical measurements (e.g., temperature value, pollution concentration value, etc., obtained from legacy sensors).

2.2. Taxonomies for the analysis of environmental phenomena

To simplify the discussion of their contributions, we apply two different taxonomies to the body of work selected in Section 2.1, adopting two different perspectives.

With the first one, which resorts to a top-down approach, we acknowledge the role that emotional reactions have in OSN-based environmental sensing. Scientific contributions have been hence classified based on the rarity and severity of the event. Natural (e.g., earthquakes) and human-generated (e.g., radioactive material spillover) hazards clearly fall in one group. Everyday ones, such as a rainy day, in another. This distinction will be adopted when discussing papers in Section 3.

With the second one, instead, we concentrate on the data categories that have been analyzed by the scientific community and on how these have influenced their results. Adopting this method in Section 4, we group scientific contributions according to the mathematical nature of the observed variable. In particular, we distinguished those works that assessed categorical variables, such as the detection of an event (e.g., thunderstorm), opposed to those which measured quantitative ones (e.g., temperature and pollution concentration values).

2.3. Data science techniques

The analysis proceeds with an overview of the initial set of data and the analysis of the algorithmic components considered in each work. To this aim, we break down our analysis in terms of the:

1. Environmental phenomenon of interest;