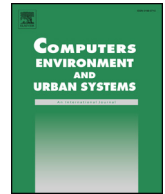




Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Review

Mining location from social media: A systematic review

Kristin Stock

Massey University, Private Bag 102904, North Shore, Auckland 0745, New Zealand

ARTICLE INFO

Keywords:
Social media
Geographic information

ABSTRACT

During the last ten years, a large body of research extracting and analysing geographic data from social media has developed. We analyse 690 papers across 20 social media platforms, focussing particularly on the method used for extraction of location information. We discuss and compare extraction methods, and consider their accuracy and coverage. While much work has adopted location information in the form of coordinates in message metadata, this approach has very limited coverage in most platforms and reports on posting location rather than message location or the location that the message refers to (geofocus). In contrast, a wide array of other approaches have been developed, with methods that extract place names from message text providing the highest accuracy. Methods that use social media connections also provide good results, but all of the methods have limitations. We also present analysis of the range and frequency of use of different social media platforms, and the wide range of application areas that have been addressed. Drawing on this analysis we present a number of future areas of research that warrant attention in order for this field of research to mature.

1. Introduction

The potential for social media to provide useful geographic information to either replace or augment traditional methods of data collection has been recognised for some years. In that time, a large number of research efforts have explored this potential with applications including health, disaster management, tourism and recreation, environmental monitoring, crime, civil unrest and marketing.

In this paper we provide a systematic literature review of papers across the field, identifying 690 papers within scope, analysing their content in order to compare different aspects of the research and identifying gaps and future research potential, particularly focussing on three aspects. Firstly, we review the different social media platforms that have been used for extracting geodata in the published literature. There is a clear preference for Twitter over other platforms, and we discuss the reasons for this and the potential for the increased use of other platforms to extract data that is not currently being used. Secondly, we explore the methods used to extract location information from social media. While use of metadata geotagging is the most common method, it has a number of limitations, and other methods including text mining, user profiling and different kinds of inference have been developed. We discuss these methods, their use and advantages and disadvantages, analysing accuracy and coverage achieved by each method. Thirdly, we review the impressive array of applications that have been addressed with data extracted from social media, and discuss the dominance of different application areas. Finally, we

propose future research directions to cover gaps in the current work, and to enable this research field to reach maturity.

The organisation of the paper is as follows. Section 2 discussed previous reviews that have been conducted on social media location data and related areas. Section 3 describes the methodology used for the systematic literature review and presents the research questions. Section 4 provides analysis of the social media platforms used to extract geographic data. Section 5 discusses and compares specific methods of location extraction, providing detailed discussion about the alternative approaches. Section 6 analyses the application domains used in the research papers surveyed and Section 7 discusses future research directions.

2. Literature review

A number of reviews have previously been completed in the field of social media location extraction, exploring different aspects. In this Section, we discuss firstly those reviews that address social media generally (not specifically spatial data), then those addressing location-based services and image analysis. We then consider those that focus more directly on extraction and analysis of geographic information, including reviews that focus on analysis focussed reviews, those that address VGI and those that review papers that address the use of specific application areas to extract geographic information. This summary then indicates the gap that we intend to fill with this paper.

Reviews that have looked at social media generally (not specifically

E-mail address: k.stock@massey.ac.nz.

<https://doi.org/10.1016/j.compenvurbysys.2018.05.007>

Received 16 March 2018; Received in revised form 17 May 2018; Accepted 18 May 2018
0198-9715/ © 2018 Elsevier Ltd. All rights reserved.

at spatial aspects) include [Batinca and Treleven \(2015\)](#), who provide a survey of social media technologies, methods for analysis and APIs as a tool for social scientists, but differ from our work in that they do not provide any detail about location, and [Hua et al. \(2012\)](#), who focus on approaches to extraction of content from Twitter, providing a section on detection of current location, briefly reviewing a few papers.

In the area of mobile and location-based social media, [Kaplan \(2012\)](#) surveys mobile marketing, and distinguishes between the different ways that marketing strategies consider location and time, but does not describe any aspects of location extraction in detail, and [Bao et al. \(2015\)](#) review recommendation systems for location-based social networks. Their review is mainly focussed around the approaches and methods used for making recommendations, and on location-based social networks (LBSN) in which places (e.g. venues) are first class citizens, enabling user check in. Although they refer to the ways in which geotags and other forms of location are used, they do not discuss this in detail.

Reviews that focus on image analysis include [Liu \(2011\)](#), who reviews a set of papers that use geographic information to analyse images, sometimes in combination with other information (e.g. image tags); [Yanai \(2015\)](#) provides a summary of analysis of web images, touching on the use of image analysis to infer location in the context of Flickr and [Luo et al. \(2011\)](#) discuss some approaches to determine the location of photos, reviewing several interesting papers in this area. [Zheng et al. \(2011\)](#) also discuss georeferencing from an image point of view, reviewing approaches to location landmarks and more general locations from photos.

Moving towards a greater emphasis on geographic information exclusively, [Stieger et al. \(2015\)](#) conduct a systematic literature review on the use of Twitter for geospatial analysis. They investigate 92 papers, examining discipline of authors, application domain, time of publication, type of data extracted and broad category of paper, and then look in more detail at the kind of analyses performed. They address the ways in which location information is extracted to some degree, but focus on the ways in which the location information has been analysed.

[Senaratne et al. \(2017\)](#) comprehensive review into quality issues in VGI points out some of the issues involved in using both active and passive (i.e. social media) VGI, and some of the approaches that have been developed to deal with these issues; and [Yap et al. \(2012\)](#) describe some of the requirements for a successful VGI-based LBS, including privacy, trust and information classification functions. They do not discuss details of location representation or extraction. [Goodchild and Li \(2012\)](#) propose three different approaches to dealing with quality: crowdsourcing, in which other contributors correct the errors of their peers; social, in which moderators police or verify contributions and geographic, in which spatial patterns can be used to identify unlikely or inconsistent contributions. Although this latter work applies more generally to all VGI, these approaches could be applied to data extracted from social media.

A number of studies review literature on the use social media in particular application areas, including [Guy et al. \(2011\)](#), who address the use of social media in disease surveillance. They do not discuss methods for extraction of location from social media in this context, but they acknowledge the need to “...determine the effectiveness of geo-location in garnering real-time estimates of ILI (influenza-like illness)” (p.5). Similarly, [Velasco et al. \(2014\)](#) explore the use of social media type approaches in disease surveillance, but do not discuss methods to extract location. [Horita et al. \(2013\)](#) discuss the use of VGI in disaster events and provide an overview of disaster types and the phase in which VGI is used. Their study is wider than social media, also including more active methods of crowdsourcing, with 6 of the 21 papers they summarise using social media. [Imran et al. \(2015\)](#) also focus on disaster events, providing a review of methods for processing social media messages in disaster situations. They only address location briefly.

[Klonner et al. \(2016\)](#) review papers looking at the use of VGI in the preparedness and mitigation phases of a disaster, but do not discuss extraction of location information. [Leung et al. \(2013\)](#) review papers on the use of social media for tourism and hospitality, but do not address location. [Yue et al. \(2014\)](#) describe data collection options to study trajectory-based travel behaviour, one of which is social media, and identify several studies.

Most relevant to our work, [Ajao et al. \(2015\)](#) conduct a review of location extraction approaches that have been used in Twitter, identifying seven different types of location indicator (tweet content, geotag, social networks; user profile; geotag; third party sources (for geocoding and reverse geocoding); time zones and web snippets. They then discuss the way Natural Language Processing (NLP) (specifically Named Entity Recognition [NER]) and gazetteers have been used to extract location. Our work is very closely related to this previous survey, and builds on it.

Our work differs from the previous work in that we consider a wider view, looking across social media rather than only focussing on Twitter; identifying the differences and gaps across social media platforms and studying location extraction approaches in detail. We also provide a systematic (quantitative) review which offers figures regarding the use of different approaches. Finally, we summarise the range of applications to which this approach has been applied. We also go beyond much of the previous work in identifying future research directions required in order to make use of a broader range of available data and fully realize the potential of this research field.

3. Methodology

Our systematic literature review follows the methodology described in [Kitchenham and Charters \(2007\)](#) and [Kitchenham et al. \(2009\)](#). In addition, even though we do not consider our work to constitute a scoping study, our review shares some goals in common with [Arksey and O'Malley's \(2005\)](#) four reasons for conducting a scoping study: namely that we aim to “examine the extent, range and nature of research activity” (p.21) on the use of spatial data in social media and that we aim to identify gaps in the existing literature, specifically in relation to potential ways to exploit social media that have not yet been considered.

We address the following broad research questions:

RQ1. : From which social media platforms has geographic data been extracted?

RQ2. : What methods have been used to extract location information from social media?

RQ3. : Which domains, sub-domains and research questions has geographic data extracted from social media been used to address?

The selection of these research questions was motivated firstly by the goal of maximising the opportunities offered by social media data for geographic mapping and analysis. Our intention was to determine whether there were social media platforms that are popular among users, and that contain significant amounts of data, but that have been neglected in the literature, and similarly to determine whether there were obvious gaps or under-presentation in particular methods for extracting location. A scan of the existing literature suggested the dominance of geotagged data from Twitter, but we wanted to confirm whether this was the case, and to highlight opportunities to exploit, compare and evaluate other data sources and location extraction methods. Similarly, we were interested in gaps in the research in particular application areas, and whether there were opportunities for new investigations that had not yet been addressed.

Our initial selection of candidate papers was achieved through

Download English Version:

<https://daneshyari.com/en/article/6921830>

Download Persian Version:

<https://daneshyari.com/article/6921830>

[Daneshyari.com](https://daneshyari.com)