# Spatiotemporal Analysis of User-Generated Data on the Social Web

Soha A. E. A. Mohamed[1] and Alia I. Abdelmoty[1]

[1]Cardiff University, School of Computer Science and Informatics, Queen's Buildings,

5 The Parade, Roath, Cardiff CF24 3AA, UK

Tel: +44 (0)29 2087 4812 | Fax: +44 (0)29 2087 4598

S.A.A.Mohamed@cs.cardiff.ac.uk

**Summary:** The vast increase in using GPS-enabled devices and the exponential growth in using social media sites such as Twitter[1] are enabling users to geographically annotate their personal resources as well as their opinions and experiences on the Web. The result is a large volume of ever-increasing spatiotemporal records of information related to users and their existence in space and time. Mining this data to extract the implicit semantics would give insight into the characteristics of users and place and interactions between them. This work reviews recent works in this area and presents some results from an experiment to analyse both the spatial and temporal aspects in this type of data.

## 1. Introduction

Data on the Social Web is a rich source of dynamic spatiotemporal information related to user's association with geographic places over time. Popular social networking applications, such as Twitter, Facebook and Foursquare are allowing users to geo-tag their data and resources, register their presence in a geographic place as well as record their activities and experiences. As a result, vast amounts of data are now being collected by these applications. Inherent in this data are user and place-related information that can be analysed to understand characteristics of users, places and their inter-relationships.

The value and availability of this data have recently spawned a surge in research works aiming to analyse the data with a view of understanding user behaviour and place characteristics. Several challenges need to be addressed, related to the nature of the data and the collection processes. For instance, geo-tagging is normally enabled by the user who controls when and in which applications the function is used, resulting in possibly intermittent recording and incomplete data. Also, GPS locations recorded by the applications may only allow a certain degree of accuracy to be visualised by other users of the application, resulting in vague association of users with the region they are located in.

---

[1]www.twitter.com
[2]www.yelp.co.uk

In this paper, a review of related work in the area of spatiotemporal analysis of Social Web data is given. Works are categorised into those that focus on user analysis and those that are concerned with place understanding. Furthermore, to demonstrate the content and value of this data, an experiment was developed to analyse data collected from Twitter over the period of the London Olympics in July and August 2012. Some sample data analysis is described to illustrate the relevance of the content of the data to the geographic locations they are generated in. This work is part of a project aiming at integrated mining of this type of social data to derive an understanding of relevance of place to users on the Web.

## 2. Related Work

Three principal dimensions of Social Web data can be studied: users, who generate the data and who communicate with other users; place, where the data are created by users, and time, when the data are created by users. Related work can be classified according to the emphasis of analyses on these three axes to those concerned with studying place and those focussing on user analysis.

### 2.1 User-related Analysis

An empirical study of user activity patterns in Foursquare[3] was developed in (Noulas el al., 2011), where checkin dynamics were analysed and spatiotemporal patterns of user activity were examined by comparing the number of checkins across the ten most popular place categories during weekdays and weekends. It was shown how checkins provide means to uncover daily and weekly patterns of user behaviour and recurrent transitions between different activities.
In a similar study of Foursquare data, (Chorley el al., 2011) and (Colombo el al., 2012) analysed the relationships between users and their checkin patterns and reported a correlation between visited places and friendship networks. In addition, a logical conclusion was noted that a significant percentage of users tend to make frequent and regular visit to a limited number of venues.

The above approaches considered only the presence of users in a place. Other works considered the content of data generated by users in a place in an attempt to analyse more deeply the activities carried out by users in the places they visit. In order to build a corpus of potential activities for each location, (Dearman & Truong, 2010) aimed to identify a set of potential activities that are supported at a person's current location and other locations they visit by processing community-authored reviews on Yelp[2].

Hapori is a context-driven local search framework for mobile phones that adapts to users' needs by analyzing context such as time, weather and the activity of the user. Hapori also builds behavioural profiles of users and extracts behavioural similarities with others in the broader community of local search users (Lane et al.,2010).

Similarly, (Abel et al. 2011) investigated user modelling strategies for inferring personal interest profiles from Social Web interactions. Twitter was used for conducting the analysis of individual activities over time. Data were collected on the Egyptian revolution (January 25, 2011) and were used to analyse how the interests of users in a topic changed over time.

---

[3]www.foursquare.com

## 2.2 Place-related Analysis

Here approaches to analysing the social data focus on understanding place and its characteristics, as well as related events in a place, using both the user and time dimensions of the data. (Tatsuya et al, 2010) attempts to discover "folks' behaviour patterns" using data from Twitter. An observation system, capable of collecting geo-tagged and time stamped posts from a specific region was developed and used to analyse the change in user population in different regions of space over time.

Mapping of the city can be achieved to reflect the collective behaviour of its residents by analysing patterns of people's movements through a city, (Cranshaw J. el al., 2012). In addition, (Kim el al., 2011) proposed a method to recognize spontaneous clusters of users associated with a location in real time. The proposed scheme collects and analyses geo-tagged Tweets and identifies clusters using the k-means algorithm.

Inference of "home" locations using the tweet message contents was presented in (Mahmud el al. ,2012). An algorithm for predicting locations of Twitter users from tweet contents and external geographic knowledge sources is proposed that uses an ensemble of statistical and heuristic classifiers to predict locations that can be considered as home.

Social data are also a rich resource of information regarding events happening in a place. proposed a scheme to discover events occurring in a place by frequent analysis of geo-tagged tweets and then feeding back the information on spontaneous events emerging in a place to users in the vicinity of the place. (Calabrese el al., 2010) observed the correlation between the origin of people attending an event and the type of the event. Based on the analysis of 1 million cell phone traces, results showed that people who live close to an event are preferentially attracted by it. In (Gruzd el al., 2011) and (Thelwall el al., 2011) sentiment analysis of data associated with events was carried out and was shown to be a useful indicator and differentiator of event characteristic.

Research in this area is fairly new and many research questions still remain open. In particular, understanding the semantics of the data and its dimensions that are relevant to a user and to a place. How to model the change in these data elements over time and reflect the change in both user and place information? What inter-relationships can be inferred using this data between users and places?

# 3. Twitter Experiment

The main goal of this experiment is to analyse the spatio temporal data through collecting data from twitter. This experiment is an initial step for characterising places from content of the tweets, and the tweet behaviour. Besides, it will help us to find the relationship between the users, the places they visit and the time of their visits.

Data were collected over the period of the London Olympics event (from 25-Jul-12 to 24-Aug-12). Different sporting events were taking place in several cities in Great Britain. A large international community of users as expected to visit both cities in this period, in addition to the local population. Data was collected frequently (every hour) for specific sporting events from different venues in both cities. Each record of the database contains userID, tweet, longitude, latitude and timestamp.

## 3.1 Analysis

Hashtags or # symbols are used in Twitter to mark keywords or topics in a tweet, where a Tweet is a message of not more than 140 characters. It was created originally by Twitter users as a way to categorize messages. Here, Hashtags were analysed in the tweets collected from three different venues.

Figure 1 shows a comparison between the hashtags extracted from the tweets made in three venues. The first venue is the Millennium Stadium in Cardiff, mainly hosting football games, the second venue is Excel in London, hosting boxing, fencing, judo, taekwondo, table tennis, weightlifting, and wrestling games and the third venue is Wimbledon which hosts tennis.

Excluding the common event hashtags like 'olympics2012', 'Olympics', and 'London2012', the figure reveals how the place aspects are prominent in the data collected. In addition, reference to individual sporting events is noted and can be distinguished in all tweets. Consequently, geographic place names were also used as tags to explicitly distinguish venues.
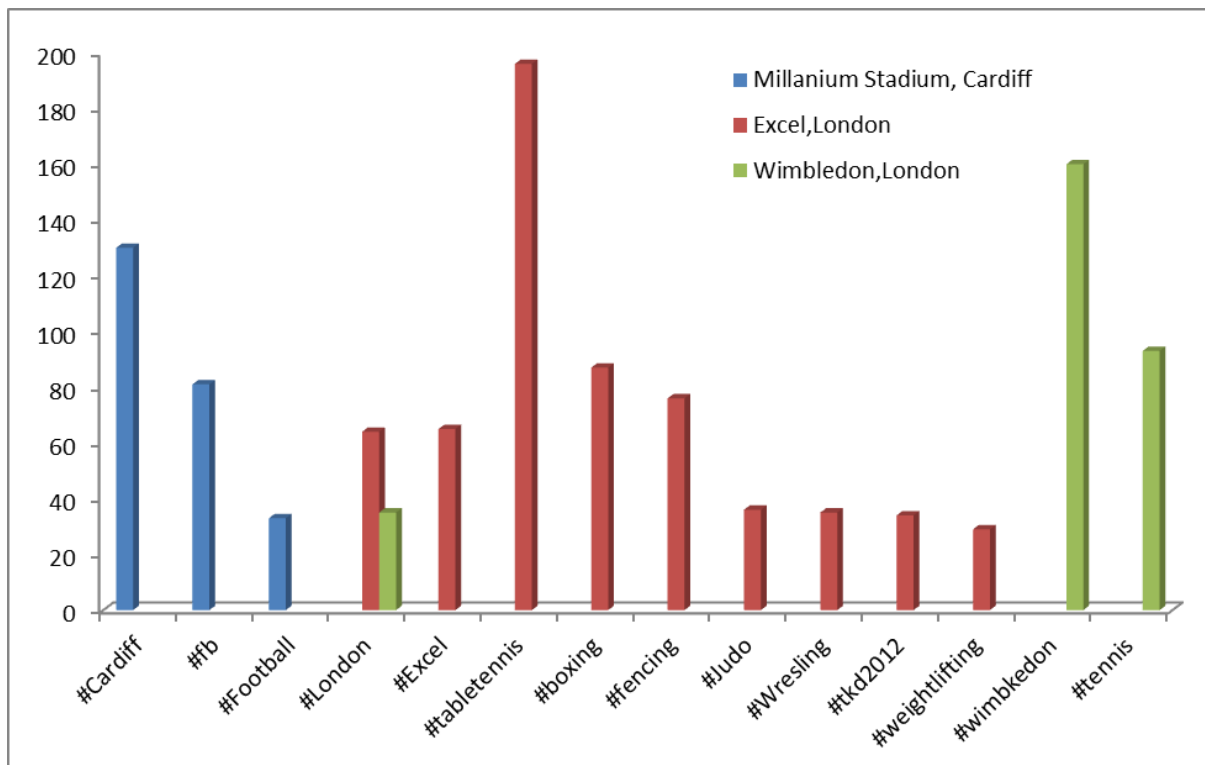


**Figure 1.** Frequent hashtags in three different places, with evidence of reference to place name and place-specific activity.

To demonstrate the effect of the temporal aspects in the tweets, the most frequent hashtags were analysed during the week 30/07/12 to 05/08/12. Tags that referred to the event in general persisted all week such as,'London2012', 'Olympics' and 'Great Team'. Other specific tags to the venues evolved over time. For example, the first medal for Great Britain was in rowing on the first of August: hashtags 'rowing', '1stgold' and 'gold' were detected then. Another example is on 05-08-12, where tags included '100final' and `Bolt', when the100m marathon was won by the Jamaica player 'Bolt'.(Figure 2.)
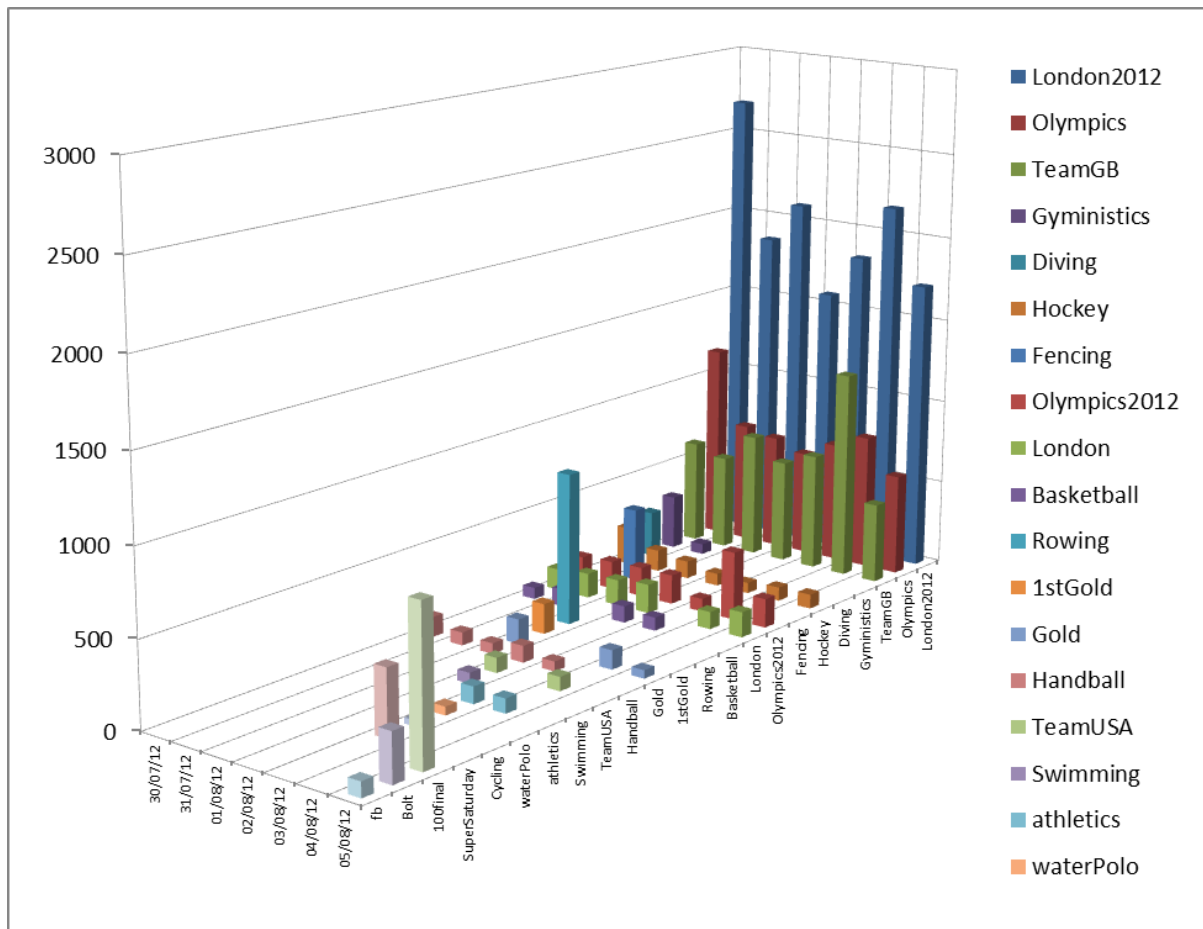
**Figure 2.** Frequent hashtags during a week in different sports venues

## 5. Conclusion and future work

The aim of this research work is to study the spatiotemporal aspects of Social Web data and analyse their value in understanding user behaviour and characteristics of places visited by users. In an initial experiment with Twitter data, it was shown how the tags in the data are rich in information about place, time and activities being carried out by users as well as how these change over time. Future work will aim to do further analysis of the data contents through keyword and entity extraction, not only hashtags. Moreover, we aim to develop methods to model place and user and their relationship.

## 6. Acknowledgements

## References

Abel F. Gao, Q. Houben, G. J. & Tao K. (2011). Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web.

Calabrese F., Pereira F., Di Lorenzo G., Liu L., & Ratti C. (2010). The geography of taste: analyzing cell-phone mobility and social events. *Pervasive Computing*, 22-37.

Chorley M., Colombo, G. B. Williams, M. J. Allen, S. M. & Whitaker R. M. (2011). Checking out checking in: observations on foursquare usage patterns. In *International Workshop on Finding Patterns of User Behaviors in NEtwork and MObility Data*.

Colombo G. B., Chorley, M. J. Williams, M. J. Allen, S. M. & Whitaker, R. M. (2012). You are where you eat: Foursquare checkins as indicators of human mobility and behaviour. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*(pp. 217-222). IEEE.

Cranshaw J., Schwartz R. Hong J. I., & Sadeh N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city.*Association for the Advancement of Artificial Intelligence*.

Dearman, D. & Truong K. N. (2010). Identifying the activities supported by locations with community-authored content. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 23-32). ACM.

Gruzd A., Doiron S. & Mai P. (2011). Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (pp. 1-9). IEEE.

Kim T., Huerta-Canepa, G. Park, J. Hyun, S. J. & Lee D. (2011). What's Happening: Finding Spontaneous User Clusters Nearby Using Twitter. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)* (pp. 806-809). IEEE.

Lane N. D., Lymberopoulos D., Zhao F. & Campbell A. T. (2010). Hapori: context-based local search for mobile phones using community behavioral modeling and similarity. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 109-118). ACM.

Mahmud J., Nichols J. & Drews C. (2012). Where Is This Tweet From? Inferring Home Locations of Twitter Users. *Proc AAAI ICWSM*, *12*.

Noulas A., Scellato S., Mascolo C., & Pontil M. (2011). An empirical study of geographic user activity patterns in foursquare. *ICWSM'11*.

Tatsuya F., L. Ryong et al. (2010) Discovery of user behavior patterns from geo-tagged micro-blogs. Proceedings of the 4th International Conference on Uniquitous Information Management and Communication. Suwon, Republic of Korea, ACM.

Thelwall M., Buckley K. & Paltoglou G. (2011). Sentiment in Twitter events.*Journal of the American Society for Information Science and Technology*, *62*(2), 406-418.

## Biography

*Soha Mohamed: First year PhD student in the school of computer science and informatics, Cardiff University.*
*Alia Abdelmoty: Lecturer in the school of computer science and informatics, Cardiff University.*