INFERRING ACTIVITIES FROM SOCIAL MEDIA DATA

Emmanouil Chaniotakis
Centre for Research and Technology Hellas – Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., 57001, Thermi, Thessaloniki, Greece
Tel: +30 2310 498 495 Fax: +302310 498 269; Email: chaniotakis@certh.gr

Constantinos Antoniou
Professor, Chair of Transportation Systems Engineering
Technical University of Munich
Arcisstraße 21, D-80333, Munich, Germany
e-mail: c.antoniou@tum.de

Georgia Aifadopoulou
Centre for Research and Technology Hellas – Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., 57001, Thermi, Thessaloniki, Greece
Tel: +30 2310 498 451 Fax: +302310 498 269; Email: gea@certh.gr

Loukas Dimitriou
Department of Civil and Environmental Engineering, University of Cyprus,
75 Kallipoleos Str., P.O. Box 20537, 1678 Nicosia, Cyprus
Tel: +35722892286*; E-mail: lucdimit@ucy.ac.cy*

Word count: 4,486 words text + 9 tables/figures x 250 words (each) = 6736 words

Submission Date: 01-08-2016

ABSTRACT

Social Media have been found to produce an unprecedented amount of information that can be extracted and used in transportation research with one of the most promising areas to be the inference of individuals' activities. While most studies in the literature focus on the direct use of Social Media data, this study presents an efficient framework for the inference of users' activities from Social Media data following a user–centric approach. The framework is applied to data from Twitter, combined with inferred data from Foursquare that contain information about the type of location visited. Then, the user data is classified using a density-based spatial classification algorithm that allows for the definition of commonly visited locations and the individual-based data is augmented with the known activity definition from Foursquare. Based on the known activities and the Twitter text, a set of classification algorithms is applied for the inference of activities of all tweets. The results are discussed upon the types of activities recognized and its classification performance. The classification results allow for a wide application of the framework in the exploration of the individuals' activity space.

Keywords: Social Media, Activity Space, Data Enrichment, Data Generalization, Classification, Data Collection

INTRODUCTION

From the rise of Social Media, they have received attention from the scientific community, forming a potentially new stream of research in many fields. The reasons behind this attention can be summarized upon the unprecedented amount of information that can be extracted and the opportunities that Social Media platform use can provide on direct communication with users. The statistics of Social Media use are astonishing: as of April 2016, on Twitter around 250,000 tweets are posted each minute; while almost 300,000 statuses are updated and about 136,000 photos are uploaded on Facebook. Studies indicate that around 76% of the online adults use Social Media, with 83% of them to be using their smart-phones in order to access and interact in Social Media platforms (pewinternet.com). Another interesting fact is that according to online ranking websites (alexa.com) Facebook and Twitter are within the 10 most visited websites (3rd and 10th position respectively in April 2016) while most of the rest most visited websites (4 out of 8) are search engines. Within the Social Media research context, several definitions have arose with the most inclusive to be found to be the one from Kaplan and Haenlein *(1)* who define Social Media as "a group of Internet-based applications that are built on the ideological and technological foundations of web 2.0, and allow the creation and exchange of user generated content". Probably for the first time and with the use of Social Media data–sets and platforms, scientists are able to sense peoples' thoughts, feelings and actions and interact with them in platforms characterized by ease of use and capabilities of equal and public exchange of opinions.

Focusing on the use of data, the information posted by users in Social Media that can be exploited for research and application, is diverse *(2)* and depends on the Social Media platform functionalities and data availability. However, as studies in transportation illustrate *(3, 4)*, the existence of spatial and temporal characteristics can by default provide information concerning the transportation system. Not limited to that, textual and user related information (available either from posts or from the user profile) enrich the information provided and increase the potential to be utilized towards newer solutions to transportation related problems *(5, 6)*. Finally, posts in Social Media can include integrated links (URL) from other Social Media platforms such as Foursquare, Instagram and Facebook, further enriching the information to be acquired.

One of the major transportation modelling necessity for which the incorporation of Social Media is appealing is the development of activity–based models. Although at the frontier of transportation demand modelling, activity–based models present many challenges especially concerning the data required for representing individuals activities *(7)*. In this direction data from Social Media can be utilized to provide the information on the activities users perform.

Pertinent literature suggests a wide spectrum of Social Media analysis in transportation. However, to the authors' knowledge, only a few focus on users' activities *(8–10)*, with most of the studies to focus on the extraction of information following what we will refer to as a post–centric approach: In post–centric analysis, each Social Media post is viewed as an entity, ignoring in most cases the user perspective. This is believed to disorient from the necessity to merge the online and offline space that could potentially offer useful information towards a better representation of the transportation system and specifically individuals' activities. On the other hand, user–centric approaches can provide valuable interpretations on elemental research questions, such as why a user posts on a Social Media platform, which activities are users performing and how online activity is linked with users' offline life.

This study provides an efficient user–centric framework for the inference of users' activities from Social Media data. The framework is applied in data from Twitter, inferred with data from Foursquare that contains information about the type of location visited. Then, the users' data is classified using a density based spatial classification algorithm that allows for definition of

commonly visited location and the individual based data is inferred with the known activity definition from Foursquare. Based on the known activities and the Twitter text, a set of classification algorithms is applied for the exploitation of activities from Twitter text that discussed upon the types of activities recognized and its performance.

The remainder of the text is organized as follows: first the related work on Social Media for transportation is presented, focusing on the potential uses of Social Media in transportation research. Next, in the Methods section the data collection and data enrichment methodologies proposed are presented and discussed, while in the Analysis section the results of its application are presented. Finally, in the Conclusions section, a discussion takes place on the applicability of the proposed method including some insights on the applicable future work.

RELATED WORK

Information from Social Media for transportation is mainly examined on the spatial component (geotag) that posts sometimes include *(11, 12)* and in some cases on the ways that the Social Media platforms are used for direct communication with customers *(13)*. The use of the Social Media data can be categorized in two categories: the use of the streaming data that Social Media commonly provide or the use of the historical Social Media data.

Streaming data in the majority of studies originates from Twitter and has been used for identification of events mainly focusing on social events or system disruptions *(12, 14, 15)*. Data can also originate from other Social Media platforms oriented towards the provision of information concerning social events *(16)*. In these works, the proposed methodologies use the textual and spatial information and in most cases, aggregation is performed on the spatial data and text identification or topic modelling is performed on the textual information, in order to augment incidents that can be detected.

Historical data from Social Media is mainly exploited upon the enhancement of transport models and the extraction of information concerning the transportation system. The spectrum of applications is wide. Some studies explore the spatial and temporal mobility patterns *(17, 18)*. The scale of those analysis can vary from a global level *(17)* to the very local level of a specific location in a city *(18)*. In most cases models are defined to explain the dynamics of online posts in regards to the offline locations. Additionally, the investigation of the applicability of data from Social Media for travel demand modelling is explored *(4, 8)*. This research stream steers in some studies towards the definition of the urban settings and related characteristics, such as Points of Interest (POIs), Urban Boundaries, Land uses *(19, 20)*, while some studies explore the potential of either directly deriving or inferring Origin Destination matrices from Social Media *(21, 22)* or the identification of user activities *(8–10)*.

This work builds on the finding of Lee et al. *(8)* and Hasan and Ukkusuri *(10)* on activity investigation, from Social Media. The former investigated the activity space of Twitter users and implemented Latent Class Analysis for the definition of major activities location aiming at identifying among others the optimal data collection period. The latter proposed a topic–modelling framework for the definition of multi-day activity patterns of individuals to identify user–specific activity patterns and the users, who mostly contribute to the definition of such a pattern. Although both works focus on users' activities, they only use a very small fraction of information available in Social Media due to constrains towards either the use of only Twitter–originated data or the use of only Foursquare–originated data. Also, activity inference in the Lee et al. *(8)* does not focus on the extraction of the type of activity, while Hasan and Ukkusuri *(10)* follow an activity–centric approach which does not identify locations of high density, something that can be very useful in terms of inferring a complete per user activity space.

METHODS

Data Collection and Data Enrichment

The Social Media data examined in this work originates from Twitter and was collected using the Twitter Application Programming Interface (API) service, for the collection of geotagged tweets. A two–phase data collection methodology is applied, where first data is conventionally collected by performing Twitter API queries for the last 200 geotagged tweets. This first data collection phase allows the creation of a user database; essentially forming a users' sample. Then for each user, n number of tweets are collected by performing queries to acquire their timeline *(11)*. It should be noted that the application of the second phase of the data collection methodology from Twitter is not bounded in a specific area as it collects all the tweets that individuals post.

Based on the Twitter dataset, extraction of information from Foursquare posts integrated in Twitter is performed to achieve the enrichment of activities–related information. In order to parse the information, the methodology presented in Figure 1 is applied. This includes the extraction of URL patterns using regular expressions, the extension of shortened links to long links and Fousquare queries. The queries' response is of JavaScript Object Notation (JSON) format and it can be either a venue, or a check–in. In some cases it requires for the link to be (within the Foursquare query) resolved. Finally, the response is parsed in tabular form and stored with the initial tweet information. It should be noted, that the required queries are subject to the limitations that the Twitter and Foursquare APIs impose. At the time this methodology was applied (June 2016), the maximum number of queries was 500 queries per hour. The application of the methodology for enriching twitter data was performed in R *(23)* using related packages (ThinkToStartR, stringr, rjson, RCurl, httr). The only exception was the URL extension component that was applied in Java.

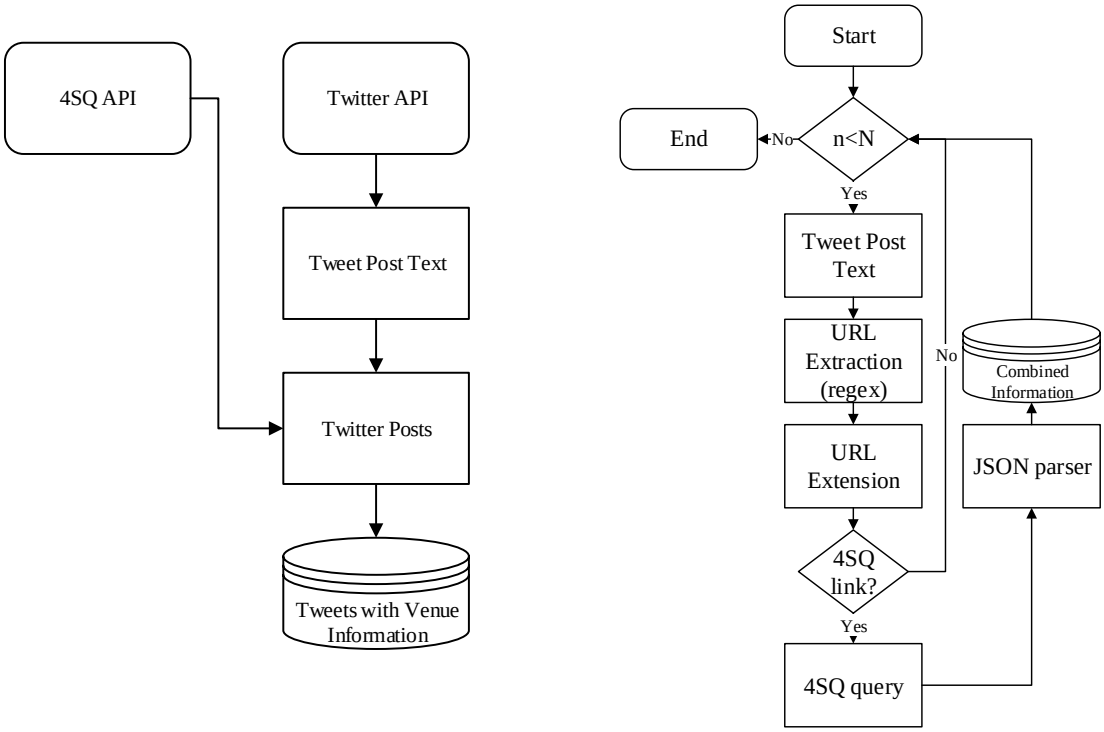

FIGURE 1 Twitter data enrichment from Foursquare: high–level methodology (left); detailed parsing methodology (right).

User–Centric Activity Enrichment

The resulting dataset from the data collection and dataset definition methodology is used for the extraction of information on the nature of activities users perform and post about in Social Media. The methodology is based on the notion that there are places that users visit frequently to perform activities (e.g. home, work) namely recurring activities, and other places which are scarcely visited or only visited once for non–recurring activities (e.g. leisure). This distinction of places visited and –as a consequence– performed activities is considered of high importance, for the data generalizations that are required aiming at both data reduction – in terms of the identification of locations that are commonly visited – and enrichment – in terms of genalization of activities identified.

The high level representation of the suggested methodology is presented in Figure 2. First, locations are characterized based on their spatial distance and their density, in order to specify locations that are visited by each user more than a specified number of times. The classification algorithm used is the Density–based Spatial Clustering of Applications with Noise (DBSCAN) *(24)*, which characterizes locations visited less than a specified number of times and under a specified density as noise. This is very convenient for the purposes of the study, as it allows for the distinction of locations that are frequently visited and can be all together characterized by only one Foursquare post, in a robust and fast way. The classified locations receive a characterization based on the venue categories they belong to. The characterized – based on recurrence – tweets combined with venue and temporal characteristic are then used for the definition of document–term matrices specified to each type of activities that can be performed in venues. This allows for direct classification of each tweet describing an activity, in an activity cluster using classification algorithms. It should be noted that not all tweets are used to describe activities.

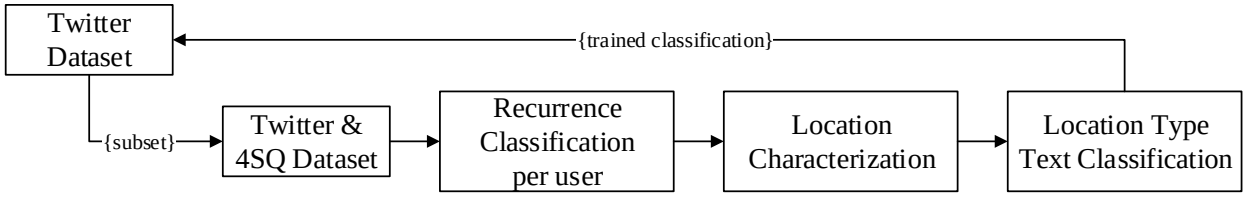

FIGURE 2 User–Centric Activity Enrichment methodology (high level)

Dataset Definition

The first phase twitter data collection method was used for continues data collection in a period of one year for the greater London area. The resulted dataset consists of 482,883 unique users and approx. 4.5 million tweets. For a random sample of those users (90,000 users) the last 200 tweets were collected using the second phase of the Twitter data collection methodology. In total, the database included 11,060,814 tweets. From those, 8,141,996 were tweeted in the greater London area. From the geotagged tweets in the greater London area, 3,764,230 (46.2%) included a link that could be parsed with venue information, 220,118 of which originated from Foursquare (2.7%).

The spatial and temporal characteristics of the sampled tweets are presented in Figure 3. Concerning the spatial distribution, it is clearly evidenced that the urban structures can be identified, as tweets seem to also be concentrated in the urban areas around the city of London. On the temporal level, the distributions presented indicate a peak among evening hours, while the highest number of unique tweets takes place during the weekends and specifically Saturday afternoon. This trend is found to be similar in other cities examined in the literature *(3, 11)*

indicating a tendency towards posting during non–working hours.

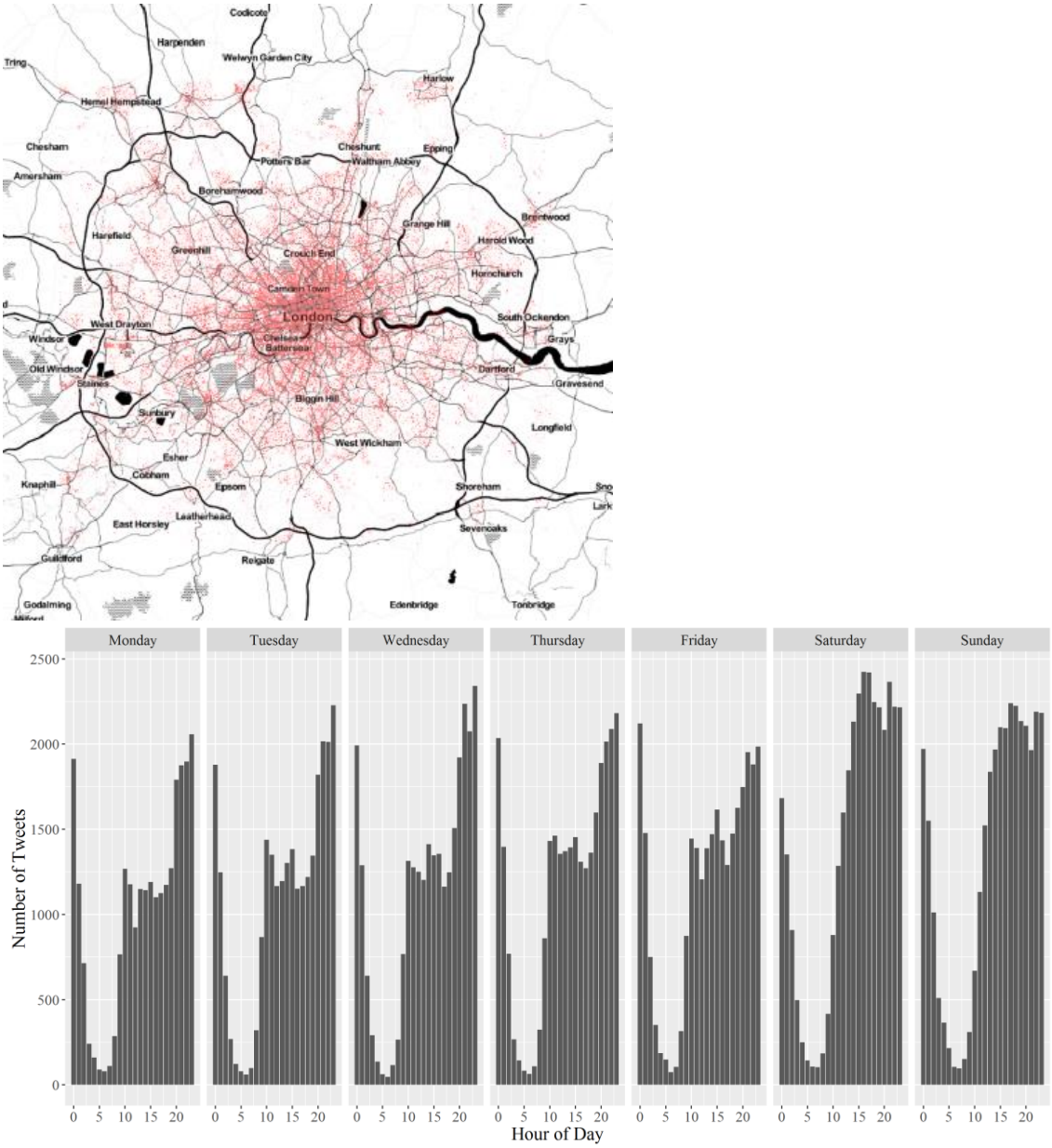

FIGURE 3. Spatial (top) and temporal (bottom) characteristics of collected Twitter data for London

As clarified in the literature, data collected from Twitter contains a large number of users who have either only a few tweets or post automated messages from the same location *(4)*. For that reason, a subset of users characterized based on their above average posting activity and a threshold of average twitter trip length (average direct distance between tweets of a user's determined as the average of the distance matrix) was selected, for the application of the User–Centric Activity Enrichment methodology. The subseting data flow and the subset procedure is presented in the Figure 4. The average number of tweets was found to be 22 and the average distance threshold was defined to be 1 km. The subset included in total 50,344 users who have posted in total 5,080,362 geotagged tweets.

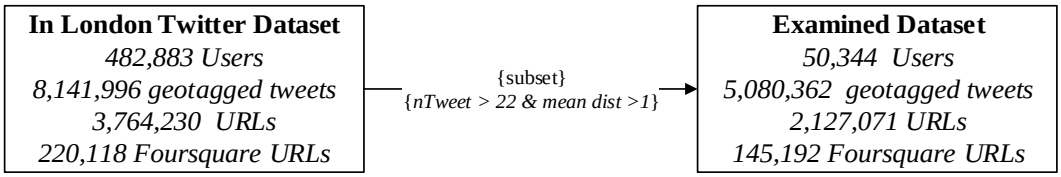

FIGURE 4 Subset characteristics, in the greater London area and concerning users with above average Social Media use

ANALYSIS

User Characteristics and Recurrence Classification

The methodological framework presented above, was applied for the London city dataset. The selected users with above average posting activity were examined upon their posting characteristics and the clustering characteristics. First, on the posting characteristics the average number of posts each user was posting was found to be 101. 8 with a standard deviation of 311.4, indicating that there is a number of users that are posting a very large number of tweets. The standard deviation of the percentage of tweets posted per day was found to be 0.8% suggesting a close to uniform distribution and indicating that on average the users selected are frequent Twitter users. From the initially selected 50,344 a fraction of 4,185 users (8.3%) were found to have posted at least one tweet that include a foursquare link. The average percentage of posts including activities for these users was found to be 39.0%.

On the classification for location recurrence, the analysis was performed in R using the *dbscan* library which implements the algorithm presented by Ester et al. *(24)* for the tweets of each user. The parameters were selected after examination of various settings taking into account the GPS accuracy *(25)* and the number of tweets that each individual posts. For this reason, the neighborhood of a point parameter (*Eps*) was defined to be 0.002 and the minimum number of points to be 5. On the following figure (Figure 5), an example of the classification results is presented for two cases: one for which the locations visited cannot be clustered and a second one for which the algorithm identifies two classes.

FIGURE 5 Example of classification using Density Based Spatial Clustering of Applications with Noise: no cluster identified (left) and two clusters identified (right)

The average number of clusters was found to be 2.41 with a standard deviation of 3.18. The percentile distributions of the locations that belong in a cluster and those characterized as noise per user are presented in Figure 6. It is indicated that a significant amount of users tend to only post in locations that were not be included in any cluster. This can be used in profiling of individuals who use Social Media based on the type of things that they tend to post.

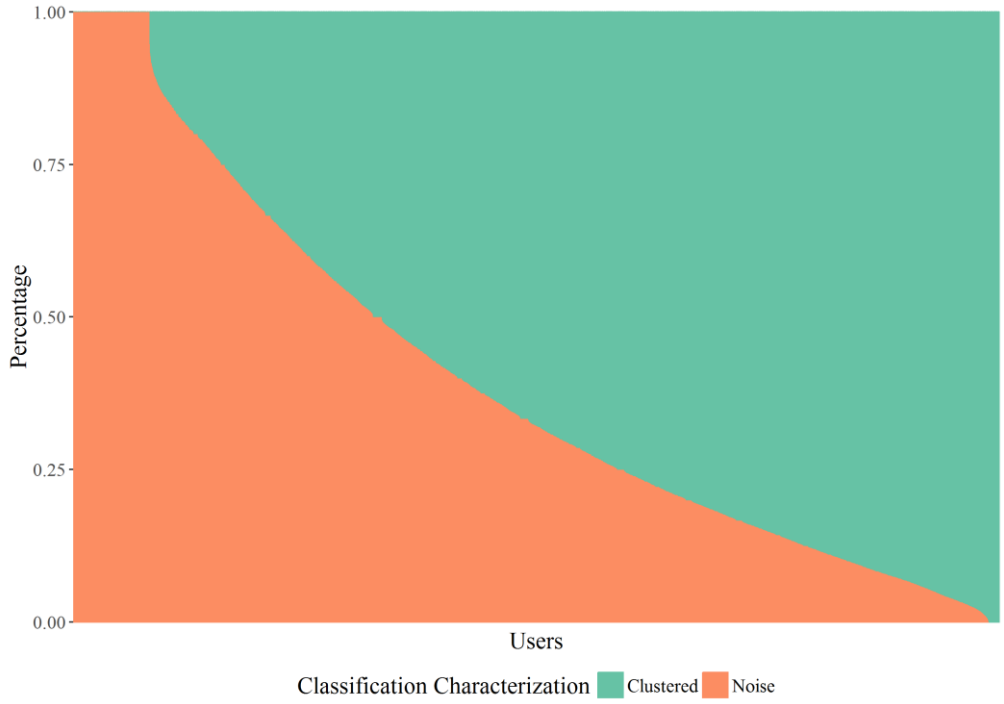

FIGURE 6 Example of classification using Density Based Spatial Clustering of Applications with Noise: no cluster identified (left) and two clusters identified (right)

Activity Data Enrichment

The categories provided by Foursquare on the locations visited by users were manually aggregated in 14 representative categories and examined upon their temporal characteristics for verification. As it is illustrated in Figure 7 there was a tendency towards leisure activities that was present for all days of a week. Activities such as education and work were clearly higher represented during week days and less represented during weekend days. The activities with the highest representation were found to be the "Bar – Pub" and "Restaurant" activities.

The application of the data enrichment methodology increases the information concerning the activities performed by individuals. First the identified as clustered destinations enriched were enriched based on location characterization (activity). More specifically, for the dataset of the above average users defined; from the initial number of Foursquare posts the number of posts that were assigned in recurrence clusters was found to be 172,675. This is based on the initial 65,806 tweets that included known location characterization to be associated with a known cluster. It should be noted that this data enrichment step further enriched the classification process presented below, as it allows for posts that were not meant to denote visitation at a specific location to be associated with a cluster location characterization and as a consequence an activity.

The last step of the data enrichment methodology is the training of classification model that can represent the activity space based on the Twitter text that individuals post. This approach

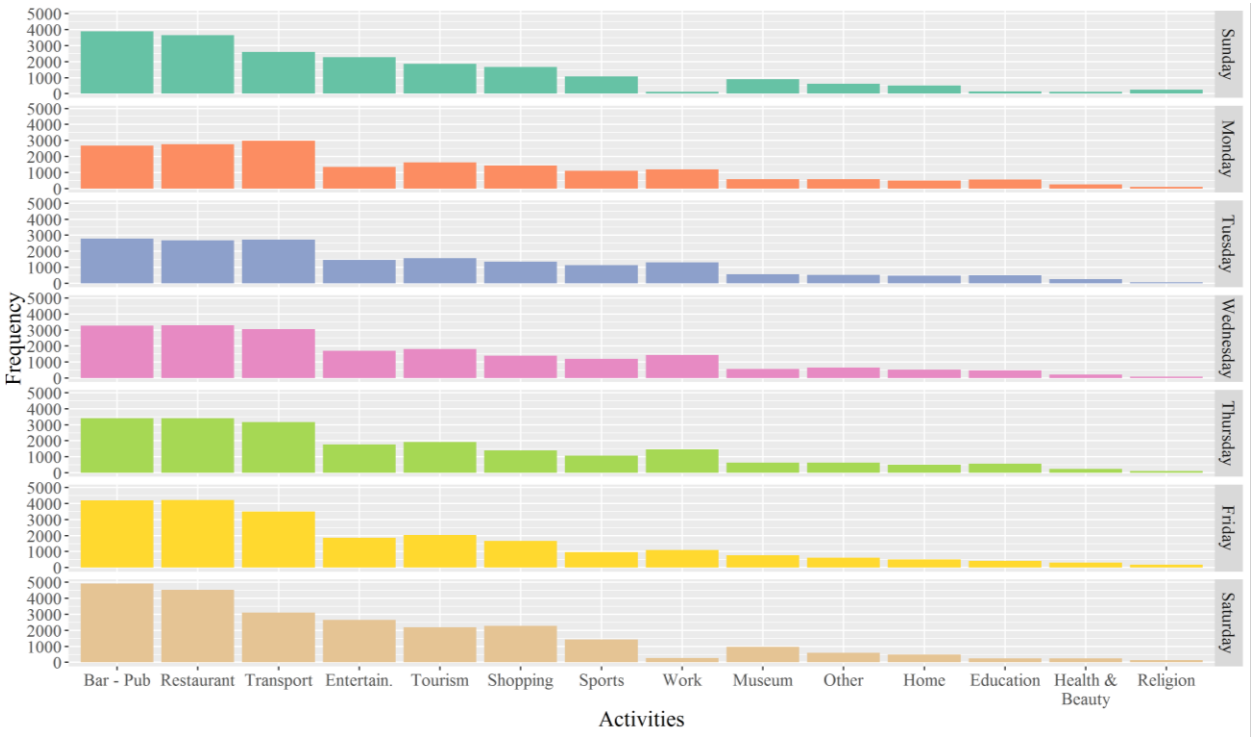

FIGURE 7 Temporal distribution of Foursquare labeled activities

would greatly improve the data availability on which researchers build their analysis. A set of classification methods believed to better fit the methodological framework proposed were selected and tested: a) Support Vector Machine (SVM), b) Generalized Linear Model via Penalized Maximum Likelihood (GLMPML) and c) Maximum Entropy (MAXENT). Given computational issues faced when training the classifiers for the whole dataset, the test of the best algorithm took place 10 times on a subset of data that included the major activities identified (Bar–Pub, Entertainment, Restaurant, Sports, Shopping, Other, Work, Tourism, Transport) and for 20,000 randomly selected cases. Each time, the dataset was split in training (85% – 17,000 entries) and testing set (15%, 3000 entries). The overall accuracy results from the classification are presented in the Table 1 and Figure 8. Table 1 includes a summary of the results from the examined classification methods on the metrics of precision, recall and the F–score. Results in Table 1 suggest that both the Maximum Entropy and the Generalized Linear Model via Penalized Maximum Likelihood are considered capable to be used for the identification of activities from Tweets.

TABLE 1 Classification Performance

	Average Precision	Precision Standard Deviation	Average Recall	Recall Standard Deviation	Average F–Score	F–Score Standard Deviation
SVM	0.6374	0.0104	0.5808	0.0132	0.6006	0.0117
MAXENT	0.8060	0.0112	0.7773	0.0081	0.7881	0.0080
GLM	0.8392	0.0078	0.6752	0.0068	0.7249	0.0064

Furthermore the precision of both algorithm indicate that the results would provide a sufficient accuracy, especially taking into account the short number of characters allowed in each tweet. Figure 8 on the other hand provides an overview of the classification in terms of confusion for the GLM and MAXEXT case. As it is clearly indicated for both classification methods, the pair that is mostly confused is the Bar–Pub and restaurant pair, which is logical when taking into account that both location categories might involve same activities performed by users (as commonly met for example drinking wine).

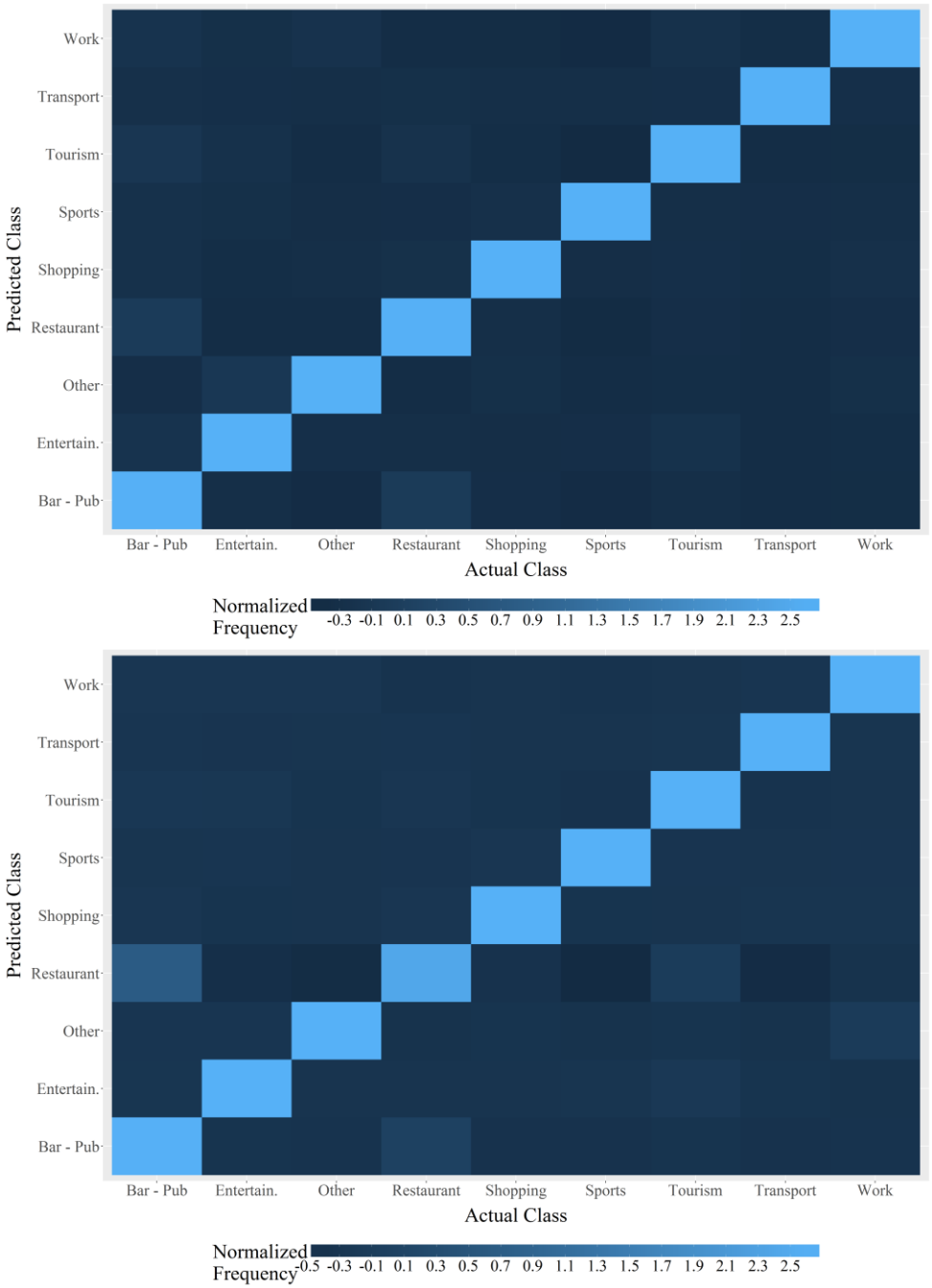

FIGURE 8: Confusion Matrix for the Performance of Classification for the two selected performing classification algorithms: Max Entropy (top) GLM (bottom)

CONCLUSIONS

In this research, we present a methodological framework for the characterization of transport–related activities for data originating from Twitter, using inferred data from Foursquare that defines a combined dataset. The resulted combined data is then used to derive classification methodologies for activity characterization using Twitter text. The methods to infer location characterizations (and, as a concequence, activities) and to enrich data, based on the user–centric activity data enrichment, have been applied for a large dataset from London. The results of the analysis indicate the capabilities of the proposed methods to derive activity patterns for individuals and to further use data from Social Media in transportation research and more specifically in activity–based models. Furthermore, the classification performance indicates that, for a sufficient number of cases, the data enrichment can indicate activities performed for all tweets recorded, based on the posted text, which –given the use of endogenous posting data– suggests that the generalization of the proposed methodology in other languages is possible.

Although the results of the analysis seem promising, there are some drawbacks that might be raised in a wide application of the methodology in other study areas and especially upon the final text classification for the inference of activities. To begin with, the methodologies are –of course– bounded to the wide–spread use of Foursquare in the area examined, which might not be the case in some countries or regions. Additionally, the classification has been performed only on posts that include a Foursquare URL or have been clustered in commonly visited venues by user locations. Within this data, only a small fraction was attributed to non–activities related posts (such as drinking wine and running denoted as activities, as opposed to –for example– posting or commenting on news). It is believed that for a proper representation of the activity space of individuals, based on text classification, non–activity related tweets should be detected (either using manual labeling or classification algoritms) to allow for a more complete representation of the reasons that users post on Social Media. Finally, although the examined algorithms seem to perform well, the performance of additional classification algorithms and the sensitivity of their parameters should be examined.

LITERATURE

1 A. M. Kaplan and M. Haenlein, ‘Users of the world, unite! The challenges and opportunities of Social Media’, *Business Horizons*, vol. 53, no. 1, pp. 59–68, Jan. 2010.

2 G. Cao, S. Wang, M. Hwang, A. Padmanabhan, Z. Zhang, and K. Soltani, ‘A scalable framework for spatiotemporal analysis of location-based social media data’, *Computers, Environment and Urban Systems*, vol. 51, pp. 70–82, 2015.

3 E. Chaniotakis, C. Antoniou, and E. Mitsakis, ‘Data for Leisure Travel Demand from Social Networking Services’, in *4th hEART Symposium*, 2015.

4 J. H. Lee, S. Gao, and K. G. Goulias, ‘Comparing the Origin-Destination Matrices from Travel Demand Model and Social Media Data’, in *Transportation Research Board 95th Annual Meeting*, 2016, no. 16–0069.

5 A. Gal-Tzur, S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, and I. Shoor, ‘The potential of social media in delivering transport policy goals’, *Transport Policy*, vol. 32, pp. 115–123, 2014.

6 S. M. Grant-Muller, A. Gal-Tzur, E. Minkov, S. Nocera, T. Kuflik, and I. Shoor, ‘Enhancing transport data collection through social media sources: methods, challenges and opportunities for textual data’, *IET Intelligent Transportation Systems*, 2014.

7 E. Cascetta, *Transportation systems analysis: models and applications*. 2009.

8 J. H. Lee, A. W. Davis, and K. G. Goulias, 'Activity Space Estimation with Longitudinal Observations of Social Media Data', in *Transportation Research Board 95th Annual Meeting*, 2016.

9 A. Noulas, C. Mascolo, and E. Frias-Martinez, 'Exploiting foursquare and cellular data to infer user activity in urban environments', in *2013 IEEE 14th International Conference on Mobile Data Management*, 2013, vol. 1, pp. 167–176.

10 S. Hasan and S. V Ukkusuri, 'Urban activity pattern classification using topic models from online geo-location data', *Transportation Research Part C Emerging Technologies*, vol. 44, pp. 363–381, 2014.

11 E. Chaniotakis and C. Antoniou, 'Use of Geotagged Social Media in Urban Settings: Empirical Evidence on Its Potential from Twitter', in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2015, vol. 2015–Octob, pp. 214–219.

12 Y. Gu, Z. (Sean) Qian, and F. Chen, 'From Twitter to detector: Real-time traffic incident detection using social media data', *Transportation Research Part C Emerging Technologies*, vol. 67, pp. 321–342, 2016.

13 S. Bregman, *Uses of social media in public transportation*, vol. 99. Transportation Research Board, 2012.

14 S. Jiang, A. Alves, F. Rodrigues, J. Ferreira Jr., and F. C. Pereira, 'Mining point-of-interest data from social networks for urban land use classification and disaggregation', *Computers, Environment and Urban Systems*, 2015.

15 A. Kumar, M. Jiang, and Y. Fang, 'Where not to go?: detecting road hazards using twitter', *Proceedings of the 37th international ACM conferance*, vol. 2609550, pp. 1223–1226, 2014.

16 F. C. Pereira, F. Rodrigues, E. Polisciuc, and M. Ben-Akiva, 'Why so many people? Explaining Nonhabitual Transport Overcrowding With Internet Data', *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1370–1379, 2015.

17 B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, 'Geo-located Twitter as proxy for global mobility patterns', *Cartography and Geographic Information Science*, vol. 41, no. 3, pp. 1–12, 2014.

18 Y.-S. Cho, G. Ver Steeg, and A. Galstyan, 'Where and Why Users" Check In".', 2014.

19 J. Li, Q. Qin, J. Han, L.-A. Tang, and K. H. Lei, 'Mining Trajectory Data and Geotagged Data in Social Media for Road Map Inference', *Transactions in GIS*, vol. 19, no. 1, pp. 1–18, Feb. 2015.

20 B. Jiang and Y. Miao, 'The evolution of natural cities from the perspective of location-based social media', *The Professional Geographer*, vol. 67, no. 2, pp. 295–306, 2015.

21 F. Yang, P. J. Jin, X. Wan, R. Li, and B. Ran, 'Dynamic origin-destination travel demand estimation using location based social networking data', in *Transportation Research Board 93rd Annual Meeting*, 2014, no. 14–5509.

22 P. Jin, M. Cebelak, F. Yang, J. Zhang, C. Walton, and B. Ran, 'Location-Based Social Networking Data: Exploration into Use of Doubly Constrained Gravity Model for Origin-Destination Estimation', *Transportation Research Record: Journal of the Transportation Research Board*, no. 2430, pp. 72–82, 2014.

23 R. C. Team and others, 'R: A language and environment for statistical computing', 2013.

24 M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, 'A density-based algorithm for discovering clusters in large spatial databases with noise.', in *Kdd*, 1996, vol. 96, no. 34, pp. 226–231.

25 M. Schaefer and T. Woodyer, 'Assessing absolute and relative accuracy of recreation-grade and mobile phone GNSS devices: a method for informing device choice', *Area*, vol. 47, no. 2, pp. 185–196, 2015.