Volatile Classification of Point of Interests based on Social Activity Streams

A. E. Cano, A. Varga and F. Ciravegna [1]

OAK Group, Dept. of Computer Science, The University of Sheffield, UK
{A.Cano, A.Varga, F.Ciravegna}@dcs.shef.ac.uk

Abstract. Location sharing services(LSS) like Foursquare, Gowalla and Facebook Places gather information from millions of users who leave trails in locations (i.e. chekins) in the form of micro-posts. These footprints provide a unique opportunity to explore the way in which users engage and perceive a point of interest (POI). A POI is as a human construct which describes information about locations (e.g restaurants, cities). In this work we investigate whether the collective perception of a POI can be used as a real-time dataset from which POI's transient features can be extracted. We introduce a graph-based model for profiling geographical areas based on social awareness streams. Based on this model we define a set of measures that can characterise a location-based social awareness stream as well as act as indicators of volatile events occurring at a POI. We applied the model and measures on a dataset consisting of a collection of tweets generated at the city of Sheffield and registered over three week-ends. The model and measures introduced in this paper are relevant for design of future location-based services, real-time emergency-response models, as well as traffic forecasting. Our empirical findings demonstrate that social awareness streams not only can act as an event-sensor but also can enrich the profile of a location-entity.

Keywords: Points of Interest, social awareness streams, social data mining, citizen sensing, emerging semantics

1 Introduction and Motivation

Recent studies in user profiling have proposed the use of social activity streams for modelling users' interest, activities and behaviour [11][1][3]. These studies explore a user's comments in windows of time for revealing hidden features; which can aid in profiling the user in real-time. Although people-entities have started to be modelled in real-time, little has been done in modelling other entities involved in the environment in which a user is immersed. One example of these entities is Location.

In terms of location-awareness, a Point of Interest (POI) has been so far modelled as a set of static data (e.g. name, address, geo-coordinates) and classified according to the type of services it provides. Nonetheless, there are diverse latent (or hidden) features which can describe volatile and temporal aspects of it. For example, in normal conditions London, UK can be classified as a city labelled as: Urban, Tourism, Fashion. However during the London riots(Aug 2011), the collective opinions gathered through social activity streams (i.e. Twitter) regarding this city, started profiling this place with the following tags: looting,unrest,police. These tags clearly provide a temporal reclassification of this venue labelling it as for example: Political, Uprising, Violence.

In this paper, we investigate whether the supplement of situational knowledge extracted from social activity streams can be used to infer higher level contextual information, which can induce a transient representation of a venue. Given the real-time and volatile nature of events happening at a venue, providing an accurate classification of these events involve different challenges including the variation of the vocabulary and classes in which an event could be classified in time.

The contributions of this paper are as follows:

- *GeoLattice Awareness Streams*: We introduce a graph-based model for profiling geographical areas based on social awareness streams.
- *Approach to derive a transient semantic classification of a POI*: We present a novel approach for dynamically classifying POI based on location-based social footprints and DBPedia structured data. We define a set of measures that can characterise a location-based social awareness stream as well as act as indicators of volatile events occurring at a POI.
- *Empirical Study*: We applied this methodology in a dataset consisting of a collection of tweets generated at the city of Sheffield and registered over three week-ends.

The model and measures introduced in this paper are relevant for design of future location-based services, real-time emergency-response models, as well as traffic forecasting.

2 Related Work

Little work has been done in classifying POIs based on location-based social activity streams. However, there are several research directions closely related to POI classification. Analysing the contextual meanings of places has long attracted attention by researchers in fields like social interaction, environmental psychology, ubiquitous computing and spatial data mining. Researchers on social interaction and environmental psychology have documented the way in which mobile users tend to provide information about location when they are asked about their current activity [7][12]. Schegloff [10] noted that during a conversation, attention is exhibited to: 1) 'where-we-know-we-are'; 2) 'who-we-know-we-are'; 3) 'what-we-are-doing-at-this-point-in-conversation'; from which a '*this* situation' can be translated in some '*this* conversation, at *this* place, with *these* members, at *this* point in its course'. This contextual knowledge has been used to infer a users' situational features including a person's level of availability or interruptibility.

The role of geography and location in online social networks has recently attracted increasing attention. Experimental work done on location awareness has shown that location sharing services (LSS) (e.g. Foursquare) are used to express not only users' whereabouts but also their moods, lifestyle and events [2]. In their work, Barkhuus et al. allowed users to tag areas and build a repartee in a group. They pointed out four different types of location labels that participants used in their study, including: 1) geographic references, 2) personal meaningful place, 3) activity-related labels, and 4) hybrid labels.

Cheng et al.[4] modelled the spatial distribution of words in Twitter's user-generated content for predicting user's location. Following a top-down approach they propose a probabilistic framework for estimating a Twitter user's city-level location based on the content of the user's tweets even on the abscence of any geospatial cues. Although their approach is content-based and can automatically indetify words in tweets with a strong geo-scope, they don't provide a topical categorisation of a given geo-scope.

Further work from Cheng et al [13] study mobility patterns of users in location sharing services (LSS), they correlate social status, geographic and economic factors with mobility and perform a sentiment-based analysis of post for deriving unboserved context between people and locations.

Lin et al [8] derive a taxonomy of different place naming methods, showing that a person's perceived familiarity with a place and the entropy of that place (i.e. the variety of people who visit it) strongly influence the way people refer to it when interacting with others. Based on this taxonomy, they present a machine learning model for predicting the place naming method people choose. Ireson and Ciravegna [6] study toponym resolution (i.e. the allocation of specific geolocation to target location terms) using Flickr data. They construct an SVM classifier for predicting location labels associated to a location term. Their model makes use of information context features including geo-tag media, users' contacts'related tags.

Regarding place descriptions based on location sharing services (LSS), Hightower [5] redefines a place as an evolving set of both communal and personal labels for potentially overlapping geometric volumes. He highlights that a meaningful place can capture the venue's demographic, environmental, historic, personal or commercial significance.

Our work is in line with Hightower's definition of a place, however rather than study location-sharing practices we aim to study how location-based generated content can be modelled for discoverying topics or categories that classify a place on time.

3 GeoLattice Awareness Stream

Following the Tweetonomy model suggested by Wagner and Strohmaier[11], we introduce a formalisation for describing the comments related to a geographical region in time; we refer to it as GeoLattice Awareness Streams.

The W3C POI Working Group [1] defines a POI as a human construct which describes information about locations. According to their definition, a POI is not limited to a set of coordinates and an identifier but also can include a more complex structure like for example a three dimensional model o a building, opening and closing hours etc.

As mentioned in the previous section, location sharing services provide a classification of their points of interest according to the type of service they provide (e.g. Food, Nightlife Spots), however these categories are static and do not reveal any information about the type of events occurring in a given venue. The key idea of our approach is to enrich a POI by associating transient categories emerging from social activity streams regarding this POI.

Definition 1. *A GeoLattice Awareness Stream can be defined as a sequence of tuples* $\mathtt{S} := (Poi_{q1}, C_{q2}, R_{q3}, Y, ft)$ *where*

[1] W3C POI Working Group, http://www.w3.org/2010/POI/

- $\texttt{Poi}, \texttt{M}, \texttt{R}$ *are finite sets whose elements are called Points of Interest, Messages and Resources;*
- *Each of these sets is qualified by* $\texttt{q1}$, $\texttt{q2}$ *and* $\texttt{q3}$ *respectively (explained below);*
  - *The qualifier* $\texttt{q1}$ *for a Point of Interest (poi) includes for example name, geographical-bounding area, and geo-coordinates.*
  - *The qualifier* $\texttt{q2}$ *for a message* m *considers for example the message's source (e.g Facebook, Twitter) and it's geo-coordinates.*
  - *The qualifier* $\texttt{q3}$ *for a resource* r *considers:* $\texttt{R}_{cat}$ *(category),*$\texttt{R}_k$ *(keywords),* $\texttt{R}_h$ *(hashtags).*
- $\texttt{Y}$ *is the ternary relation* $\texttt{Y} \subseteq \texttt{Poi} \times \texttt{M} \times \texttt{R}$ *representing a hypergraph with ternary edges. The hypergraph of a GeoLatice Awareness Stream* $\texttt{Y}$ *is defined as a tripartite graph* $\texttt{H}(\texttt{Y}) = \langle \texttt{V}, \texttt{E} \rangle$ *where the vertices are* $\texttt{V} = \texttt{Poi} \cup \texttt{M} \cup \texttt{R}$*, and the edges are:* $\texttt{E} = \{\{poi, m, r\} \mid (poi, m, r) \in Y\}$.
- f_t *is a function that assigns a temporal marker to each Y;* $f_t : Y \rightarrow T$.

Given a GeoLattice awareness stream S, a POI awareness stream can be defined as the sequence of tuples from S where:

$\texttt{S}(\texttt{Poi}') = (\texttt{Poi}, \texttt{M}, \texttt{R}, \texttt{Y}', ft)$, and $\texttt{Y}' = \{(\texttt{poi}, \texttt{m}, \texttt{r}) \mid \texttt{poi} \in \texttt{Poi}' \vee \exists \texttt{poi}' \in \texttt{Poi}', \tilde{\texttt{m}} \in \texttt{M}, \texttt{r} \in \texttt{R} : (\texttt{poi}', \tilde{\texttt{m}}, \texttt{r}) \in \texttt{Y}\}$ i.e., a POI Awareness Stream is the aggregation of all messages which are related to a certain set of points of interest $\texttt{poi} \in \texttt{Poi}'$ and all resources and further points of interest related with these messages.

4 Transient Semantic Classification of a POI

4.1 Problem Statement

Comments extracted from social activity streams can be described as semi-public, natural-language messages produced by different users and characterised by their brevity. Given these characteristics and the variation in the vocabulary appearing on a POI awareness stream comments, finding relevant categories that can accurately qualify a comment is a challenging task.

Definition 2. *We define a temporal classification of a Point of Interest as the aggregation of* $\texttt{R}_{cat}$ *category resources qualifying messages contained in a specific window of time denoted by* $[t_s, t_e]$*. An* $\texttt{S}(\texttt{Poi}')[t_s, t_e]$ *is defined as* $\texttt{S}(\texttt{Poi}')$ *where* $ft : Y \rightarrow T, t_s \leq ft \leq t_e$.

Given the above definition, our task consists on obtaining category resources $\texttt{R}_{cat}$ which can classify a poi within a window of time $[t_s, t_e]$. In this section, we introduce a strategy for categorising points of interest.

The POI categorisation within a window of time could enable reactive services (e.g. targeting advertisements to users based on a users location and the POI categorisation, emergency-response).

4.2 Entity-Based Discovery of Transient Categories

Our intuition is to use the categorisation of the messages' resources generated from a Point of Interest awareness stream ($\texttt{S}(\texttt{Poi}')$) taken in windows of time ($[t_s, t_e]$), to induce a categorisation function. Figure 1 presents an overview of our approach.

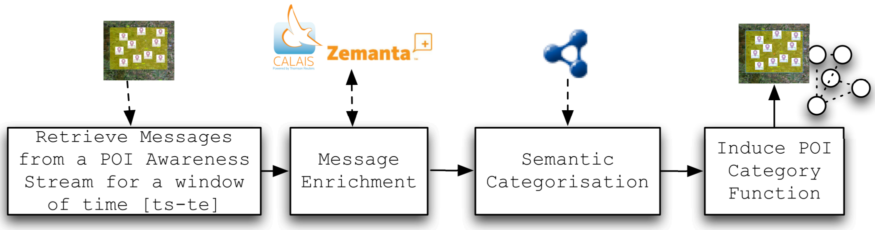

Fig. 1. Category Induction Pipeline: Messages are retrieved from a POI awareness stream. DBPedia categories are derived for each enriched message. These set of categories are used to induce a transient categorisation of a Point of Interest.

Message Enrichment Given a message from a POI awareness stream $\mathtt{S}(\mathtt{Poi}')$, we perform a lightweight *message enrichment* by using Zemanta [2], and OpenCalais [3]. These services perform entity-extraction on the input message identifying resources which can be qualified as: $\mathtt{R}_o$ (organisations – entities recognised as an organisation), $\mathtt{R}_p$ (people –entities recognised as a person), $\mathtt{R}_l$ (location – entities recognised as a location) and $\mathtt{R}_{li}$ (links resources). These services also provide DBPedia concepts relevant to the message. Consider the example in Figure 2, where the extracted entities and DBPedia concepts for a Twitter message are shown.

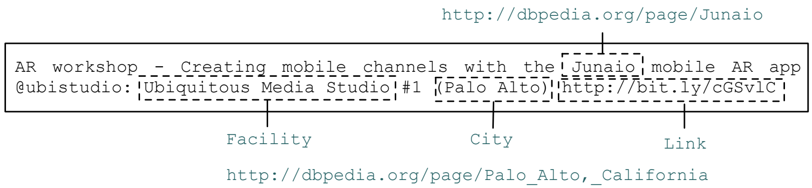

Fig. 2. Message Enriched with Zemanta and OpenCalais services. These service return entity labels as well as DBPedia concepts related to the message

Semantic Categorisation In order to semantically categorise a POI stream's message (m), we search for DBPedia concepts which are relevant to the extracted entity-based resources, and aggregate these concepts to those already suggested by the message enrichment services. Given a resource (r) we extract DBPedia *categories* and *broader categories* from the DBPedia Linked Data Graph (D) using the following construct:

[2] Zemanta, http://www.zemanta.com/
[3] OpenCalais, http://www.opencalais.com/

$$\begin{aligned} \mathtt{R}_{cat}(r) = \{x_{cat} \cup x_{broaderCat} | \\ < r, \text{dcterms:subject}, x_{cat} > \\ \wedge < x_{cat}, \text{skos:broader}, x_{broaderCat} > \in D\} \end{aligned} \quad (1)$$

For each resource (r) we SPARQL query DBPedia retrieving the collection of categories (dcterms:subject) and parent categories (skos:broader) of r. Using the previous construct, we derive the categories presented in Table 4.2 for the resource Palo_Alto contained in the example of Figure 2. These categories become a resource category R_{cat} of the POI awereness stream ($S(Poi')$).

Entity	Category
Palo_Alto (of type City)	`dcterms:subject` *Palo_Alto,_California*
	`skos:broader` *Populated_places_in_Santa_Clara*
	`skos:broader` *University_towns_in_the_United_States*
Junaio (of type Thing)	`dcterms:subject` *Augmented_reality*
	`skos:broader` *Mixed_reality*

Table 1. Categories and broader categories derived for the entities extracted from the comment in Fig 2

Induce Category Function After applying the *semantic categorisation* technique to all messages belonging to a POI stream taken from a window of time $[t_s, t_e]$, we need to weight them in order to identify the relevant categories.

In order to do so, we utilise the resource category stream ($S(R'_{cat})$) of a POI stream ($S(Poi')$), which is the collection of all category resources classifying the POI stream's messages. For characterising the POI stream ($S(Poi')$) based on the category resources we propose two metrics:

1. *Category Entropy of a Stream*, which indicates the topical diversity of the stream. We defined the category entropy in terms of the POI stream's vocabulary as :

$$\mathtt{CE(c)} = -\sum_{w \in R_k} P(w|c) * log(P(w|c)) \quad (2)$$

where w is a word in the POI stream's vocabulary ($S(R'_k)$), and c is a category in the POI stream's categories ($S(R'_{cat})$). Low category entropy levels reveal that a stream is dominated by few categories, while a high category balance reveals a higher topical diversity. In normal conditions (i.e. no special events happening), we would expect for example to obtain a low category entropy levels for a POI stream referring to a Restaurant, since the messages would be classified within a limited set of categories related to Food. While for a POI stream referring to a city

in normal conditions (no particular events happening), we would expect to observe higher category entropy levels since the topical diversity would be higher.
However if normal conditions are broken, and unexpected (or volatile) events start to happen, we would expect to observe an increment in the category entropy levels of Restaurant POI stream, and a decrement in the category entropy levels of a City POI stream. The category entropy acts in this way as an indicator of volatile events.

2. *Mutual Information (MI)*, measures the information that two discrete random variables share. In this work we consider the following:
   - *Categories-Hashtags (MI)*

$$\mathtt{I(C;H)} = \sum_{c \in R_{cat}} \sum_{h \in R_h} p(c,h) * log \frac{p(c,h)}{p(c)p(h)} \tag{3}$$

   where c is a category in the POI stream's categories ($S(R'_{cat})$) and h is a hashtag in the POI stream's hashtags ($S(R'_h)$) and p(c,h) is the joint probability distribution function of C and H, with marginals p(c) and p(h).
   - *Categories-Keywords (MI)*

$$\mathtt{I(C;K)} = \sum_{c \in R_{cat}} \sum_{w \in R_k} p(c,w) * log \frac{p(c,w)}{p(c)p(w)} \tag{4}$$

   where c is a category in the POI stream's categories ($S(R'_{cat})$) and w is a word in the POI stream's keywords ($S(R'_k)$).
   - *Hashtags-Keywords (MI)*

$$\mathtt{I(H;K)} = \sum_{h \in R_h} \sum_{w \in R_w} p(h,w) * log \frac{p(h,w)}{p(h)p(w)} \tag{5}$$

   where h is a hashtag in the POI stream's hashtags ($S(R'_h)$).

   The higher the mutual information, the more one random variable is relevant to the other.

5 Experiments

In this section we discuss our approach for evaluating the accuracy of the strategies proposed in Section 4 by using the formalisation introduced in Section 3. In order to identify a transient categorisation of a point of interest we decided to investigate a POI stream $S(Poi')$ in windows of time of one week-end.

5.1 Dataset

The corpus used for our study consists of Twitter messages taken over three week-ends in the city of Sheffield. Since we aim to study patterns emerging from volatile events we registered a week-end in normal conditions (i.e. no events happening) from 2011-06-10 to 2011-06-13 as control and two more week-ends in which especial events occurred.

The especial events were the Sheffield Food Festival (from 2011-07-08 to 2011-07-11) and the Sheffield Tramlines Music Festival (from 2011-07-22 to 2011-07-25). The data was collected using the Twitter Streaming API[4] with the public firehose and filtering by geographical area (using Sheffield's bounding geo-coordinates).

For each week-end dataset we removed stop words and applied the approach presented in Section 4.2, extracting hasthags, keywords and entity resources as well as DBPedia categories for these resources. The statistics for each stream is summarised in Table 2.

Week-End	Tweets	Users	Hashtags	Links	GeoTagged	RT	Reply
Common	5853	649	9%	5%	27.11%	2.8%	40.6%
Food Festival	11203	726	18%	4.2%	40.7%	4.2%	40.7%
Tramlines	13381	899	9%	24%	14.8%	9%	39.3%

Table 2. General Statistics, percentages of messages containing hashtags, links, geotagged, RT (retweeted) and Reply (tagged as a reply-tweet)

Week-End	Hashtags	Resources[a]	Categories [b]
Common	9%	1475	9495
Food Festival	18%	2681	830
Tramlines	9%	1912	9770

[a] DBPedia resources derived from the messages
[b] DBPedia categories derived from the resources

Table 3. Streams hashtags, and categories.

5.2 Results and Discussion

First we analyse the most frequent hashtags in the three datasets presented in Table 4. Although trends in hashtags are useful for detecting changes in a stream, hashtags tend to present high ambiguity, and a frequent use of abbreviations. These are some of the reasons why hashtags are not enough to provide a categorisation by themselves.

We calculated the categories' entropies for each of the three datasets' categories. The categories entropy distributions are shown in Figure 3. We can observe that the stream taken from Sheffield in normal conditions (labelled as "Week End" in the graph) presents denser regions in higher entropy levels.

[4] https://dev.twitter.com/

Order	Common	Food Festival	Tramlines
1	ff	ff	tramlines
2	sheffdocfest	foofighters	ff
3	blogsmoda	sheffield	buskersbus
4	ofs	notw	replacewordinamoviewithgrind
5	bbcf	totb	sheffield
6	blkstg	bbcf	amywinehouse
7	nosleeptilleadmill	titp	swfc
8	underwearshongs	swfc	allabouttonight
9	articmonkeys	sonishphere	hallamfm
10	beards	believe	forgetramlines

Table 4. Top 10 Most Frequent hashtags

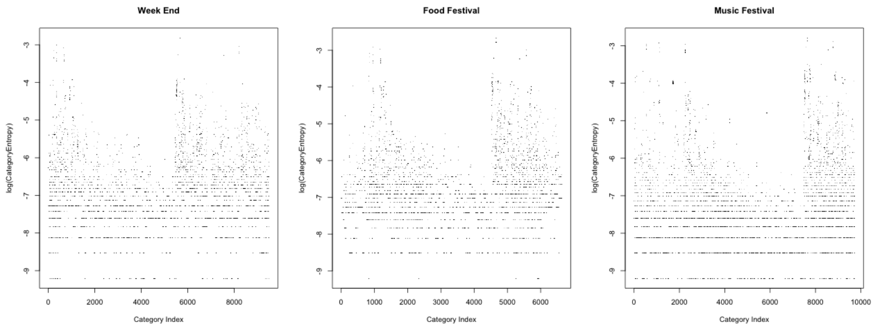

Fig. 3. Category Entropies vs. Category Index

Since lower category entropy levels provide a better information gain, we pick a category entropy threshold from which to pick categories. For these data sets and following Figure 3 we picked -9 as a threshold obtaining: 255 categories for the common week-end, 28 categories for food festival, and 562 categories for Tramlines. Table 5 shows the top 21 categories for each stream.

It is important to notice that we are not biasing the results by picking a priori hashtags relevant to the week-end events, but rather the categories emerge from category entropy analysis. From Table 5, very disparate categories appeared for the week-end in normal conditions ("common"), while for the Food Festival week-end we find categories which appear to be related either to external events or future events (Music_Festivals), as well as categories related to a current event (Food_companies_of_the_United_Kingdom). Incidentally for the food festival week-end we found two sets of semantically coherent categories, the first (categories from 13-17) matches an external event related to the 2012 Olympic tickets sales, while the second (categories 18-23) appears to be closely

Order	Common	Food Festival	Tramlines
1	History_of_the_Middle_East	Music_festivals_by_country	Arts_occupations
2	Mediterranean	American_Roman_Catholics	Music_industry
3	Near_East	American_people_by_ethnic_or_national_origin	Disco
4	Western_Asia	Food_companies_of_the_United_Kingdom	Dance_music_by_subgenre
5	Geography_of_Iraq	Public_opinion	DJing
6	Geography_by_country	Youth	Electronic_music
7	Cultural_history	Students	New_York_culture
8	Argentine_culture	Education	New_York_City
9	Argentine_society	Adolescence	Rock_music_genres
10	Nicaraguan_culture	Sport_and_politics	Rock_music
11	Languages_of_Colombia	Athletic_culture_based_on_Greek_antiquity	Underground_culture
12	Zambian_culture	Athletics_in_ancient_Greece	Postmodernism
13	Ike_&_Tina_Turner	Olympic_culture	Types_of_subcultures
14	Sun	Olympics	Youth_culture_in_the_United_Kingdom
15	Social_groups	Sport_and_politics	British_culture
16	Corporate_groups	Olympic_competitors	Youth_culture
17	Cognition	Sports_competitors_by_competition	Pejorative_terms_for_people
18	Prejudice	La_Liga	Slang
19	Critical_thinking	People_associated_with_Glasgow	Stereotypes
20	Social_class_subcultures	Football_in_Spain	European_Union_member_states
21	Romani_loan_words	Footballers_in_Spain_by_club	European_Union

Table 5. Top 21 Categories (sorted by category entropy (decreasing order))

relevant to an event involving Spanish football. We can observe that the categories obtained for the Tramlines Music Festival are more semantically coherent compared to the other two week-ends. This could be due to a higher relevance of the tramlines event compared to other events occurring at the same time in the city or externally.

Although some of the categories emerging from the category entropy analysis give an insight of endemic events, there are also other categories which provide information of events occurring externally. Hence, a Point of Interest considered as a Location-Entity presents the “meformer” and “informer” patterns observed by Naaman et al. [9] in Person-Entity activity streams. In this case the “Meformer” pattern refers to a self focus of a Location-Entity, presenting information about endemic events, while the “Informer” pattern refers to an information sharing of external events, not necessarily related to this Location-Entity.

In order to provide a context in which the category is being used, we use the mutual information between categories and hashtags (see Equation 3), from which we obtain a set of hashtags that can be used to further derived related keywords (see Equation 5)

Category	Hashtag	Keywords
heightSlang	#jobs, #jheeze, #rihanna, #neversayneverdvd	earth, swag, concert
Music_Industry	dance_music	party,music,record

Table 6. Hashtags and Keywords derived for two category using mutual information (see Equation 3)

6 Conclusions and Future Work

The identification of category resources $\mathtt{R}_{cat}$ from a POI awareness stream $\mathtt{G_a(P')}$ can be considered as a multi-class, multi-label classification task. This becomes challenging when no assumptions can be made a priori on the type of classes that will classify future events. Our approach semantically enriches the information of the social stream by providing a DBPedia based categorisation.

We have presented a formalisation for describing geographically bounded social awareness streams, we have also provided an approach for deriving transient categorisations of points of interest. We have applied our methodology on a data set and we have presented an empirical analysis of our results.

Future work includes a quantitative evaluation of this methodology by using larger datasets in which events have been identified a priori, and against which we can evaluate the emerging categories resulting from our approach.

Questions still remain on how we could determine a semantic coherence metric, which could induce broader category clusters. A semantic cluster of these categories can provide a better insight to the kind of events to which they refer to. Take for example the categories found for the Tramlines event, although we know these categories are related to music, we still haven't inferred the broader category "Music_Festival".

Acknowledgements A.E. Cano is funded by CONACyT, grant 175203.Andrea Varga is funded by the SAMULET project, co-funded by TSB and Rolls-Royce plc/

References

1. F. Abel, Q. Gao, G. Houben, and K. Tao. Semantic enrichment of twitter posts for user profile construction on the social web. In *In proceedings of Extended Semantic Web Conference 2011*, May 2011.
2. L. Barkhuus, B. Brown, M. Bell, S. Sherwood, M. Hall, and M. Chalmers. From awareness to repartee: sharing location within social groups. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 497–506, New York, NY, USA, 2008. ACM.
3. A. Cano, S. Tucker, and F. Ciravegna. Capturing entity-based semantics emerging from personal awareness streams. In *Proceedings of the Workshop on Making Sense of Microposts (MSM2011)*, May 2011.
4. Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 759–768, New York, NY, USA, 2010. ACM.
5. J. Hightower. From position to place. pages 10–12, 2003.
6. N. Ireson and F. Ciravegna. Toponym resolution in social media. In *Proc., 9th International Semantic Web Conference*. ISWC 2010, 2010.
7. E. Laurier. Why People Say Where They Are During Mobile Phone Calls. *Environment and Planning D: Society and Space*, 2000.
8. J. Lin, G. Xiang, J. I. Hong, and N. Sadeh. Modeling people's place naming preferences in location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 75–84, New York, NY, USA, 2010. ACM.

9. M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: message content in social awareness streams. In *CSCW '08: Proc., 2010 ACM conference on Computer supported cooperative work*, pages 189–192, 2010.
10. E. Schegloff. Notes on a conversational practice: formulating place. *in Studies in Social Interaction Ed D Sudnow (Free Press)*, 1972.
11. C. Wagner and M. Strohmaier. The wisdom in tweetonomies: Acquiring latent conceptual structures from social awareness streams. In *Proc. of the Semantic Search 2010 Workshop (SemSearch2010)*, april 2010.
12. A. Weilenmann. "i can't talk now, i'm in a fitting room": formulating availability and location in mobile-phone conversations. *Environment and Planning A*, 35(9):1589–1605, 2003.
13. K. L. D. Z. S. Zhiyuan Cheng, James Caverlee. Exploring millions of footprints in location sharing services. In *5th International Conference on Weblogs and Social Media (ICWSM)*, ICWSM '11. ACM, 2011.