Analysis of Content and Activity in Geo-Social Networks for Place Recommendations

Anastasios Noulas
Computer Laboratory
University of Cambridge
name.surname@cl.cam.ac.uk

Cecilia Mascolo
Computer Laboratory
University of Cambridge
name.surname@cl.cam.ac.uk

Introduction While the area of Geo-Social networking (GSN) is growing rapidly, with the rates of user adoption and appearance of startups experiencing exponential growth during 2010, there are series of research questions that arise synchronously to this technological shift. We identify one of the primary ones to be the division of frameworks that will be able to analyse and model data sourced from those systems. The necessity for such direction primarily originates from the fact that the data generation process features properties that may require special attention. First, it is *quantity diverse* from the point of view that user participation in those systems may vary, for example at an individual level as one may forget to checkin at a place or may avoid it for privacy reasons. Quantity of data reception may also be across different geographical areas (e.g. cities) that have different levels of adoption. Second, GSN data is *quality diverse* as one may observe mobility (visits at exact locations) trajectories of users, activity (visits at types of places) trajectories or text. The latter, may be user communication content or tips and tags assigned to locations. Further in this document, we describe our first steps towards addressing some of the requirements in creating a layer towards the unification of GSN data that will facilitate the access of applications and services to it.

Modelling Geographical Areas We initially present a framework for the representation *geographical areas* by means of nearby points of interest. Given a geographic point p encoded through a (latitude,longitude) pair, we retrieve all nearby a locations l within radius r and their respective categories c. The constitution of the considered area in terms of places can be indicative of what a user could encounter in the area. Moreover, we incorporate social activity to our methodology, by observing the number of *checkins* at a location. As an experimental case that utilises the above representation, we apply a clustering algorithm over the city of London and we discover groups areas where similar activity patterns.

Topic-Modelling and Sentiment Analysis applied on Communication Content With respect to user generated content, such as text in Geo-social networks, we propose the application of two natural language processing based paradigms, each of which will serve a different purpose. Topic Modelling, applied on micro-blogging data featured in those systems, could provide information about what is a place or a general area about. In addition, sentiment analysis techniques could provide information about the emotional consensus of the user community across the geographic plane. We envision a system, that could employ this analysis in order to provide automated reviews on places and areas as people move by and discuss about them.

An eye to future applications Models of data generated in Geo-Social Networks could be useful to a wide range of applications and services. Urban planners could discover more about a city, since the dynamics of human activity on it are revealed as people checkin at places. Augmented reality applications and location based services could also benefit as new data recourses can be accessed at massive scales. In addition, social scientists could learn more on how social systems evolve as patterns of individuals and social interactions emerge. To close, we present an example of how GSN data could be used in an application context by considering a place recommendation scenario for tourists at two metropolitan cities.