Detecting Event Visits in Urban Areas via Smartphone GPS Data

Richard Schaller[1], Morgan Harvey[2], and David Elsweiler[3]

[1] Computer Science, Univ. of Erlangen-Nuremberg, Germany
richard.schaller@fau.de
[2] Faculty of Informatics, University of Lugano, Switzerland
morgan.harvey@usi.ch
[3] I:IMSK, University of Regensburg, Germany
david@elsweiler.co.uk

Abstract. In geographic search tasks, where the location of the user is an important part of the task context, knowing whether or not a user has visited a location associated with a returned result could be a useful indicator of system performance. In this paper we derive and evaluate a model to estimate, based on user interaction logs, GPS information and event meta-data, the events that were visited by users of a mobile search system for an annual cultural evening where multiple events were organised across the city of Munich. Using a training / testing set derived from 111 users, our model is able to achieve high levels of accuracy, which will, in future work, allow us to explore how different ways of using the system lead to different outcomes for users.

Keywords: detection of visited locations, GPS, location-based services.

1 Introduction and Background

Evaluation of search engine performance is important, has long since been a contentious issue and recent work has lead to the insight that it is important to measure more than simply whether returned documents are relevant or not. Equally important is how the user interacted with the search system and the actual outcomes: did the user watch any of the recommended TV programmes? Were any of the found recipes actually cooked, eaten and enjoyed? Did the exploratory search to find out more about a political party change the way the searcher voted?

Of course some of these situations will be easier to measure than others - the point is that if we could glean more information about outcomes of search sessions, it would inform more about the utility of the search system than the interaction data with the system alone. In this work we try to make it possible to investigate outcomes in one particular context - geographical information needs, which represent nearly one third of all mobile search needs [2]. Such needs are dependent on location in some way. Examples include, “where is the nearest bank?” or a request for directions to a place or event. We want to establish if

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 681–686, 2014.

search results associated with particular locations are actually visited. Focusing on the problem of evaluating search and recommendation systems for events, we derive and evaluate a model to estimate, based on user interaction logs, GPS information and event meta-data, which events were visited in the context of the Long Night of Music - an annual cultural event organised in the city of Munich. During one evening in May, pubs, discos, churches and museums host diverse musical events, opening their doors to over 20,000 people. Search and recommender systems exist for the night and it would be helpful to evaluate which events found using such systems were visited in which situations.

2 Related Work

Geographical information is increasingly important in IR research, as demonstrated by the introduction of a TREC track dedicated to contexual suggestion using locations [4]. There has been significant work on estimating the geographic location of images [8]. Location information is often used in social media for providing localised content and location-aware recommendations. Users can be clustered based on geographic patterns [7], which can be used, for example, to improve friend suggestion based on geographic proximity [3].

In geographical searches location is not only a vital input to search algorithms, but could be an important component in evaluating the success of such systems. If we can determine whether a user actually visited the place suggested or made it to the destination they requested then this tells us something about the success of our system i.e. the results were appealing and / or relevant enough for the user to try them out. This applies most strongly to recommender systems for travel or tourist situations [9] and this situation is the one we focus on in the remainder of the paper.

Establishing points of interest (POIs) from GPS data is not straightforward and has been investigated as a research problem [6,1]. Typical approaches tend to apply a fixed distance radius from POIs to establish whether or not the user's GPS signal originates from nearby and determine a visit [1]. Other approaches additionally account for the prominence of POIs and opening hours [6].

While the methods proposed to date function well in the contexts for which they have been applied e.g. distinguishing between activities such as shopping and dining [6] or between POIs in different cities [1], the models are not easily transferable to other usage contexts or where the necessary granularity and accuracy is very fine. We present a new model which accounts for similar and additional features to those in the literature that can be used for our desired task of establishing the outcome of search / recommendation sessions.

3 Experimental Setup

We developed an Android app with search, recommendation and browse features to help visitors discover events of interest to them during the *Long Night*. Once the user has chosen events of interest, the system can create a plan and guide the user between chosen events using a map display and textual instructions [see Figure:left 3.1]. We examined user behaviour by recording all user interactions and positional data from the app. The app was available on the Google Play Store and was advertised on the official event web page. It was downloaded approximately 1300 times and 1159 users allowed us to record their interaction data. GPS positional updates were recorded if users had GPS turned on resulting in GPS data from 180 users between 5pm and 5am on the day of the event.

The main aim of this work is to determine, for each user, the sequence of events visited and the duration of time spent at each venue. We split the whole process into two parts: Extraction of the visited events (and their order) and estimation of the start and end time of each event visit. For both the human labellers and the later automatic methods we make the following assumptions: Only one event can be visited at any given time, only open events can be visited and minimum dwell time at an event is 5 minutes.

3.1 Gold Standard Dataset

The main source of information used for deciding whether a user visited a certain event or not is the recorded GPS track. We therefore concentrate further analysis on the 111 users for whom we had GPS tracks between 5pm and 5am that were sufficiently reliable to establish the events visited, as determined by manual analysis. In addition to the GPS track and accuracy of GPS signal[1] we can also consider properties of the event (such as position, name/description and opening hours) and user interactions with the event (whether the event description was viewed and read and whether the event was rated for possible tour inclusion).

We developed a tool which displays all this information on a map and allows a (high-speed) replay of the evening (see Figure:right 3.1). We instructed two

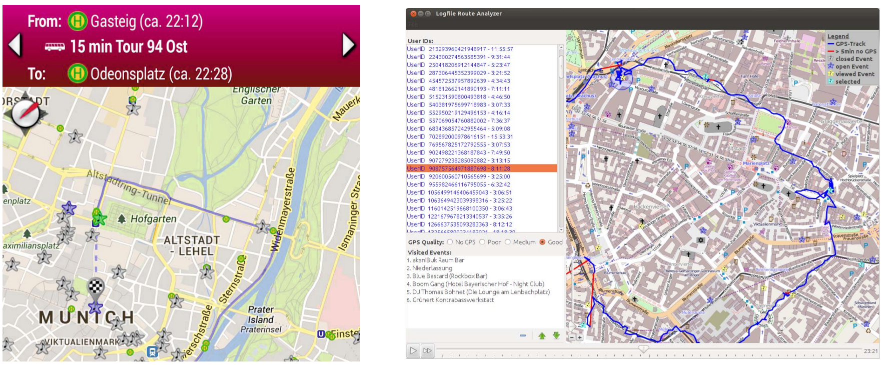

Fig. 1. Left: mobile app map view, Right: GPS track labelling interface

[1] As reported by the Android device. A circle around the reported location which contains the true position with 68% confidence.

human labellers to use the tool to mark the sequence of events each user most likely visited. Labellers were asked to consider all the information listed above, especially in case of multiple events at the same event location.

There are a number of aspects that make the GPS track difficult to interpret: GPS sensors are of very different quality, users may quit/restart the app or turn GPS on/off at any time, GPS does not work indoors or in streets flanked by high buildings and events are represented as a single coordinate with no information about the area of the event location given.

To compare the event sequences of both labellers we used two metrics: Jaccard-Similarity (JS) which returns a score between 0 and 1 indicating the overlap between two sets and Levenshtein-Distance (LD) which compares sequences by calculating the minimum number of edit operations (insertions/deletions) necessary to get from one sequence to the other. We compared the sequences of events from the labellers to determine what level of agreement was reached and obtained a JS of 0.818 and a LD of 1.09 event visits, suggesting generally very good overlap. Despite the manifold potential issues when working with GPS data we consider the itineraries generated by the labellers to be sufficiently accurate approximations of the true itineraries and therefore use these as a source of gold standard data for later analysis.

3.2 Experimental Systems and Their Performance

As features the model uses the same information as the labellers: GPS track and accuracy, event position and opening hours and user interactions with the event (whether the event description was read and whether the event was rated by the user). The model can also consider the popularity of the event as a prior, determined by how often all users have rated the event. For ease of computation we discretise the evening into time slots of 10 seconds. The model then calculates, for each such time slot, a probability of a visit for each event. This probability of event E having been visited given GPS measurement M and X, the true location of the user, can be formulated as:

$$P(E|M) = P(E) \int_x P(x|E)P(x|M)\, dx$$

The estimate $P(E|M)$ can be separated into the prior probability of visiting an event $P(E)$ and an integral over all possible locations $\in X$. $P(x|M)$ depends on the error of the GPS sensor which, according to the Android API documentation[2], can be modelled as a 2-dimensional normal distribution where the vector of means μ is the location received by the GPS sensor and σ is a scalar representing the reported accuracy. We can also model the event location probability $P(x|E)$ in a similar fashion, with μ being the location of the event and σ accounting for the variance in location when someone visits the event, which we assume to be equal for all events.

[2] developer.android.com/reference/android/location/Location.html

Thus we can solve the integral as

$$\int_x P(x|E)P(x|M)dx = \frac{1}{2\pi(\sigma_{\text{ev}}^2 + \sigma_{\text{gps}}^2)} \cdot \exp -\frac{(\mu_{\text{ev}} - \mu_{\text{gps}})^2}{2(\sigma_{\text{ev}}^2 + \sigma_{\text{gps}}^2)}$$

Note that $(\mu_{\text{ev}} - \mu_{\text{gps}})^2$ is the squared distance between the logged GPS coordinate and the event location. To factor in the extension of the building the event is located within we reduce this distance by a constant value S_{building}.

We calculate prior probability $P(E)$ of visiting an event based on a sum of 4 components controlled via scalar weights which must be trained. Three of these components are user dependent: The fraction of total event views for event E and whether of not event E was rated and/or selected for tour inclusion, normalised over the total number of events rated and selected by the user. The final component is the overall popularity of the event over all users, estimated as the number of times E was rated over the total number of ratings.

After calculating $P(E|M)$ over the set of all events, the event with the maximum probability is selected. Since it is possible that a user was not visiting any event at a given time we only predict a visit if $P(E|M)$ is over a threshold. Given an event visit (or no visit) for each 10 second time slot we then look at contiguous time slots with the same visited event: If the time between the first and the last time slot is longer than 5 min (our assumption of the shortest visiting time) then we consider this event to be visited during these time slots.

For evaluation we use the gold standard data described above (comprising 441 visits in total), split into 3 testing/training folds. We consider two problems: Determining a visited location and the more difficult task of determining the exact event visited (multiple events can take place at the same approximate location). We compare our model with a baseline system inspired by approaches in the literature in which a radius of 50m is set around each event[3]. An event is considered visited each time a user's GPS signal is within this radius of it for a minimum of 5 minutes. Both models were trained in order to maximise F1 score.

Table 1. Comparison of the baseline with the described model

	Location specific			Event specific		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Baseline	72.7%	78.6%	75.5%	54.8%	**77.8%**	64.3%
Our model	**81.6%**	**85.6%**	**83.6%**	**71.4%**	77.1%	**74.1%**

Table 1 shows the results from both models as measured by precision, recall and F-score. The proposed model outperforms the baseline in all but one case and always in terms of precision, arguably a more important metric than recall for this task. Our model delivers particularly strong performance (relative to the baseline) for the more difficult problem of detecting which individual events were visited, achieving an F1 score improvement of over 13.2%, highlighting the extra predictive power afforded by the extra features considered.

[3] This value is reported by [1] and our tests also indicated this to be an optimal value.

4 Conclusions

In this paper we have explored methods of using GPS and interaction data to determine whether or not geographical search results were visited during the Long Night of Music. Using log data from a naturalistic study of a mobile search system, we evaluated the performance of two models. The first, a baseline, derived from models for similar problems in the literature was outperformed by a second, new model, accounting for additional features. This model is able to achieve high levels of accuracy, offering huge potential to understand the success of different features of the system and of different user behaviour patterns. In future work we will investigate how user behavioural strategies related to concrete outcomes for the evening, using our model to derive metrics such as the number event visits, the popularity and diversity of events visited, as well as the time spent travelling. We believe the method we have used could be tailored to be equally fruitful for similar problems such as geographical queries in mobile search.

Acknowledgments. This work was supported by the Embedded Systems Initiative (`http://www.esi-anwendungszentrum.de`).

References

1. Bohte, W., Maat, K.: Deriving and validating trip purposes and travel modes for multi-day gps-based travel surveys: A large-scale application in the netherlands. Transportation Research Part C: Emerging Technologies 17(3) (2009)
2. Church, K., Smyth, B.: Understanding the intent behind mobile information needs. In: Proceedings of the 14th International Conference on Intelligent User Interfaces, pp. 247–256. ACM (2009)
3. Cranshaw, J., Toch, E., Hong, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Ubicomp (2010)
4. Dean-Hall, A., Clarke, C.L., Kamps, J., Thomas, P., Voorhees, E.: Overview of the trec 2012 contextual suggestion track. In: Proceedings of TREC, vol. 12 (2012)
5. Hauff, C., Houben, G.J.: Geo-location estimation of flickr images: social web based enrichment. In: Baeza-Yates, R., de Vries, A.P., Zaragoza, H., Cambazoglu, B.B., Murdock, V., Lempel, R., Silvestri, F. (eds.) ECIR 2012. LNCS, vol. 7224, pp. 85–96. Springer, Heidelberg (2012)
6. Huang, L., Li, Q., Yue, Y.: Activity identification from gps trajectories using spatial temporal pois' attractiveness. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. ACM (2010)
7. Scellato, S., Mascolo, C., Musolesi, M., Latora, V.: Distance matters: geo-social metrics for online social networks. In: Proceedings of the 3rd Conference on Online Social Networks, p. 8. USENIX Association (2010)
8. Serdyukov, P., Murdock, V., Van Zwol, R.: Placing flickr photos on a map. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 484–491. ACM (2009)
9. van Setten, M., Pokraev, S., Koolwaaij, J.: Context-aware recommendations in the mobile tourist application compass. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 235–244. Springer, Heidelberg (2004)