Predict User In-World Activity via Integration of Map Query and Mobility Trace

Zhengwei Wu
Baidu Research, Big Data Lab
wuzhengwei@baidu.com

Haishan Wu *
Baidu Research, Big Data Lab
wuhaishan@baidu.com

Tong Zhang
Baidu Research, Big Data Lab
zhangtong10@baidu.com

ABSTRACT

People often resort to map search engine or other location-based services for location information when planning long trips or local navigation, and their map queries as well as mobility trace will be accumulated and stored in user log. These data offers valuable information for studying the mechanism of human mobility pattern, furthermore, map query data enable us to sense users' real-time interests towards locations, and even to forecast their in-world activity in the near future.

In this paper, we unveil the connection between users' map queries and their in-world explorations, and prove the predictability of query-activity formation using two large-scale datasets, the complete map query log and mobility traces of 4 million Baidu map users, which comprise of 118 million map queries and 6.5 billion GPS location records during consecutive 3 months. To the best of our knowledge, it is the first attempt to extensively assess the unique qualities of map query data and predict whether queries about one location would actually lead to in-world visits using heterogeneous data sources. We first characterize the properties of these two datasets, then extract interesting features to quantify their correlation, finally we construct gradient boosting model for prediction, and describe applications empowered by our findings, such as mobility modeling and urban flow estimation.

General Terms

Human Mobility, Urban Computing, Spatial-temporal Data Analysis, Query Log Analysis

Keywords

Human Mobility, Urban Computing, Map Query Log analysis

1. INTRODUCTION

*corresponding author

Pervasive usage of location-based services and smart phones around the globe has contributed to vast and rapid accumulation of geolocation data. Imagine the following scenario: you search for the location of a new restaurant from Google maps on your smartphone and plan to meet a friend there; two hours later you drive from office to the restaurant via navigation service provided by Google maps; you then share current location by checking-in at the restaurant through Twitter and FourSquare app. After dinner, you issued a query about the airport and hotel information in a new city where you intent to go for a trip in the forthcoming vacation. Meanwhile, your map queries and GPS coordinates are captured by these application providers with your consent.

As a result, huge amount of map query and trajectory records were produced. In China, Baidu, the largest Chinese search engine, also possesses a dominant advantage in mobile map applications and share over 60% in the domestic mobile map market. Our research is supported by the large-scale datasets collected from anonymous Baidu mobile app users.

Such geocoded map queries imply users' strong intent of visiting the queried location, while the captured mobility traces characterize the mobility pattern and preference of historically visited places[11][20]. Integrating and mining such two data sources has a great potential for various commercial applications. Previous researches show that mobile users tend to perform location search queries when they plan to visit more unfamiliar places[18]. If we are able to validate whether the user visits the searched places by checking historical traces, we could model the relationship between location queries and mobility behavior. Furthermore, we are not only able to predict future mobility from location histories, but also anticipate users likelihood of visiting new places after user issued map queries. This is fundamental to many location-based services such as travel package recommendation[12], precise customer targeting [15] and predictive search[7] etc.

In this paper, we emphasize on investgating the unique properties of map query data, and assessing its strength to affect in-world activities, via integration with mobility trace data. We made three contributions in this work. First, we model the connection between users map query intents and in-world location exploration, on two large-scale datasets which comprise of 4 million Baidu mobile map users during 3 month as shown in Figure 1. To the best of our knowledge, it is the first time to predict mobility behavior by combin-

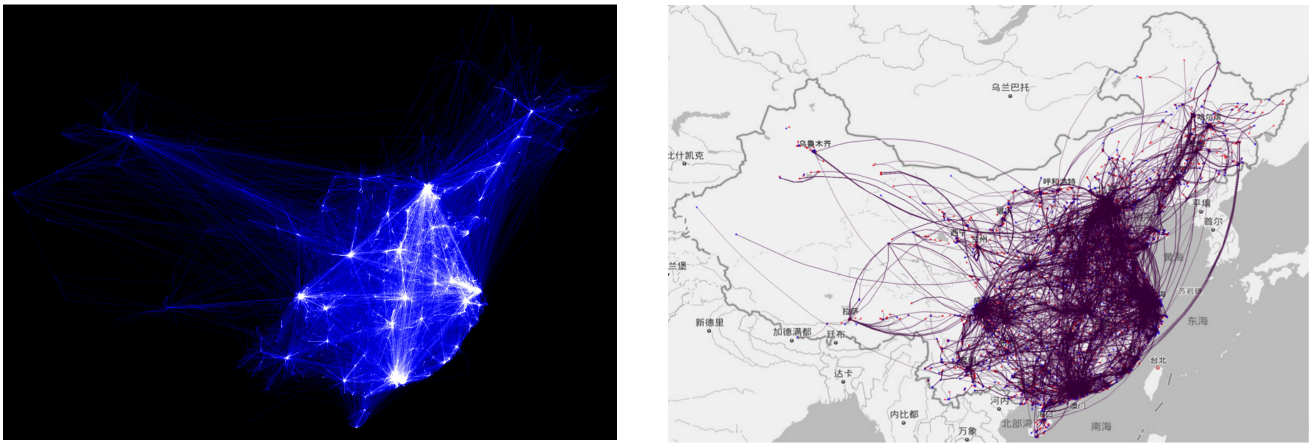


Figure 1: Left: visualization of mobility trajectories of 8751 users. Right:visualization of map query OD(origin-destination) network of 8751 users. Blue color indicates the location where the query was issued, red color indicates the queried location, and curves indicate the queried locations that were actually visited.

ing such two heterogeneous data sets. Second, We extract extensive query-activity features to characterize the correlation between online location query and offline in-world visit. Finally using extracted features as inputs, we build gradient boosting model for prediction task.

The rest of paper is organized as following: first we survey through recent work on human mobilty, geocoded query mining and urban computing in Section 2, and elaborate the contribution of our work; afterwards we discuss two large-scale datasets, i.e., map query data and mobility trace data, then discuss how to integrate them for analysis in Section 3. We address the details concerning how to extract important features, then propose and evaluate gradient boosting model for query-activity prediction task in Section 4 and 5, respectively. Finally, we conclude our findings and discuss future work and promising applications enabled by our research.

2. RELATED WORK

From theoretical perspective, Song et al. [14] quantified the predictability of human mobility using three different entropy measures, showing that 93% of human mobility behaviors are predictable. Numerous models have been proposed to predict future human mobility, but few have incorporated users' map query data as additional source.

Search engine is the gateway to Internet information, and massive search queries are capable of crowdsourcing real-time mass interests and help forecast future trends of a vast population. Many applications have utilized Google search queries to predict economy and disease outbreaks.[3] Teevan et al.[17] provided a detailed analysis on users mobile local search behavior, showing that local searches were highly contextual. West et al.[18] performed a thorough analysis on geocoded query logs and introduced a statistical model to infer the interplay of location queries and users' in-world activities. Their research shows that geocoded queries hold valuable information for user future movement prediction. Yang et al.[19] mined geocoded mobile queries and applied their model on predicting users future visit to healthcare utility.

These models, as the authors admitted in [18], were not able to quantify the factors that influence visit-after-query behavior, as they lack the ground truth of actual movement to validate whether user did visit the queried locations. Besides, compared with the geocoded data used in the work mentioned above [10] [17] [18] [19], location and routing queries on mobile map service that we used here consist of both geographic and POI description, thus reflecting users' intents towards future movement in a more direct way.

There is also some work worth noting which studies individual mobility patterns or urban dynamics using geotagged social media data.[5] [21]

3. DATA PROCESSING AND INTEGRATION

The whole framework is illustrated in Figure 2. In this section we focus on the data processing and data integration part.

3.1 Data processing

3.1.1 *Map Query*

When a user searches for a place or plans a route using Baidu map app, the query keyword, i.e. the destination, and user's current location will be recorded and stored to Baidu's database server provided the user's consent. We here collected more than 118 million historical map queries of 4 million users over 3 months since January 1, 2015, that is, 29 map queries per user on average. Each map query record has the following information: anonymized user ID, timestamp, latitude and longitude of the location where the query was issued, query keyword, property of point of interest (POI) which was resolved from query keyword, including POI address, category, detail information. Besides, when user plans a route, the location of specified destination, origin and selected transportation mode including walking, driving, and transit will also be recorded. About 10% map queries that cannot be resolved into valid GPS coordinates mainly due to erroneous user input, are discarded during preprocessing. We observed that in route planning, most cases are that the origin is the location where the map query is issued, while

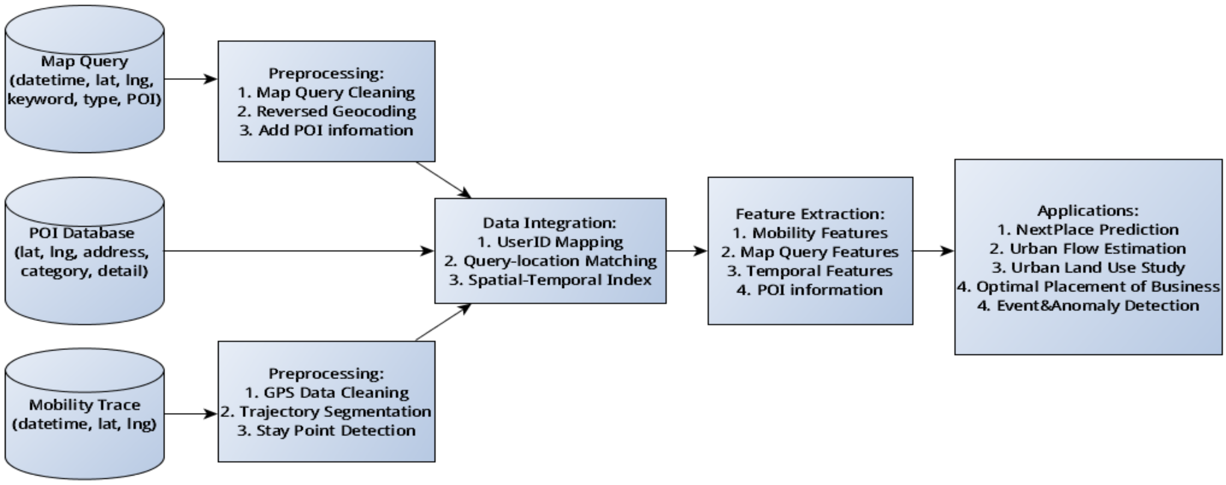


Figure 2: Diagram of proposed system

almost 10% of the origin direct to other places rather than user' current location. The distribution of average origin-destination distance of different transportation mode also follows power law distributions while different transportation modes exhibit different long tails.

3.1.2 *Mobility Trace*

We collected mobility trace data of the same 4 million Baidu mobile map users, while each trajectory records 3-month duration historical locations captured by Baidu mobile geolocation SDK since January 1, 2015. Location data are stored on the server if the user consents to the privacy policy of Baidu, and each record contains hashed user ID, timestamp, latitude, longitude, reverse geocoded address and location precision radius. Specifically, user's location may be determined by various means including GPS, wireless network spot or cell tower and the average accuracy radius is less than 60 meters. Incorrect location records due to network failure were removed during preprocessing.

We then identify the stay locations from historical mobility trace data. First, we classify mobility mode into three states, that is, stay, pause and move, based on movement distance and time interval between two consecutive location records. Afterwards we extract the location records with state of stay, and employ DBSCAN[6], a efficient density-based spatial clustering algorithm, to identify user's stay points. It performs best in our experiments compared with other classic clustering methods like MeanShift algorithm[4] and K-Means algorithms. Finally, each user has 20 stay points on average, and the number of stay points range from 1 to 158, influenced by user's mobility regularity level and GPS record density. We determine user's home and workplace leveraging on the visit freqency and hour-of-day distribution of each stay point, generally the top two most frequently visited stay points correspond to user's home and workplace.

3.2 Data integration

To quantify the relationship between location queries on map and actual movement in mobility traces, one essential step is how to integrate them together. We first established correspondence between them via hashed user id. We set the maximum time interval and minimum distance threshold of query-activity to be 15 days and 200 meters, respectively. In other words, we assume a queried location is actually visited by the user if there is at least one location record in user's next 15-day mobility trace matched at a distance less than 200 meters. Figure 5 shows the distributions of distance from the location where a map query was issued to the queried location where the user indeed visited afterward. As observed, about 20% of OD(origin-destination) distances are above 100km, indicating that people tends to plan their routes in advance for long distance travels. Another fact that we observed is that people usually queried a general keyword like 'Beijing' or 'airport in Beijing' before they plan a travel in a new city. To better handle such cases, the minimum distance threshold is designed to scale with origin-destination (OD) distance, and here we set the scaling factor to be 0.05, i.e. when user queries a location 1000 km away, the query-trace match is established if OD distance is less than 50km. Thus, 40-50% user map queries can be confirmed as actually visited afterwards by matching with their future mobility traces, the acutal percentage varies due to different levels of trace density, since some users might issue queries about a place and visit there later, but we could not observe this through sparse user traces. Figure 3 illustrate the spatial-temporal relationship between map queries and mobility traces.

4. CHARACTERIZE MAP QUERY AND MOBILTY TRACE

4.1 Query-Activity Relationship

4.1.1 *Time Lag between user's map query and in-world activity*

We quantify the relationship between map query data and mobility trace data after integration as detailed in previous section. Figure 4 shows the distribution of time interval

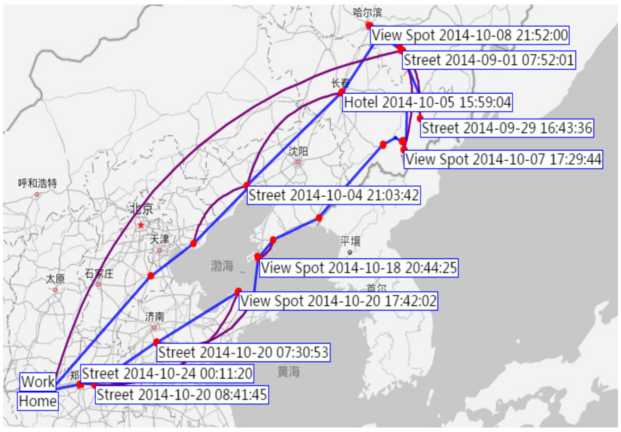


Figure 3: Queries and trajectory timeline of a user's journey during October, 2014: blue straight lines indicate mobility trajectories and curves indicate the queried routes. Each query, which contains the query timestamp and POI category of queried location, are labeled on the location where it was issued. We see that his query OD lines are always one-step ahead of his in-world trajectory.

between the time user searched for the place and the time he/she visits there. Not surprisingly, it is well characterized by heavy-tail distribution: about 25% matched query locations are visited instantly within 30 minutes, and over 50% within 2 hours, while 81% are within one day and 95% within one week. One interesting finding of this distribution is that there is a local extrema at around 6 and 24 hours. One possible interpretation is that if a trip takes more than 6 hour (the longest flight in domestic China is less than 6 hour), people tends to visit this place tomorrow or even later.

Figure 5 shows the distributions of distance between queried destination and user's home/workplace, respectively. We found that over 50% queried locations are confined within 50 km away from users' home or workplace, while there is no significant discrepancy between home and workplace(54.5% and 51.1%,respectively). By contrast, when we compared the frequency of map queries that performed in home and workplace, we found that the frequency of home-based queries is more than 2 times higher than that of workplace (23% and 9.4%, respectively). This indicates that users tend to navigate for new places when they are at home or nearby.

4.1.2 *Influence of Query Type*

We then quantified the influence of route planning service on query-activity behavior. Intuitively, querying direction information indicates stronger intention of visiting the queried location. Figure 6 demonstrates that route planning service obviously increases visiting probability, while the influences of three transportation modes (e.g. walking, transit and driving) provided by mobile map are different. Interestingly, users are more likely to visit the queried location after checking walking direction than transit or driving. This might be because the user is much closer to the destination in such case.

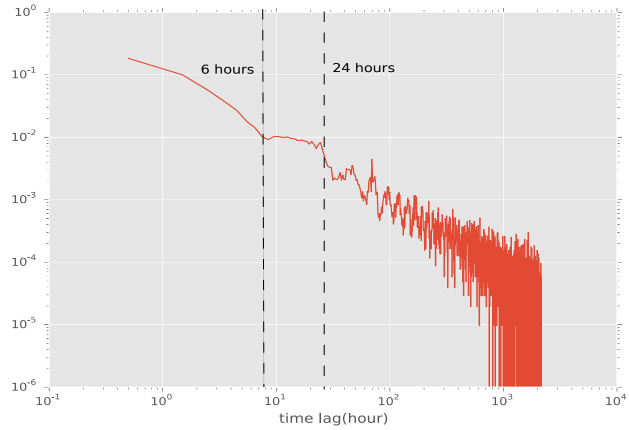


Figure 4: Distribution of the time lag between users' map queries and their actual in-world visits. It is worth noticing that there is a local extrema around 6 hours and 24 hours.

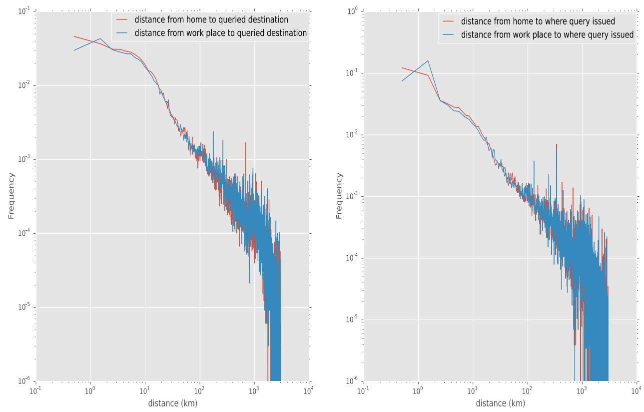


Figure 5: Left: Distribution of distance from users' home and workplace to the destination location of their queries. Right: Distribution of distance from users' home and workplace to the origin location of their queries.

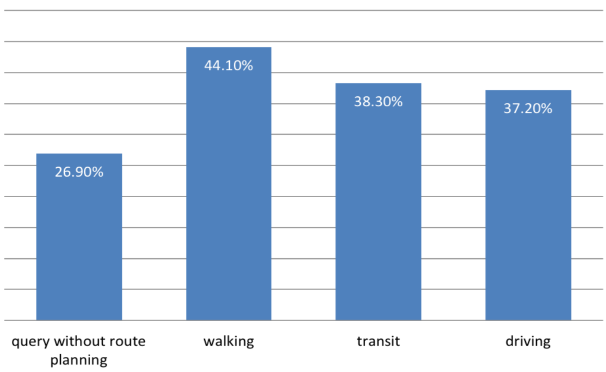


Figure 6: Visiting probability of queried location after using route planning service

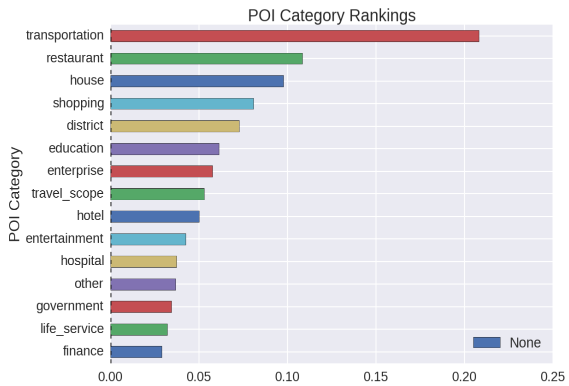


Figure 7: POI category rankings of map query destinations. Top 2 are transportation and restaurant, respectively.

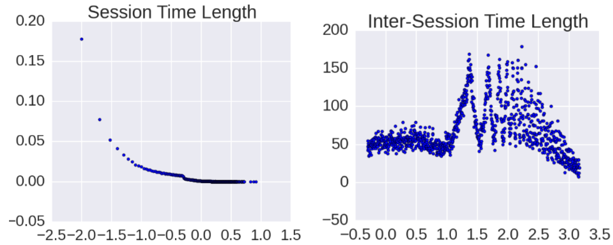
4.2 Point-Of-Interest Information of Map Query

Another factor that could influence such relationship is the POI information of queried locations. Most queried locations in our dataset are classified into 14 POI categories such as transportation, restaurant, hotel, etc. As for the remained 20% queried locations without category information, we first employ a CRF-based algorithm[13] to segment the keywords into tokens, then vectorize the tokens into one-hot encoding features, and use them as input to train a multi-label SVM classifier, which can automatically assign POI categories. This simple method can achieve overall accuracy of 92% when test on the labeled query records. We computed the percentage of each category for the queried locations which were actually visited by users. As shown in Figure 7, top 2 categories, of which the percentage are significantly higher than other types and almost account for 36% of all matched queries, are transportation (such as airport, railway stations etc.) and restaurant. It is also interesting that different categories of queries often exhibit distinct dynamics when compared under hour-of-day time interval, e.g. in Figure 8, restaurants are most queried during dining time, and hospitals are most queried at the begining of daily reception hour. Therefore we later use queries' POI category information of queried location as a feature to feed our prediction model.

4.3 Query Sessions

Extensive research has discovered that a variety of human behaviors express bursty nature and can be well-characterized by heavy-tailed distribtion. As we observed in users' map query log, one or several location queries within a short time often follow a long period of inactivity. Therefore it is reasonable to view a user's historical map queries as consecutive sessions, and monitor their characteristics and in-session/inter-session interactions.

We apply 30 minutes as time window to separate a series of queries into sessions. If time interval between two consecutive queries are shorter than 30 minutes, they are identified as the same session, otherwise would fall into different ones. We also record important features of each query session, such as number of query items, session duration and inter-session time.



(a) In-session time(log hour) (b) Inter-session time(log)

Figure 9: Property of Query Sessions

Map queries are much more concise than general search queries, because they express user's instant and direct interest towards locations. User would in often times rephrase query keywords in order to obtain better results from search engine, and these queries natually form a query session.

One interesting fact is users' repeated query behavior, which mostly occurs within one session and can be classified into three types: the first one is similar as what observed in general search engine[16], users frequently rephrase their location queries in short time period to obtain the optimal query feedback, such as trying full name or abbreviation for the same place; And the second one, user intends to visit one or certain categories of locations, and search for all the places that could suffice his/her current needs, then choose one among them to visit after comparison; The third one is that users usually search for the same location on Baidu map before leaving for the destination (for planning new route) or being very close to the destination (for checking the exact location). Such behavior indicates that the user is very inclined to visiting this place, or is already on the way. Intuitively, the more frequent users repeatly search for one location, the higher probability that he/she will visit this place in the near future.

According to our observation, 37.7% sessions consist of only one map query and have 0 session time length. For other query sessions, the session time length ranges from a few seconds to 8 hours, and is around 13 minutes on average. The overall distribution follows a long-tail distribution, as shown in Figure 9, Left, there is a minor gap near 30 minutes(10e-0.3 hour) due to the inter-session time window setting. And the inter-session time length distribution shows a periodic peak almost every 12 hours, indicating temporal regularity in query behavior.

4.4 Feature extraction

Based on the discoveries above, we extract the following features for query-activity prediction.

4.4.1 Map query features

Map query features are defined as following:

- Query session features: We segment user queries into consecutive query sessions via a inter-query time window of 30 minutes, and calculate session size(number of in-session queries), session time length and inter-

(a) Restaurant (b) Transportation (c) Hotel (d) Hospital

Figure 8: Queries of different POI categories exhibits distinct temporal dynamics in Hour-of-Day popularity: (a): daily popularity of restaurant queries peaks at 1am, 12pm, 18pm, mostly during dining time. (b): daily popularity of traffic queries are highest after midnight, when other needs, such as food and shopping, become dorment. (c): daily popularity of hotel queries are higheest during night time. (d): daily popularity of hospital queries peaks at 8am, when most hospitals in China begin to receive patients.

session time length, to depict the property of this session.

- Query OD distance: this is a vital factor that affects user's decision of visiting a location or not.[18]. In our validation, we observed that the average OD distance of visited queries is significantly shorter than that of non-visited query locations.
- Distance between query origin, destination and home, workplace, the nearest stay point. As perceived, a large portion of map query requests occurs close to users' stay points(home, workplace), but once users deviate from their familiar daily routes, they tend to issue queries more frequently to navigate their surroundings. And when the user has a list of candidate locations in mind, the nearest place often stands out, either nearest to the current location, or nearest to the locations he/she is already familiar with(stay points).
- Query type: search, walking, driving or transit. They represent different transportation mode and have shown distinct query-activity matching percentage.
- Query index and number of matched queries in the past: context information contains features about user's past query behavior, and generally when a user issued queries about many places and visited them afterwards, he/she is more likely to repeat this behavior on current query session.

4.4.2 *Mobility features*

With mobility trace and calculated stay points in hand, we compute the following features for query-activity prediction:

- Hour-of-day movement distance: we compute the average daily movement distance for each hour-of-day, extreme cases, such as a single long trip that exceeds 100 km, are excluded during calculation.
- Hour-of-day GPS records percentage: we compute how a user's GPS records are distributed on hour-of-day, since the generation of GPS records indicates that the user is currently using Baidu map, this feature helps describe the variations of user's daily activity level.
- Radius of gyration: a widely-used measure for quantifying the typical range of a user's trajectory.
- Commute distance: most people's daily movements are formed of home-work commutes, and commute distance represents their normal movement and radius of gyration level.
- Travel frequncy: frequency of long travels that exceed 100 km, this feature is included to complement the information of hour-of-day movement.
- Temporal features: Human mobility presents particularly high temporal regularity, i.e. people tend to stay at certain locations during a specific time period. Meanwhile, it is perceived that the probability of visiting queried locations with long distance during weekend or holiday is much higher than that of weekdays. Thus we should include temporal information of the query as features: hour of day, day of week, weekday/weekend/holiday.

4.4.3 *POI features*

- POI category: transportaion, restaurant, real estate, shopping, district, education and university, enterprise and industry zone, travel spots, hotel, entertainment, hospital, government, life services, finance and banks. As discussed above, POI category characterizes the function of queried location and reflects user's purpose of visit, therefore can be a valuable input in query-activity prediction.
- POI query number, visit number, and popularity ranking: based on each POI location's total historical frequency of being queried and historical visits, we calculated the popularity ranking of the queried location.
- POI category matched ratio: the ratio of visit number and query number towards each category of POI location. e.g. hotel POIs have the highest ratio of 52%.

5. PREDICT IN-WORLD ACTIVITY VIA MAP QUERY

5.1 Datasets

Our datasets consists of two parts: users' aggregated map query log and their mobility traces. First we have 4,080,283 Baidu map app users who had issued map queries in Beijing City on January 1, 2015, afterwards we collect and integrate their entire map query log and mobility traces un-

Table 1: Query OD City Percentage

Model name	query number	precentage
Beijing->Beijing	64,278,560	54.2%
Beijing->Others	3,407,144	2.9%
Others->Beijing	9,242,424	7.8%
Others->Others	37,890,702	35.1%

til March 31, 2015, altogether 90 days. Finally when invalid map queries and location records are excluded, we are able to conduct quantitative analysis using 118,671,390 map queries and 6,493,688,641 user GPS location records, on average 29.1 map queries and 1591.5 GPS records for each user, respectively. According to our validation with user mobility tracse, 42.6% map queries lead to users' in-world activity at 4,994,373 unique POI locations throughout China.

Among all the query records, 57.26%(67,953,813) are issued from Beijing City, since we start data collection process from Beijing local users, and 61.95%(73,520,984) query destinations are within Beijing.

All query and mobility data are collected with users' content, and their IDs in our experiments are indexed and profile information omitted to guarantee anonymity. Compared with similar data sets used in other related work, our data is significantly more vast and consecutive, therefore reducing the common limitation of data scarcity problem in characterizing human mobility pattern.

5.2 Gradient Boosting Model for Prediction

Map query data provides rich information to infer future travel intents. As shown in section 4, we investigated the relationship between map query and in-world visit behavior, and extracted a set of related features to characterize such behavior. Taking these features as inputs, we here build probabilistic models to predict that in which new place the user will appear and when.

5.2.1 Problem Formulation:

Given user's current location query, historical mobilty trace, and historical map query log, including:

$$\{(t_0, O_0, D_0), \cdots, (t_c, O_c, D_c)\}$$

where O, D denote origin location and queried destination location, respectively, our goal is to infer whether user would actually visit D_c in the future, i.e. user being observed less than 200 meters from D_c within next 15 days' mobility trace records.

5.2.2 Model Settings:

We model this task as a binary classification problem, and our goal is to predict whether the user will visit a place after he issued queries on Baidu map. We propose to use gradient boosting classifier [8], a powerful additive model that has shown great success in a variety of academic problems and industrial applications.

Here we resort to XGBoost(eXtreme Gradient Boosting), an open source gradient boosting library which also provides an optimized distributed version, implemented by [2]. It has been widely adopted in many data mining competitions, and is best-known for its outstanding performance and efficiency for training.

Table 2: Query-activity Prediction Model Performance

Features	Error Rate	AUC
Query Features	0.206	0.841
Mobility Features	0.218	0.830
POI Features	0.226	0.817
All	**0.189**	**0.869**

Two most important parameters in the model, the max depth of boosted trees and the number of trees, are tuned to 8 and 20, respectively. Another useful features is that XGBoost supports probability values in [0,1] as label. Since the validation of query-activity is not definite, i.e. matched queries may be only passed-by instead visited, and unmatched queries may be just unable to confirm due to trace sparisity, we therefore assign probabilistic labels 0.99 and 0.01 to positive and negative samples, respectively.

In order to avoid overfitting and ensure model's robustness from possible errors in extracted features and query records, we evaluated our model via 5-fold cross validation.

5.3 Experiments

5.3.1 Evaluation metrics

We use test error(logistic loss) and AUC score(area under the ROC curve) as metrics to quantify the performance of boosting model, and the predictability of user's in-world activity based on his/her map query and historical mobility trace features.

All features are grouped into 3 categories, and are used as input separately to further evaluate their influence on model's prediction accuracy. We also calculate the detailed importance score for each feature in the complete training model.

5.3.2 Results

As shown in Table 2, on query-activity prediction task, our model can achieve test error of 0.189 and AUC score of 0.869 using all the features as training input. We think it convincing that query-activity formations are predictable at individual level.

5.3.3 Feature Importance:

All the features used in our model were extracted based on our observation as detailed in Section 4. We here evaluated the importance of each feature as demonstrated in Figure 10.

- **Query features:** Query features prove to be the most important factor on determining user's in-world visit behavior. This verifies our empirical assumption mentioned in Section 4 that the more frequently one user issued queries about the same location, the higher probability that the user will visit the queried place in

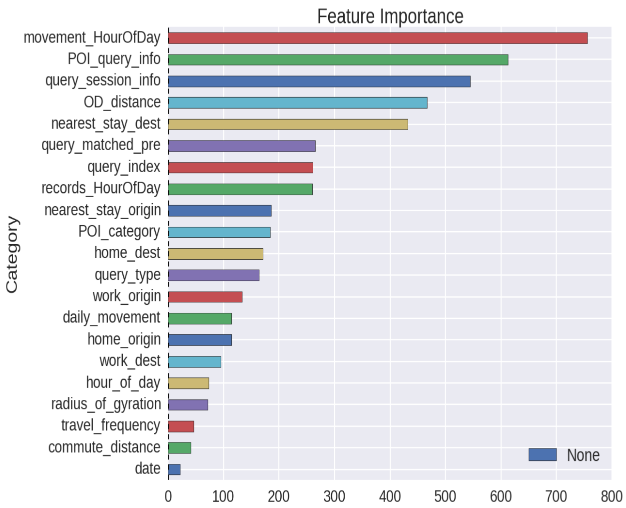


Figure 10: Feature importance rankings in query-activity prediction model

the future. Besides, the time interval between repeat queries also expresses the intensity of user's interest towards visiting the queried location.

Origin-destination distance and home-destination distance are also useful features. As described in Section 4, the OD distance of visited queries are significantly shorter than unvisited ones.

- **Mobility features:** Mobility features such as the HourOfDay movement distance, number of HourOfDay location records, daily movement distance, radius of gyration and travel frequency account for a large portion of model performance in total. This indicates that future mobility behavior, including novelty seeking behavior, is also greatly influenced by one's daily mobility patterns.
  Temporal features including hour-of-day, day-of-week, holiday are less important than query features and mobility features. One possible reason is that novelty seeking behaviors are much more diverse in temporal range although human mobility exhibits temporal regularity.

- **POI features:** POI category information of queried location, not surprisingly, also plays an important role. This is because the category information indicates the function of queried location and is therefore directly related with people's in-world activities.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we integrate large-scale map query data and mobility trace data from Baidu map users, to characterize the interplay between users' queries and their in-world activity. After study of important traits of map query and mobility trace, we build gradient boosting model to predict future in-world activity of users, based on their current query and well-designed features, then evaluated the experiment results and assessed the contribution of each feature. Our research has great potentials for a large spectrum of applications ranging from POI recommendation, accurate customer targeting, to urban planning and modeling.

Next Place Prediction: Most existing work on next place prediction focus on modeling people's mobility regularity and inferring next location based on their historical transitions between stay points.[14][9] Though it yields good performance, but this mechanism fails to incorporate the novelty-seeking behavior of people. Therefore map queries, which best reflect people's location demands and interests towards novel places, can offer valuable additional information of users' future in-world activities, and further enhance the performance of traditional next place prediction models.

Urban Flow Estimation: Most flow estimation methods solely rely on historical or real-time collective user mobility traces as input, then assess the density zones and traffic flow[1]. This kind of system is sufficient in many scenaios, and capable of flow estimation under fine time granularity, but it is not quite robust when event or anomoly occurs. For example, when a public stadium is about to host a big event tonight, unusual amount of map queries about this location would emerge and soon be detected as the event time approaches, even though the real-time mobility flow in this region has only shown very little fluctuations. With the boost of map query information, we can easily sense and quantify the massive location demands from the crowd, thus calibrate and forecast urban mobility flow beforehand.

7. REFERENCES

[1] F. Alhasoun, A. Almaatouq, K. Greco, R. Campari, A. Alfaris, and C. Ratti. The city browser: Utilizing massive call data to infer city mobility dynamics.

[2] T. Chen and T. He. xgboost: extreme gradient boosting. 2015.

[3] H. CHOI and H. VARIAN. Predicting the present with google trends. *Economic Record*, 88:2–9, 2012.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

[5] J. Cranshaw, R. Schwartz, J. I. Hong, and N. M. Sadeh. The livehoods project: Utilizing social media to understand the dynamics of a city. In *ICWSM*, 2012.

[6] 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*, volume 96, pages 226–231, 1996.

[7] T. Fiatal. Predictive content delivery, Jan. 2 2009. US Patent App. 12/348,136.

[8] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[9] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[10] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel. Analysis of geographic queries in a search engine log.

In *Proceedings of the first international workshop on Location and the web*, pages 49–56. ACM, 2008.

[11] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[12] A. Majid, L. Chen, G. Chen, H. T. Mirza, I. Hussain, and J. Woodward. A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, 27(4):662–684, 2013.

[13] F. Peng, F. Feng, and A. McCallum. Chinese segmentation and new word detection using conditional random fields. In *Proceedings of the 20th international conference on Computational Linguistics*, page 562. Association for Computational Linguistics, 2004.

[14] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[15] J. Steenstra, A. Gantman, K. Taylor, and L. Chen. Location based service (lbs) system and method for targeted advertising, Aug. 31 2004. US Patent App. 10/931,309.

[16] J. Teevan, E. Adar, R. Jones, and M. A. Potts. Information re-retrieval: repeat queries in yahoo's logs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 151–158. ACM, 2007.

[17] J. Teevan, A. Karlson, S. Amini, A. Brush, and J. Krumm. Understanding the importance of location, time, and people in mobile local search behavior. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 77–80. ACM, 2011.

[18] R. West, R. W. White, and E. Horvitz. Here and there: goals, activities, and predictions about location from geotagged queries. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 817–820. ACM, 2013.

[19] S.-H. Yang, R. W. White, and E. Horvitz. Pursuing insights about healthcare utilization via geocoded search queries. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 993–996. ACM, 2013.

[20] Y. Ye, Y. Zheng, Y. Chen, J. Feng, and X. Xie. Mining individual life pattern based on location history. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*, pages 1–10. IEEE, 2009.

[21] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie. We know how you live: exploring the spectrum of urban lifestyles. In *Proceedings of the first ACM conference on Online social networks*, pages 3–14. ACM, 2013.