Travel Purpose Inference with GPS Trajectories, POIs, and Geo-tagged Social Media Data

Chuishi Meng[1], Yu Cui[2], Qing He[2], Lu Su[1], Jing Gao[1]

[1]*Department of Computer Science and Engineering, SUNY Buffalo*
[2]*Department of Civil, Structural and Environmental Engineering, SUNY Buffalo*
Email: {chuishim, ycui4, qinghe, lusu, jing}@buffalo.edu

Abstract—In our daily lives, travel takes up an important part, and many trips are generated everyday, such as going to school or shopping. With the widely adoption of GPS-integrated devices, a large amount of trips can be recorded with GPS trajectories. These trajectories are represented by sequences of geo-coordinates and can help us answer simple questions such as "where did you go". However, there is another important question awaiting to be answered, that is "what did/will you do", i.e., the trip purpose inference. In practice, people's trip purposes are very important in understanding travel behaviors and estimating travel demands. Obviously, it is very challenging to infer trip purposes solely based on the trajectories, because the GPS devices are not accurate enough to pinpoint the venues visited. In this paper, we infer individual's trip purposes by combining the knowledge from heterogeneous data sources including trajectories, POIs and social media data. The proposed dynamic Bayesian network model captures three important factors: the sequential properties of trip activities, the functionality and POI popularity of trip end areas. Extensive experiments are conducted on real-world data sets with trajectories of 8,361 residents and the 6.9 million geo-tagged tweets in the Bay area. Experimental results demonstrate the advantages of the proposed method on correctly inferring the trip purposes.

Keywords-Dynamic Bayesian Network, Trajectory, Social Media, Point of Interest, Trip Purpose

I. INTRODUCTION

In this big data era, we are able to collect a large amount of human trajectories because of the ubiquitous adoption of GPS-integrated devices. For instance, smart phones can track real-time trajectories and geo-tagged social media messages can also reveal users' trajectories. On the surface, these human trajectories record people's daily trips with a sequence of geo-coordinates. But in essence, they reveal people's activities or trip purposes, e.g., "shopping" or "eat out". With these abundant trajectory data in hand, we are curious to ask "what did/will you do when you arrive at one place?". This is the trip purpose inference problem.

The inference of people's trip purposes has many benefits for the whole society. First, people's trip purposes can help government officials understand travel behaviors and estimate travel demands which will lead to better city planning and investment decisions. In addition, it can provide customers with more accurate recommendations and better services, and this recommendation can be made even before users start their trips.

However, trip purpose inference is very challenging, because the GPS records are not accurate enough to pinpoint the venues visited, let alone revealing the purposes of the trip. Although there have been some attempts on this research topic [1], [2], [3], existing methods overlooked several important properties such that makes them less practical. Firstly, existing methods do not take fully advantage of the heterogeneous data, such as trajectories, POIs and social media. In addition, people's activities usually follow certain patterns, and there are intrinsic relationships among the sequence of activities. Take the trips shown in Figure 1 as an example. Parents may drop off their children at school before going to work, and people would eat in a restaurant after shopping. Similar activity patterns commonly reside across different users. As a result, modeling the whole sequence of activities can help us infer the past or future activities. Moreover, although nearby POI information is useful to infer the trip purposes [2], [4], their geo-graphical distribution cannot reveal how people like the venues. Obviously, not all the POIs attract equal attentions, and some of them are usually more popular than the others. Because the trip purpose is a people-centric concept, the popularity of the POIs would be useful for the inference.

In this work, we propose to infer trip purposes from a large amount of heterogeneous data sets, i.e., users' trajectories, POIs and social media messages. In order to capture the sequential property of the trips, this paper proposes a dynamic Bayesian network model in which the trip purposes are hidden variables. In addition, this paper proposes to incorporate POI popularity in the trip purpose inference, and mine the popularity of POIs from geo-tagged social media data.

In summary, this work makes the following contributions:

- We propose to infer individual's trip purposes by combining the knowledge from heterogeneous data sources including trajectories, POIs and social media data. The proposed dynamic Bayesian network model captures the sequential property of people's activities.
- We propose an effective method to extract mentioned POIs from geo-tagged social media messages, and
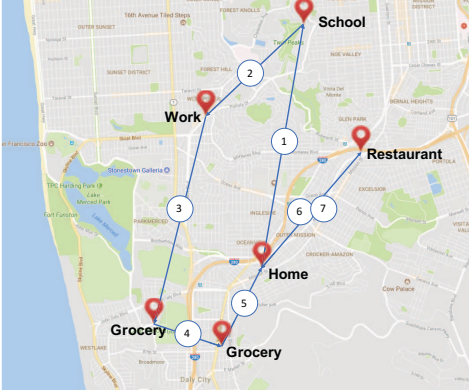
Figure 1: A typical user's daily trips

model POIs' popularities in trip end locations. This popularity knowledge can help improve the inference performance.

- We conduct extensive experiments on real-world data with trajectories of 8,361 residents and 6.9 million geo-tagged tweets in the Bay Area, CA. The results demonstrate the advantage of the proposed method on correctly inferring users' trip purposes.

The rest of the paper is organized as follows. We formally define the trip purpose inference problem in Section II. The proposed methods are detailed in Section III, and experiments are shown in Section IV. We review the related work in Section V then conclude the paper in Section VI.

II. OVERVIEW

In the following, we introduce several important concepts that will be used throughout the work, then formally define the trip purpose inference problem.

Definition 1. *A* Trajectory (Tr) *is a sequence of time-ordered spatial points, $Tr : l_1 \rightarrow l_2 \rightarrow \cdots \rightarrow l_n$ where each point l is represented by a pair of GPS coordinates, i.e., longitude and latitude.*

In this paper, we regard "trip" as the movement from one location to another, e.g., $l_1 \rightarrow l_2$, and we refer these GPS points l as "trip end locations". Take the trajectory in Figure 1 as an example. The trajectory comprises eight GPS points which follows the sequence of $l_{Home} \rightarrow l_{School} \rightarrow l_{Work} \rightarrow l_{Grocery} \rightarrow l_{Grocery} \rightarrow l_{Home} \rightarrow l_{Restaurant} \rightarrow l_{Home}$. From the perspective of trips, it has seven trips labelled by blue circles. However, in most cases, we cannot know the activities performed in a location without users' input. In other words, we don't know whether a user is shopping in a grocery store or having meal at a restaurant, given only the geo-coordinates of trip end locations.

Definition 2. Trip Purpose *is the activity that a user performed at a trip end location.*

As the name refers, it denotes the purpose of a trip. In the following sections, we use "trip purpose" and "activity"

Table I: Activity Categories

Category	Example Activities
Home	Any activities performed at home
Education	School, Class, Laboratory, Meal at college, etc.
Shopping	Groceries, clothing, gas, oil, etc.
EatOut	Drive through meals, Meal at restaurant/dinner
Recreation	Indoor or outdoor exercise, Health care, etc.
Personal	Household errands, religious activities, etc.
Work	All activities performed at the work place, etc.
Transportation	Change type of transportation, Pick up/drop off

interchangeably. As shown in Table I, we categorize all trip purposes into eight categories, i.e., "Home", "Education", "Shopping", "EatOut", "Recreation", "Personal", "Work", and "Transportation". The example activities shown in Table I are defined by the California Household Travel Survey (CHTS) [5] which collected people's daily trajectories and activities. More details about this data set will be discussed in Section IV.

Problem Definition. **Given** the trajectories of users, the points of interest and the Twitter messages near trip end locations, our **Objective** is to infer the purposes of trips.

Note that some of the trips have labels with corresponding purposes. These labels can be manually recorded by users or mined from social media messages. However, acquiring these labels is very difficult, and it is usually assumed to be unavailable for a large portion of trips.

III. TRIP PURPOSE INFERENCE WITH DYNAMIC BAYESIAN NETWORK

A trip's purpose is determined by many factors, such as other activities of the day, the category of the visited venue, and the functionality of the area. A good trip purpose inference method needs to consider all these factors and the intrinsic relationships among them. In the following sections, we will first describe the proposed method for POI popularity modeling with social media data, then demonstrate the proposed Dynamic Bayesian Network approach.

A. POI Popularity Modeling with Geo-tagged Tweets

A POI's popularity can be captured if we can accurately identify them from tweet messages. The basic idea is that a POI is more popular if it has been mentioned by more tweets. However, this is a very challenging task because social media data are short and noisy. In this section, we propose a method to learn POIs' popularities from geo-tagged tweets. It can be much easier to identify mentioned POIs from these tweets because their associated geo-coordinates give us a good hint.

In general, the proposed method works as follows. For each geo-tagged tweet, we first construct a local candidate pool with nearby POIs. Then a match index is calculated between the tweet content and each candidate POI names. Among the candidates, the POI with the largest match index above the threshold is marked as matched. Finally,

the POI popularity of a trip end area can be derived by aggregating all the identified POIs from geo-tagged tweets. In the following, we describe each component in detail.

POI Local Candidate Pool Construction. In order to identify POI mentions for a geo-tagged tweet, we first construct a local candidate pool with all the POIs near the geo-coordinates of the tweet. The range limitation should be set considering the accuracy of GPS devices. In this work, we set the range as 200 meters which usually results in candidate pools with dozens of POIs.

Calculate POI Match Index. After constructed a local candidate pool for each tweet, the next step is to find the best matched POI among all candidates. To this end, we design a match index to measure the similarity between a tweet and a POI name. This match index considers two essential factors:

- The number of matched terms. The match index should be larger if there are more terms matched between a tweet and a POI name. For example, a tweet "Was just told by a teenager working at this Jamba Juice, that I looked like a young Walter White" mentions the POI "Jamba Juice Redwood City", and two terms are matched between them. However, there is another nearby POI "Geoff White Photographers" which matches a term "white" to the tweet. In this case, "Jamba Juice" with 2 matched terms should be weighed higher than the other one with only 1 matched term.
- The rareness of the matched terms. Some terms may frequently appear in the candidate pool. For example, there is no surprise that many POI names contain "San Francisco" in the Bay area. Then these terms should have less impact on the matching index. On the contrary, terms such as "Corneta" is relatively rare. In fact, this term only appears in a restaurant named "La Corneta Taqueria". No doubt that these terms should have larger impact on the matching index. In other words, if terms like "Corneta" matched between a tweet and a POI name, we should have a high belief that the tweet mentioned the restaurant "La Corneta Taqueria".

In this work, we propose a POI Match Index which characterizes the aforementioned factors. Specifically, each Tweet T_i is represented by a set of terms, i.e., $T_i = \{u_1, u_2, \cdots, u_m\}$. Similarly, each candidate POI's name is represented by $P_j = \{v_1, v_2, \cdots, v_n\}$. The set of Matched Terms MT between T_i and P_j are

$$MT(T_i, P_j) = T_i \cap P_j. \tag{1}$$

Then we can calculate the Match Index as follows

$$
\begin{aligned}
MI(T_i, P_j) &= |MT(T_i, P_j)| \\
&\times \log \frac{N_{pool}}{1 + \sum_{k=1}^{N} \mathbb{1}(MT(T_i, P_j) \in P_k)},
\end{aligned}
\tag{2}
$$

where N_{pool} denotes the size of the candidate pool. $\mathbb{1}(\cdot)$ is the indicator function which returns 1 if and only if the condition holds. In addition, $\sum_{k=1}^{N} \mathbb{1}(MT(T_i, P_j) \in P_k)$ calculates the frequency of the matched terms MT in the

Table II: POI Category

Category	Google Place Type
Money	accounting, atm, bank, post office, finance
Leisure	art gallery, gym, movie rental, movie theater, museum, etc.
Food	bakery, cafe, food, restaurant, meal delivery
Bar	bar, night club
Care	beauty salon, hair care, spa
Store	book store, clothing store, grocery, supermarket, etc.
Trans	bus station, subway station, train station, taxi stand, etc.
Auto	car repair, car wash, gas station
Religion	cemetery, church, funeral home, mosque, etc.
Civic	courthouse, lawyer, police, fire station, city hall, etc.
Health	dentist, doctor, health, hospital, pharmacy, etc.
Improve	electrician, locksmith, painter, real estate agency, etc.
Edu	library, school, university
Lodge	rv park, lodging, campground

POI candidate pool. As shown in the Equation (2), the first term considers the number of matched terms, and the second term considers the rareness of the matched terms. Note that $|MT(T_i, P_j)|$ could be zero, i.e., there are no matched terms between T_i and P_j. In this case, the Match Term Index will be 0 which is reasonable.

After calculated the Match Index between T_i and every P_j in its candidate pool, we can return the one with the highest index as the identified POI. However, we still need to set a threshold to the MI, and return non-identified if none of MIs exceed the threshold.

POI Popularity Modeling. After extracted the mentioned POIs from social media data, we can further represent the POI popularity across different categories by counting the corresponding mentions from social media. For example, if restaurants are mentioned by 10 different tweets, we will count 10 towards the popularity of the POI category "Food". Then the counts can be normalized into a distribution across all the POI categories in Table II.

B. Dynamic Bayesian Network Construction

We propose a Dynamic Bayesian Network (***DBN***) to model people's sequential activities. As shown in the Figure 2, $a \in A$ denotes the activity performed (or trip purpose), $c \in C$ denotes the category of POI a user visited, and l is the trip end locations. All activities (A) and POI categories (C) defined in this paper are shown in Table I and Table II.

The ***DBN*** model can be interpreted in a generative process. For each trip i, a user first decides an activity a_i (or purpose) based on his previous one a_{i-1}. Then he chooses a venue category c_i based on this choice of activity. At last, he chooses a geo-location l_i to finally perform the activity a_i in venue c_i. This process continues until the last trip.

The likelihood function of the proposed ***DBN*** model is as follows

$$
\begin{aligned}
P(a, c, l) &= P(a_0)P(c_0|a_0)P(l_0|c_0) \\
&\cdot \left(\prod_{i=1}^{N} P(a_i|a_{i-1})P(c_i|a_i)P(l_i|c_i) \right),
\end{aligned}
\tag{3}
$$

Figure 2: The Dynamic Bayesian Network

(a) Trip ends   (b) Geo-tagged Tweets

Figure 3: Trip ends and Geo-tagged tweets
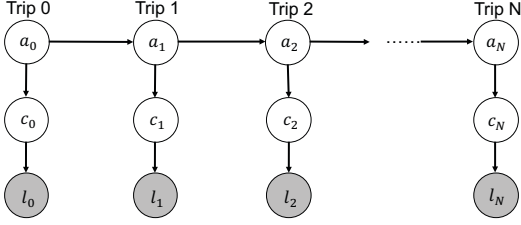
where $P(a_i|a_{i-1})$ is the probability of the activity a_i given previous activity a_{i-1}, $P(c_i|a_i)$ is the probability of the visited POI category given current activity a_i, and $P(l_i|c_i)$ is the probability of chosen location l_i given currently chosen POI category c_i. In the proposed model, the visited location l_i is always observed from the trajectory data, such that we can use Bayes's rule to approximate $P(l_i|c_i)$ as follows:

$$
\begin{aligned}
P(l_i|c_i) &\propto \frac{P(c_i|l_i)}{P(c_i)} \\
&\propto \frac{P_{POI}(c_i|l_i)P_{tweet}(c_i|l_i)}{\sum_i P_{POI}(c_i|l_i)P_{POI}(c_i|l_i)} \times \frac{1}{P(c_i)}. \quad (4)
\end{aligned}
$$

In the Equation (4), $P(c_i|l_i)$ denotes the POI category distribution given a geo-location l_i. This distribution is determined by two aforementioned factors: the functionality distribution $P_{POI}(c_i|l_i)$, and the popularity distribution $P_{tweet}(c_i|l_i)$. The first distribution is obtained from populating the nearby POIs, and the second distribution is obtained by extracting POI mentions from nearby tweets.

C. DBN Parameter Learning

There are two sets of parameters in the **DBN** model: the transition probabilities $P(a_i|a_{i-1})$ and the emission probabilities $P(c_i|a_i)$. Note that in our problem, the activities a_i and visited venues c_i are not fully observed. In other words, many activities and corresponding venues are not labelled in the data. In order to learn the parameters, we adopt the EM algorithm [6] designed for incomplete data.

D. DBN Prediction

With the learned parameters of the **DBN** model, we can infer possible activities and their corresponding probabilities for any given trip. Specifically, we can calculate the posterior probability of the j_{th} activities given a user's trajectory $d \in \mathcal{D}$, where $\mathcal{D}$ denotes all users' trajectories. This estimates the probability of activity a_j out of all possible activities set A, as shown in Equation (5).

$$
P(a_j|d) = \frac{P(a_j, d)}{P(d)}, \forall a_j \in A, \quad (5)
$$

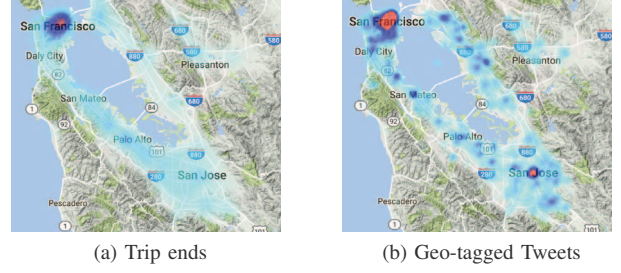The returned results are possible activities ranked by their probabilities.

IV. EXPERIMENTS

The proposed method is evaluated with real-world data sets including human trajectories, point of interests, and tweets in the Bay area, CA. In the following sections, we discuss the data sets, the baselines, and the evaluation results.

A. Data Sets and Baselines

Trajectories. The California Household Travel Survey (CHTS) [5] collected travel information from residents across California's 58 counties. The survey was designed to obtain detailed information about the household socio-economic characteristics and their travel behaviors. Among its various achievements, the survey collected $8,631$ participants' GPS trajectories for a week. In addition, each trajectory is accompanied with a detailed trip diary which records visited POIs and trip purposes. The trip purposes are labelled by users with pre-defined categories, and some of them are shown on the second column in Table I. However, the diaries are far from exhaustive and many trips recorded by GPS devices are not logged in the diary. This is a common issue for traditional surveys because it requires too much efforts from the participants. The heat map of trip end locations are shown in Figure 3 (a).

Twitter Data. We collected 6.9 million geo-tagged tweets in the Bay area from Jan 31 2013 to Feb 16 2017. They are queried through Twitter APIs and filtered by geo-coordinates. The heat map of geo-tagged tweets are shown in Figure 3 (b).

Point of Interest. The POIs are queried through Google Places API [7]. In this work, we use its Nearby Search request to get the nearby POIs of any given geo-coordinates, and the returned POI names and types are utilized by the proposed method. Some Google Place types are shown in the Table II.

In the following experiment, we perform extensive experiments on the proposed **DBN** method, and compare it with the state-of-the-art methods for the trip purpose inference, i.e., Random Forest (**RF**), Support Vector Machine (**SVM**), Artificial Neural Network (**ANN**), and K-nearest Neighbor (**KNN**).

B. Trip Purpose Inference

To compare the performance of all the methods on the trip purpose inference, we conduct experiments on the collected

Table III: Performance of Trip Purpose Inference

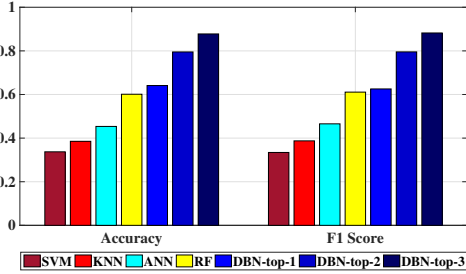| Accuracy	SVM	ANN	KNN	RF	DBN-top-1	DBN-top-2	DBN-top-3
Education	27.0%	41.9%	34.4%	40.2%	**52.3%**	**72.9%**	**78.4%**
Shopping	59.6%	65.3%	52.5%	78.6%	**80.1%**	**94.2%**	**98.4%**
EatOut	4.8%	30.0%	32.1%	60.6%	**79.0%**	**83.2%**	**85.5%**
Recreation	18.6%	39.0%	30.6%	55.5%	**62.7%**	**77.4%**	**84.7%**
Personal	7.7%	21.7%	22.5%	**50.4%**	42.2%	**65.0%**	**90.0%**
Transportation	**84.6%**	74.2%	59.1%	75.4%	68.3%	**84.3%**	**89.6%**
Average	33.7%	45.4%	38.5%	60.1%	**64.1%**	**79.5%**	**87.8%**

Figure 4: Average Performance of Trip Purpose Inference

real-world data set. The features used for training include travel mode, previous activity category, activity time and duration, nearby POI category distribution, and nearby POI popularity distribution. We randomly select 80% trips as training data, and leave the rest for testing. All the methods are evaluated by 10 times, and the average results are reported. In addition, there are about one third trip end locations in the dataset are either home or work, because the survey collected trips from people's daily lives. Using these trips to train the model would greatly bias the performance. As a result, we only test the inference results on other activities, and assume the home and work trips are known. In fact, it is also easy to infer users' home and work locations from their trajectories, because normal people stay in these locations for the most time. In other words, we are interested in non-trivial tasks of inferring non-home and non-work trip purposes.

Note that the proposed *DBN* model can output a ranked list of activities with corresponding confidence. This is a great advantage by adopting the Bayesian method. As we have discussed in Section III-D the top ranked inference results are very useful in real-world applications, especially when the classification boundaries among classes are vague. As a result, we also evaluate the *DBN* performance with top-2 and top-3 inference results in the experiment, denoted as DBN-top-2, and DBN-top-3 respectively. In these cases, we regard the inference as correct if the ground truth activity is among the top-2 or top-3 results.

The inference accuracy and F1 scores are shown in Table III, and the average results are also compared in Figure 4. We can observe that the proposed *DBN* model, including DBN-top-1, DBN-top-2 and DBN-top-3, outperforms other baselines on almost every activity category by higher accu-

racy and F1 score. This is because *DBN* model captures the intrinsic relationships among sequential activities, trip end locations' POI distributions and the popularities identified from the Twitter data. On average, the *DBN* model can reach 64% accuracy with the top-1 inference result. Moreover, the average accuracy of top-2 ranked results can reach 79.5% which is impressive considering random guess would only result in 33% accuracy. This demonstrates the top-ranked results generated by the *DBN* model are very useful in the trip purpose inference.

C. POI Mention Extraction from Geo-tagged Tweets

In Section III-A, we propose to extract mentioned POIs from geo-tagged tweets. For each geo-tagged tweet, we can accurately identify the mentioned POI with a local candidate pool and the match index. In practice, it is very hard to evaluate the performance of POI mention extraction from tweets, because we have so many nearby tweets and trips in the data set. Figure 5 shows the histogram of tweets near
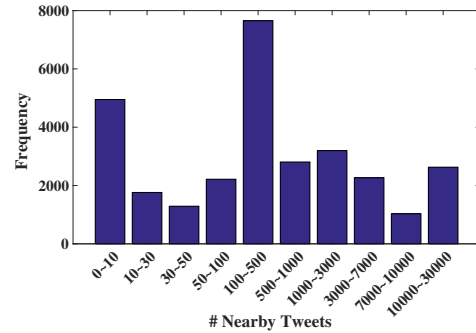

Figure 5: Histogram of tweets near trip ends

trip ends. On average, there are 2607 geo-tagged tweets near each trip end location (within 200 meters). Actually, it is impossible to be evaluated without a standard data set labelled by human workers. To this end, we recruited volunteers to label the POI mentioned tweets near 50 random trip end locations. Given a list of tweets, the volunteers are asked to judge whether those tweets mentioned any nearby POI, and whether the identified POIs are correct. In Table IV, we present the results on several example trip end locations and the average performance. For each location, we populate the total number of nearby tweets, the number of POI-mentioned tweets which are identified by human workers. Some of the tweets are formatted by third party

Table IV: POI Extraction from geo-tagged Tweets

Example Locations	Total Nearby Tweets	POI-mentioned Tweets	App-formatted Tweets	POI-identified Tweets
37.4028125, -121.881519784	575	184	116 (63.0%)	159 (86.4%)
37.3896875, -122.03080036	294	103	80 (77.7%)	93 (90.3%)
37.3278125, -121.8842176	1188	368	247 (67.1%)	305 (82.9%)
37.6715625, -122.472796763	1232	596	446 (74.8%)	495 (83.1%)
Average Recall	-	-	68.8%	83.5%

Apps, such as Foursquare and Instagram. We can easily parse POIs from these well formatted tweets, for example, "Im at Applewood Pizza in San Carlos Ca" and "Bagel time! @ Bagel Street Cafe Town Center Alameda". However, there are still many POI-mentioned tweets which are not well formatted. For instance, "Catch us at amc Mercado 20 12 am insurgent!!!!". The better performance shown in Table IV indicates the proposed method can extract mentioned POIs from tweets, no matter they are well formatted or not.

V. RELATED WORK

Trip Purpose Inference The ubiquitous adoption of GPS-integrated devices has enabled extensive studies on the trip purpose inference. In [8], they proposed to annotate geo-tagged tweets with POI categories. Detailed feature extraction methods are introduced and traditional classification methods are utilized for the annotation task. In [4], they use taxi trajectories and POIs to capture the human mobility patterns in an urban area. In addition, traditional classification methods were widely utilized to infer the trip purposes, such as random forest [2], artificial neural network [9], decision tree [10], SVM [1] and etc. Compared with the aforementioned work, the model proposed in this work incorporates heterogeneous data sets, including trajectories, POIs and social media messages. In addition, the proposed method captures the sequential properties of individual's trip activities which results in better inference results.

POI Identification from Social Media Data. In order to extract fine-grained location information from tweets, Li et al. [11] proposed a Conditional Random Field model to identify POI mentions from social media messages, and they further proposed a method [12] to link the POI name with Foursquare inventory. In addition, a supervised Bayesian Model [13] is proposed to annotate POIs with tweet information. The above research aim at identifying POIs from general social media messages, especially for tweets without geo-tags. However, in this work, we propose to extract POI mentions from geo-tagged tweets. The proposed method with local POI candidate pools can solve this problem effectively and efficiently.

VI. CONCLUSIONS

In this paper, we propose to infer people's trip purposes with heterogeneous data sets. The proposed Dynamic Bayesian Network model can capture the intrinsic relationships among sequential activities. In addition, we also propose to incorporate POIs' popularity information near trip end locations because it gives us good hints about people's activities. Extensive experiments were conducted on real-world data sets, and the results demonstrate advantages of the proposed method on accurately inferring the trip purposes.

VII. ACKNOWLEDGEMENTS

REFERENCES

[1] Z. Zhu, U. Blanke, and G. Tröster, "Inferring travel purpose from crowd-augmented human mobility data," in *Urb-IoT'14*, 2014, pp. 44–49.

[2] A. Ermagun, Y. Fan, J. Wolfson, G. Adomavicius, and K. Das, "Real-time trip purpose prediction using online location-based search and discovery services," *TRC*, vol. 77, pp. 96–112, 2017.

[3] B. Furletti, P. Cintia, C. Renso, and L. Spinsanti, "Inferring human activities from gps tracks," in *KDD'13*. ACM, 2013, p. 5.

[4] P. Wang, Y. Fu, G. Liu, W. Hu, and C. Aggarwal, "Human mobility synchronization and trip purpose detection with mixture of hawkes processes," in *KDD'17*. ACM, 2017, pp. 495–503.

[5] "California household travel survey." [Online]. Available: http://www.dot.ca.gov/hq/tpp/offices/omsp/statewide_travel_analysis/chts.html

[6] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[7] "Google places api." [Online]. Available: https://developers.google.com/places

[8] D. Falcone, C. Mascolo, C. Comito, D. Talia, and J. Crowcroft, "What is this place? inferring place categories through user patterns identification in geo-tagged tweets," in *MobiCASE'14*. IEEE, 2014, pp. 10–19.

[9] G. Xiao, Z. Juan, and C. Zhang, "Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization," *TRC*, vol. 71, pp. 447–463, 2016.

[10] M. Oliveira, P. Vovsha, J. Wolf, and M. Mitchell, "Evaluation of two methods for identifying trip purpose in gps-based household travel surveys," *TRB*, no. 2405, pp. 33–41, 2014.

[11] C. Li and A. Sun, "Fine-grained location extraction from tweets with temporal awareness," in *SIGIR'14*. ACM, 2014, pp. 43–52.

[12] Z. Ji, A. Sun, G. Cong, and J. Han, "Joint recognition and linking of fine-grained locations from tweets," in *WWW'16*. International World Wide Web Conferences Steering Committee, 2016, pp. 1271–1281.

[13] K. Zhao, G. Cong, and A. Sun, "Annotating points of interest with geo-tagged tweets," in *CIKM'16*. ACM, 2016, pp. 417–426.