Transportation Research Part C 97 (2018) 159–174

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Forecasting current and next trip purpose with social media data and Google Places

Yu Cui[a], Chuishi Meng[b], Qing He[a,c,*], Jing Gao[b]

[a] Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA
[b] Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA
[c] Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA

ARTICLE INFO

Keywords:
Bayesian neural network
Google Places
Social media
Trip purpose prediction

ABSTRACT

Trip purpose is crucial to travel behavior modeling and travel demand estimation for transportation planning and investment decisions. However, the spatial-temporal complexity of human activities makes the prediction of trip purpose a challenging problem. This research, an extension of work by Ermagun et al. (2017) and Meng et al. (2017), addresses the problem of predicting both current and next trip purposes with both Google Places and social media data. First, this paper implements a new approach to match points of interest (POIs) from the Google Places API with historical Twitter data. Therefore, the popularity of each POI can be obtained. Additionally, a Bayesian neural network (BNN) is employed to model the trip dependence on each individual's daily trip chain and infer the trip purpose. Compared with traditional models, it is found that Google Places and Twitter information can greatly improve the overall accuracy of prediction for certain activities, including "EatOut", "Personal", "Recreation" and "Shopping", but not for "Education" and "Transportation". In addition, trip duration is found to be an important factor in inferring activity/trip purposes. Further, to address the computational challenge in the BNN, an elastic net is implemented for feature selection before the classification task. Our research can lead to three types of possible applications: activity-based travel demand modeling, survey labeling assistance, and online recommendations.

1. Introduction

The key element of transportation systems is to estimate the travel demand of moving people or goods. Travel demand models aim to forecast traffic flows in the transportation system. Therefore, the activities of individuals and households and their travel decision (Bhat and Lawton, 2000) should be considered. The trip propose is the origins of these needs and is crucial to understanding travel behavior and estimating travel demand for transportation planning and investment decisions. Therefore, forecasting trip purposes is important for travel demand analysis. Moreover, activity-based models become increasingly prevalent. These models generate activities first, define the destination for the activities, determine the travel model, and finally predict route choices (Castiglione et al., 2015). Therefore, activity information is important to the activity-based model, and the trip purpose or activity purpose is an essential part of activity information.

Travel behaviors and travel patterns have become increasingly temporally and spatially diverse and have varied in recent

* Corresponding author at: Department of Industrial and Systems Engineering and Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA.
E-mail address: qinghe@buffalo.edu (Q. He).

https://doi.org/10.1016/j.trc.2018.10.017
Received 8 April 2018; Received in revised form 22 September 2018; Accepted 21 October 2018

decades. This phenomenon is a synthetic consequence of a range of factors, for instance, increasing household car ownership, the growing number of double-income families, and the rising diversity of working time and place with the appearance of part-time work and work from home (Bohte and Maat, 2008). To capture this variation, a collection of reliable data is essential to acquiring detailed travel information. A household travel survey is a widely used and traditional method. Conventional household travel surveys are composed of 1-day or multiple-day travel diaries of sampled households. Households are selected according to the proportion of certain groups of the same sociodemographic population. This subsample reflects similar travel behaviors of the entire population. Although household travel surveys are only a subsample of a population, this method is both high-expenditure and time-consuming due to the large population base. Based on traditional survey data, past studies predict trip purposes through several methods, mainly using land use, temporal information, and sociodemographics. However, major challenges still exist. The existing methods lack detailed activity-related data, including nearby points of interest (POIs) and historical choices from other travelers.

Recently, ubiquitous various sensors can capture an enormous amount of passive temporal-spatial data. Each sensor has its own advantages and disadvantages. Compared to call detail records (CDR) (Cici et al., 2014; Ponieman et al., 2013), Wi-Fi, and RFID (Schneider et al., 2013), GPS is a better way to collect travel survey data because GPS devices can generate data every second with relatively high location accuracy. In this way, more reliable data can be collected, and less burden is imposed on participants (Liu et al., 2017).

Social media is also an emerging tool that allows people to create, share or exchange information and ideas by using texts, images or videos in a virtual community platform. Using social media to keep in touch with friends is a major method of communication among modern people. Because cell phones have developed to allow easy access to social media, this enables users to generate spatial-temporal information immediately. Social media has become increasingly prevalent and ubiquitous. Not only are young people willing to share their moments and moods with their friends through social media, but elderly people also use it to record their lives and reconnect with friends (Anderson and Perrin, 2017). Merchants (e.g., Yelp) offer varying degrees of discounts or gifts when individuals check-in or post messages related to the business, which makes users willing to provide location-related information on social media. Individuals are also enthusiastic about checking-in on smartphone apps (Foursquare Swarm), which collects diverse medals or rewards and allows users to compete with their friends.

In addition to social media, POIs are favorable for extracting land use information. They can be obtained from online location-based search and discovery services. For instance, the Google Places application programming interface (API) (Ermagun et al., 2017), searches POIs according to input place names or locations and a search radius. The API returns detailed information about that place (category, location, opening hours and so on) or all detailed information about POIs within the search area of that location. With this method, we can not only query POIs in real-time and identify trip purpose automatically, but we can also reduce the burden of individuals filling out traditional household travel surveys. In this paper, we employ the function "nearby search" in the Google Place API to extract POIs. "Nearby search" will return place names and place categories with respect to input coordinates and search radius. There are several characteristics of "nearby search". There is no radius limitation for search requests. However, Google Places can only return at most 60 nearby POIs for the given geo-coordinates of each query, and a free Google Place API account has a limit of 150,000 free requests per 24 h period.

In this paper, the "nearby search" function returns a summary list of places near the queried location within a specific area, with detailed information including the category, location, and business hours. However, this list only provides what kind of POIs surround the queried place. We do not know the exact place that individual visits or the probability of entering that POI. In addition to household travel survey data, participants report the place name, and social media data provide similar information. The name of places the user visited may appear in the text of their posts. The more people that post about a place, the higher the popularity of that place. While this trip information is valuable for social media data, it is difficult to retrieve. The is discussed more thoroughly in Section 2.2.

This research is a natural extension of Ermagun et al. (2017). We make detailed comparisons between this paper and Ermagun et al. (2017) in Section 2.1. During this research, we specifically address three major challenges. First, users' GPS records alone are not sufficient to infer their trip purposes. POIs near the trip end can reveal land use, which can be related to the trip purpose. However, provided with dozens of POIs, the POI information alone cannot capture the users' preferences for visiting the area. Therefore, we propose to mine social media data to help capture popularity for each POI. The second challenge is how to extract the POIs mentioned in Tweets. It is difficult to effectively extract the mentioned POIs from Tweets because entity extraction from short text is difficult and social media data are very noisy. To address these two challenges, this paper aims to predict the current trip purposes and the next trip purpose (e.g., education, work, shopping, recreation) of individuals given their trip trajectories (or end locations), nearby POIs, and social media data. The definitions of inference of trip purpose, prediction of current trip purpose and next trip purpose are shown in Fig. 1. Trip purpose inference infers what activity has been experienced when the activity ends, given that both the activity duration and location are known. This task was conducted by Meng et al. (2017). Current purpose prediction aims to predict what a person does when arriving at a place, and the activity has not started, given that the activity location is known but not the activity duration. With regard to next trip purpose prediction, the objective is to predict what to will be done next when the current activity ends, given that the activity duration is known but not the location information about the next activity. To fill existing gaps, this paper focuses on the last two tasks.

The third challenge is feature selection. The original dataset contains too many features, and one cannot simply throw all features into the model. Otherwise, the model will require a much longer training time and have lower accuracy, and more irrelevant and noisy features. To address this problem, we employ the elastic net for feature selection. The elastic net is a feature selection method that comprises both LASSO and ridge regression. Therefore, it can overcome the limitations of both methods. For detailed information about the elastic net, please refer to Zou et al.'s paper (Zou and Hastie, 2005).

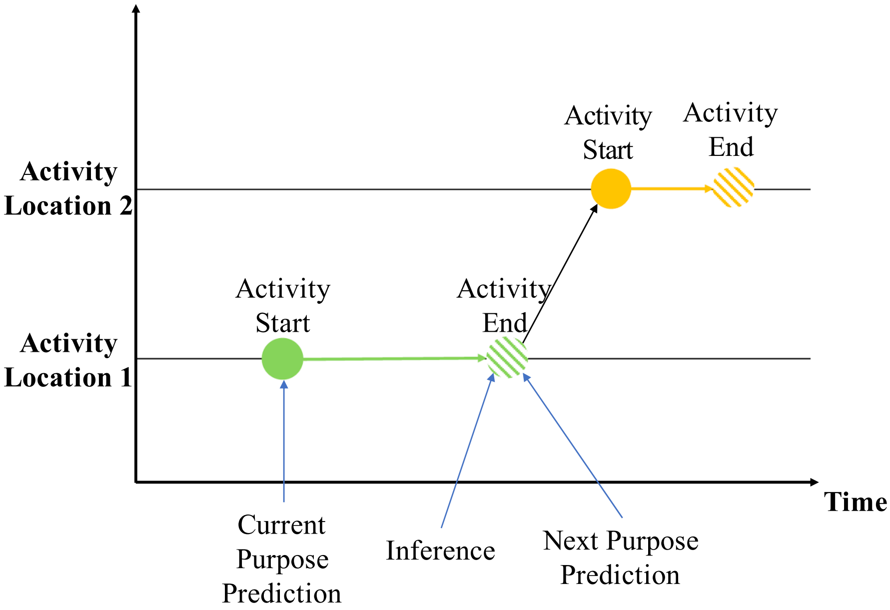

Fig. 1. Definition of inference and prediction in this paper.

The remainder of this paper is structured as follows: Section 2 summarizes previous studies for trip purpose inference and the usage of social media. Section 3 introduces the datasets and the initial analysis result of the datasets. Section 4 presents the methodology. Section 5 discusses the results of the models. Finally, this paper concludes in Section 6 by listing potential applications and Section 7 by summarizing the main conclusions and future research.

2. Literature review

2.1. Trip purpose inference and prediction

We can extract abundant information directly from GPS data, such as the trip start and end time and trip start and end location. This information can easily derive the origin-destination matrix, which is very important for traffic demand modeling. However, travel mode and trip purpose cannot be obtained directly from the original GPS data. Compared to mining trip purposes from GPS data, travel mode is more straightforward (Su et al., 2016). Trip purpose is understudied. Previous methods include rule-based methods (Wolf et al., 2001; Wolf et al., 2004), probabilistic methods (Chen et al., 2010; Oliveira et al., 2014), and machine learning and neural network methods (Ermagun et al., 2017; Liao et al., 2007; Xiao et al., 2016). Previous studies mainly used land use, temporal information, and sociodemographics obtained from household travel surveys. More literature can be found in Ermagun et al. (2017). Trip purpose can be separated into different categories. Using the three categories home, work/school, and other is the simplest and most common classification strategy. Alexander et al. divided trip purpose into home, work and other. Home is the maximum number of visit place from 7 pm to 8 am (Alexander et al., 2015). Work is the maximum distance (number of visits multiplied by distance for one place) from home, and other is otherwise. Increasing the number of categories provides more specific trip purposes. However, this also increases difficulties of inference and prediction. Despite different datasets utilized in different research, with the decision tree algorithm, Deng achieved 87.6% accuracy when differentiating between 7 trip purposes (Deng and Ji, 2010). Lu et al. obtained 73.4% accuracy when classifying 10 different trip purposes (Lu and Liu, 2012). For 12 distinguished trip purposes, the overall accuracy that Oliveira et al. reached was only 65% (Oliveira et al., 2014). Nevertheless, these accuracies are dominated by home or work trips, which account for approximately one-third of total trips (Ermagun et al., 2017). With the help of the Google Places API, which is a kind of online and real-time location-based query service, Ermagun et al. achieved 67.14% overall prediction accuracy by using the random forest algorithm when differentiating between 5 trip purposes other than home and work.

We further make a detailed comparison between this paper and Ermagun et al. (2017). Ermagun et al. (2017) considered the Google Places information, and the dataset had two components. (1) Their original survey had 115,821 trips reported by 30,286 individuals from 14,055 households in 26 counties around the Minneapolis-St. Paul metropolitan region. After data preprocessing, the final dataset contained a total of 58,503 trip purposes (including 12,217 home trips, 12,369 work trips, and 33,917 nonhome and nonwork trips). (2) They only obtained as many as 20 POIs around the trip destination by using the Google Place API. Compared to their research, the dataset in our paper consists of three components. (1) After examining the survey dataset, we finally screened 42,767 activities (including 13,198 home activities, 5425 work activities, and 25,144 activities) within the Bay area (9 counties in total). While the final dataset is comparable with Ermagun et al. (2017), our data records are denser because our research region is smaller. (2) In our paper, we acquired up to 60 records of POIs around the trip destination from the Google Place API. Therefore, we

Table 1
Comparison of relevant papers.

	Ermagun et al. (2017)	Meng et al. (2017)	This paper
Objective	Trip Purpose Inference	Trip Purpose Inference	Current Purpose Prediction and Next Purpose Prediction
Data	2010 TBI Google Places (maximum 20 results)	2010 CHTS Google Places (maximum 60 results) Twitter	2010 CHTS Google Places (maximum 60 results) Twitter
Methodology	Random Forest	Dynamic Bayesian Network	Bayesian Neural Network

can access more information about the trip destination. (3) This research also incorporates historical social media data as a new data source. In addition, this paper takes into account the prediction of the current and next trip purpose, which was not examined by Ermagun et al. (2017). Table 1 also shows the comparison of (Ermagun et al., 2017; Meng et al., 2017) and this paper.

2.2. Social media analytics and applications in transportation

Since widely utilization of social media and passion of individuals generates a huge amount of passive data, social media data attracts attention in both academic and industrial area (Lv et al., 2017). Yates et al. explored how social media technologies can influence emergency management on aspects of knowledge sharing, reuse, and decision-making in an effective and efficient way (Yates and Paquette, 2011). Not only individuals share thoughts through social media, companies (Culnan et al., 2010), merchants, agencies (Wang, 2014; Zheng et al., 2016), and even politicians (Broersma and Graham, 2012) also utilize social media to carry their points. They broadcast information and influence individuals' opinions. Moreover, social media information can also be retrieved and benefit lots of fields including transportation.

Previous social media studies in the transportation area mainly fall into two applications: traffic incident detection (Wanichayapong et al., 2011; Mai and Hranac, 2013), and traffic prediction(Gal-Tzur et al., 2014; Chen et al., 2016; Daly et al., 2013; Lv et al., 2015). Potential of social media in transportation have been exploring for several years. Several studies identified traffic accidents from tweets (Zheng et al., 2016). For example, Zhang et al. employed deep learning to detect traffic accidents from social media data (Zhang et al., 2018). The researchers considered that social media is an abundant, cost-effective, and real-time data source which can complement existing accident data source (Grosenick, 2012; Schulz et al., 2013; Gu et al., 2016). Others extracted traffic conditions (e.g. congestions, crash) from social media data which are helpful in traffic management and improve the level of service of traffic. Ni et al. utilized social media data to predict traffic flow and subway passenger flow under event conditions (Ni et al., 2014, 2017). Lin et al. modeled the impacts of inclement weather on freeway traffic speed by using social media data (Lin et al., 2015). Zhang et al. explored the correlation between twitter concentration and traffic surge (Zhang et al., 2016). Social Media can also help in understanding individuals travel behaviors. Further, Zhang et al. examined the potentials of using social media to infer the longitudinal travel behavior by using a sequential model-based clustering method (Zhang et al., 2017). Saeedimoghaddam et al. employed geo-tagged social media data to understand individual travel behavior (Saeedimoghaddam and Kim, 2017).

There are major challenges to be addressed before the use of tweets in extracting useful trip information. First, the tweet data is mainly comprised of the inherently complex and unstructured word texts, and the language ambiguity (Chen et al., 2014) in the tweet contents make them difficult to interpret. What's more, as the context of a tweet is limited to 140 words and tweet users usually intend to be concise, the common methods in the language processing studies such as support vector machine (D'Andrea et al., 2015; Schulz et al., 2013), natural language processing (Wanichayapong et al., 2011; Li et al., 2012) maybe not be adaptive. Second, the information of activity location is embedded in the names that are not straightforward to be interpreted. For example, one cannot identify "paint and pour" as the "private school" until one searches it online. Third, same as other passively collected data, social media data generally lack ground truth of individuals' travel modes and trip purpose information. Inferring the ground truth of a personal trip is very difficult due to personal privacy. To address the above challenges, this paper employs Google Place API to query trip end's categories and complement social media information.

2.3. Bayesian neural network (BNN)

Deep learning is widely used and is state-of-the-art developments in all manner of data fields, and it can deal with massive datasets (Schmidhuber, 2015; Gers et al., 2002; He et al., 2016; Dong et al., 2016; Cui et al., 2018). However, there are still some drawbacks to standard deep learning. First, neural networks compute point estimates for their weights, so that they make point predictions as well. Therefore, they may make overly confident about some classes (Blundell et al., 2015). Second, a deep neural network has huge amount number of parameters. Therefore, it needs very large datasets in order to avoid overfitting. Third, in addition to the parameters the deep neural networks need to infer, there also have lots of configuration that need to be set, such as dropout probability, learning rate, network structure, and so on. Finally, Deep neural networks are poor at representing uncertainty. There usually exist two types of uncertainties in a model. One is the aleatoric uncertainty which captures noise inherent in the observations. The other one is the epistemic uncertainty which captures our ignorance about which model generates our collected data (Kendall and Gal, 2017).

Bayesian method is a way of updating probabilities or our beliefs about hypotheses given data. A probability distribution is the best way to represent the uncertainty of our knowledge and hypotheses. A Bayesian neural network is a neural network with a prior distribution on the weights. The history of Bayesian neural network can origin back to 30 years before. Denker et al. published a paper and hinted that integrating Bayesian over network parameters (Denker et al., 1987). Then Bayesian neural networks entered its golden era around 1992. A series of papers that wrote by MacKay which described a quantitative and practical Bayesian framework for backpropagation networks (MacKay, 1992). Very few studies of Bayesian neural network can be found in the transportation field, Xie et al. employed Bayesian neural networks in transportation safety studies (Xie et al., 2007).

3. Data description and preprocessing

3.1. Survey data

This paper collected California Household Travel Survey (CHTS) data from February 2012 to January 2013 (Laboratory, 2017). The geographical range of this survey covers all 58 counties of California and three adjacent counties in Nevada. In this paper, we only utilize data from the Bay Area. Two types of data, GPS data and survey data, are contained in the CHTS dataset. In the Bay Area, 108,778 individuals belonged to 42,431 households in this survey and completed a one-day survey. For GPS data, 10,474 travelers from 5460 households carried GPS devices and reported 7 days of GPS data. Three types of GPS devices were implemented in this survey: wearable GPS devices, in-vehicle GPS devices, and in-vehicle GPS devices plus an onboard diagnostic (OBD) unit. Each participant completed a trip diary on a website or mailed a paper version of the survey to the institute. The GPS data records included trip-related information, such as coordinates of the trip start and end locations, trip start and end times, trip modes, trip durations, and trip distances. The survey data included activity-related information, such as activity place names, activity purposes, and activity start and end times. To merge these two datasets together, several rules were created.

1. Survey activity data for each individual should contain at least two trips. Because the first activity for all participants is to stay at home, they should have at least one activity other than home to generate a trip. Moreover, the last activity for all participants should also be ‘home’.
2. At least one origin or destination of the trip in CHTS data should be located within the Bay Area bounding box to coordinate with tweet data.
3. POI queried from Google Places API should be assigned to at least one POI category (money, leisure, food, bar, care, store, trans, auto, religion, civic, health, improve, edu, or lodge) because we need this POI to categorize the trip ends.

There were also several challenges in the data prepossessing procedure. First, participants provided inaccurate or false travel information (e.g., rounded up arrival and departure times from activity locations, missing trips, or wrong trip purpose). Therefore, information in the survey data may not match the GPS data. To address this issue, we first ordered one traveler’s survey data according to arrival time, activity and visit place sequence and the ordered the same traveler’s GPS data according to the trip end time, trip sequence and trip duration. Then, we merged the two datasets together. Second, one GPS trip may be divided into several records in the GPS data when vehicles pass through tunnels or encounter other circumstances when the GPS signal is blocked. Moreover, several GPS trips are connected to one trip if the interval between each trip is too small, for instance, when one picks up or drops off passengers. To address this problem, we carefully examined both the GPS data and the survey data and merged or divided GPS trips according to the survey data. After overcoming these challenges, we obtained a total of 43,767 activities, including 13,198 home activities and 30,569 nonhome activities. Fig. 2 depicts the distribution of activities conducted by individuals. Moreover, the average number of activities was 7.65 per day. According to the first data filter rule mentioned above, regardless of the first and the last trip occurring at home, we still required at least one activity other than home for all participants. Therefore, all the individuals in the database have at least three activities after filtering.

There were 40 choices when participants fill out the survey. Excluding “other”, “loop trip”, and “don’t know and refused”, there were 37 activity categories. In Ermagun et al. (2017), trip purposes were aggregated into 7 categories, including home, work, education, eat, shop, social activity, and personal business. In addition to these categories, this paper also added an activity of transportation to better coordinate with the survey data. The detailed rules of categorizing the trip purpose are shown as follows:

- Home: Trips for any activities performed at home.
- EatOut: Drive-through meals, and eating at restaurants.
- Personal: Personal business, civic and religious activities.
- Recreation: Entertainment, indoor or outdoor exercises, sports, and visiting friends.
- Education: Trips for any activities performed at school or school-related trips.
- Shopping: Routine shopping, shopping for major purchases or specialty items, and purchasing services.
- Transportation: Changes in travel modes and picking up or dropping off passengers.
- Work: Trips for any activities at the workplace, or work-related trips.

To avoid the existence of repeated measures, we also conducted repeated measures tests using SPSS. The test results show that the variables measured repeatedly are statistically significantly different.

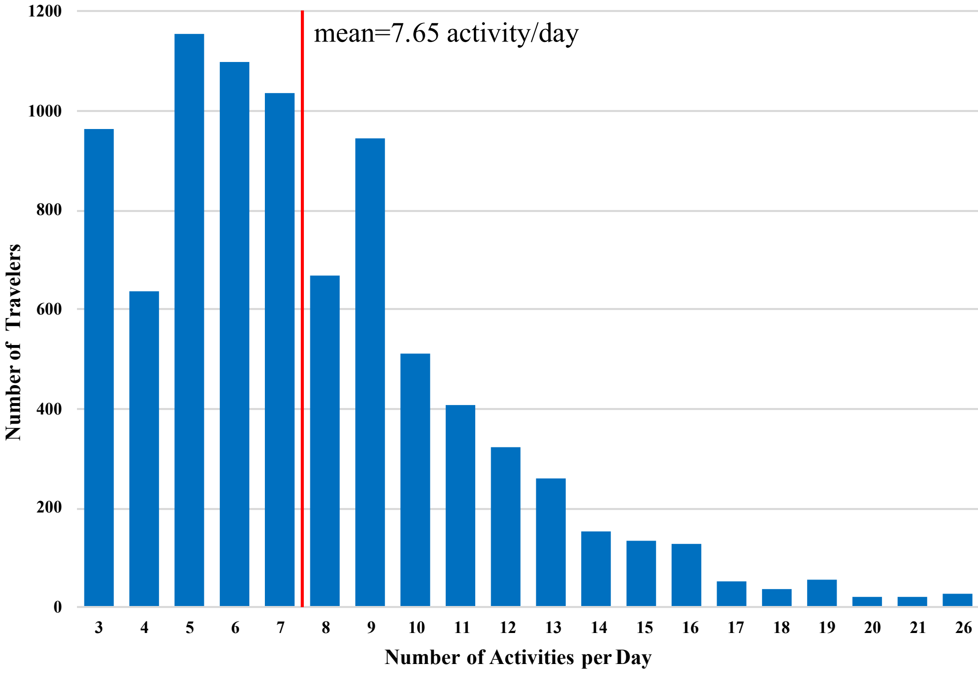

Fig. 2. Distribution of the number of activities. Note that we have deleted the days with only one "Home" activity. Thus, all individuals in the database have at least three activities after filtering including two "Home" and one other activity.

3.2. Information retrieval with Twitter data

Twitter data were collected from January 31, 2013, to February 16, 2017, in the Bay Area with a bounding box (longitude: 121.75W–122.75W, latitude: 36.8N–37.8N). There were two types of tweet data: geo-tagged tweets and non-geo-tagged tweets. In this paper, we only used the geo-tagged tweet data, and there were approximately 6.9 million valid geo-tagged tweets in total.

Since social media data are very noisy, it is very difficult to extract the named entity from a short text. In this paper, we implemented an information retrieval procedure to recognize the mentioned POIs from nearby geo-tagged tweets in (Meng et al., 2017). This method can reach up to 90% accuracy, whereas the Foursquare tweet-based method only has 2–16% accuracy.

4. Methodology

Before discussing the Bayesian neural network, we introduce the neural network. A neural network can be represented as a weighted directed graph in which nodes and directed edges with weights are connections between neuron outputs and neuron inputs. The weighted inputs are summed inside neurons and passed through an activation function. The activation function is a set of transfer functions used to scale the summations to the desired value. A neural network contains three kinds of layers: an input layer, an output layer, and hidden layers. In this paper, we build a neural network with an input layer, an output layer, and two hidden layers, which is also called a multilayer perceptron, and the structure is shown in Fig. 3. The activation function between the input layer and the hidden layer and between the two hidden layers were tanh functions (Abramowitz and Stegun, 1964). In addition, between the second hidden layer and the output layer, the activity function was a sigmoid function (Mitchell, 1997). There were also 10 neurons in each hidden layer.

Estimating the parameters for a neural network consists of two parts. The first part is a forward propagation. Afterward, the second part is backpropagation and weight updating. These two steps repeat iteratively to minimize the loss function. Given data $\mathscr{D} = \{X_i, y_i\}_{i=1}^{N}$ with input data points $X_i = \{x_i^k\}_{k=1}^{V_o}$, where V_k is the number of the types of observed random variables in the model, and $y_i \in \mathbb{R}$. The estimated parameter θ is represented by the function as follows:

$$\begin{aligned}\theta^* &= \underset{\theta}{\operatorname{argmin}} L(\theta) + \lambda \bullet \Phi(\theta)\\ &= \underset{\theta}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} L(y_i, y_i^{'}) + \lambda \bullet \Phi(\theta)\\ &= \underset{\theta}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(X_i, \theta)) + \lambda \bullet \Phi(\theta)\end{aligned} \tag{1}$$

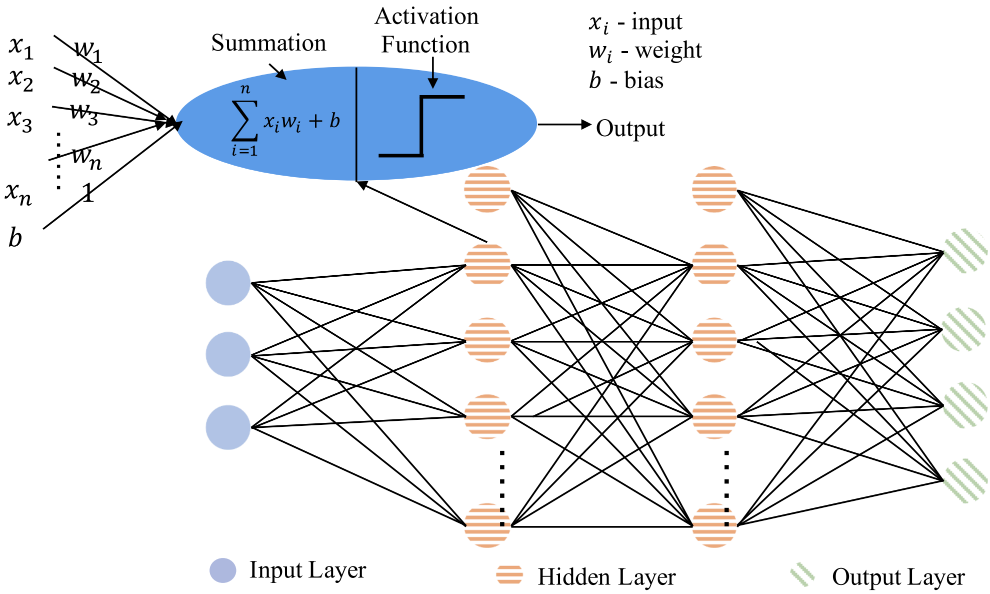

Fig. 3. Structure of a 2-layer neural network.

where $\Phi(\theta)$ is the regularization term or penalty term, $f(\bullet)$ is the activation function, and $L(\theta)$ is the loss function. There are many types of loss functions, for example, mean squared error (MSE), cross entropy, and hinge (Janocha and Czarnecki, 2017).

A traditional neural network uses point estimation for parameters that are the weights of the neural network. However, a Bayesian neural network is built upon the Bayesian inference on the weight of the neural network and treats these parameters as probability distributions over the possible value. The illustration of parameters in a traditional neural network and a Bayesian neural network is shown in Fig. 4 (Blundell et al., 2015). The goal of the Bayesian neural network is to maximize the evidence lower bound of the network (ELBO) (Yang, 2017), which is denoted below:

$$\mathscr{L}(\gamma, \nu) = c_0 \mathbb{E}_{q(\Theta)}\left[\sum_{i=1}^{N} \mathbb{E}_{q(z_i)}[\log p(y_i|Z_i, \Theta)]\right] - c_g KL[q(\Theta)||p(\Theta)] - c_l \sum_{i=1}^{N} KL[q(Z_i)||p(Z_i)] \tag{2}$$

where $KL[q(\bullet)||p(\bullet)]$ is the Kullback-Leibler divergence (KL divergence), which is a measure of how probability distribution $p(\bullet)$ diverges from probability distribution $q(\bullet)$ (Kullback and Leibler, 1951). The ELBO is equivalent to the negative KL divergence plus the first term in Eq. (2). Therefore, minimizing the KL divergence is the same as maximizing ELBO (Blei et al., 2017).

Random variables in Bayesian neural networks can be divided into three types. The first type is the observed random variable, which is the dataset $\mathscr{D}$ defined previously. The global random variable $\Theta = \{\theta^k\}_{k=1}^{V_g}$ is the second type of random variable, which is employed to measure the probabilities for all observed samples. The third type is local random variables $\mathscr{Z} = \{Z_i\}_{i=1}^{N}$, where $Z_i = \{z_i^k\}_{k=1}^{V_l}$, and this kind of variable is only implemented in autoencoding variational Bayes (AEVB). The AEVB is an approximate inference model with stochastic gradient variational Bayes (SGVB) (Kingma and Welling, 2013).

In this paper, we employ the automatic differential variational inference (ADVI) method to estimate the parameters in the Bayesian neural network (ThePyMCDevelopmentTeam, 2017). The objective of ADVI is to estimate posterior distribution $p(\Theta, \mathscr{Z}, \mathscr{D})$

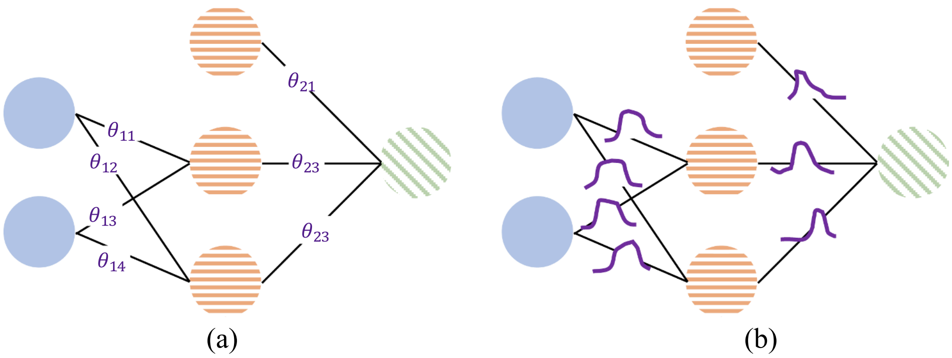

Fig. 4. Comparisons between a traditional neural network and a Bayesian neural network: (a) parameters in a traditional neural network are fixed values, (b) parameters in a Bayesian neural network are probability distributions.

by variational posterior $q(\Theta)\prod_{i=1}^{N}q(Z_i)$. Note that all the terms here are normal distributions. In ADVI, the full-rank Gaussian model is employed to generalize the mean-field (factorized) Gaussian approximation for the implicitly induced non-Gaussian variational distributions in the original latent variable set (Kucukelbir et al., 2017). More details about ADVI can be found in Kucukelbir et al. (2017). Therefore, $q(\Theta)$ depends on its means and standard deviations. These parameters can be represented as γ, which is a constant. However, the parameters of $q(Z_i)$ are determined by each observation. These parameters are represented as $\xi(y_i;\, \nu)$, where ν represents the parameters of $\xi(\bullet)$. For instance, in this paper, $\xi(\bullet)$ is the multilayer perceptron we construct.

Moreover, in Eq. (2), c_0, c_g and c_l are vectors to weight each term of $\mathscr{L}(\gamma,\, \nu)$. These terms can present as the following formats for better understanding:

$$c_0\log p(y_i|Z_i,\, \Theta) = \sum_{k=1}^{V_0} c_0^k \log p(y_i^k|pa(y_i^k,\, \Theta,\, \eta)) \tag{3}$$

$$c_g KL[q(\Theta)||p(\Theta)] = \sum_{k=1}^{V_g} c_g^k KL[q(\theta^k)||p(\theta^k|pa(\theta^k))] \tag{4}$$

$$c_l \sum_{i=1}^{N} KL[q(Z_i)||p(Z_i)] = \sum_{k=1}^{V_l} c_l^k KL[q(Z_i^k)||p(Z_i^k|pa(Z_i^k))] \tag{5}$$

where V_o, V_g and V_l are the numbers of types of observed random variables in the model, and all observed samples and random variables are utilized only in AEVB. $pa(v)$ represents the set of parent variables of v in the directed acyclic graph of the model.

5. Model results

5.1. Feature selection results

In this paper, we have 45 features and more than 20,000 records in total. To reduce the excessive computational load in the BNN, we first need to perform feature selection. In this paper, we adopt an elastic net (Zou and Hastie, 2005), which is a hybrid method combining LASSO and ridge regression. Elastic net takes advantage of both LASSO and ridge regression and overcomes their shortcomings. The formulation of the elastic net is shown as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, |Y - X\beta|^2 \quad subject \quad to(1-\alpha)||\beta||_1 + \alpha\,||\beta||_2^2 \le t \tag{6}$$

where α ranges from 0 to 1. Eq. (6) is LASSO when $\alpha = 0$ and ridge regression when $\alpha = 1$. We iterate 21 different α from 0 to 1 with a 0.05 step size. The experiment applies feature selection by using the elastic net with the artificial neural network (ANN). Moreover, in this paper, we do not perform validation of feature selection on the test set of the final model. Actually, we first separate the entire dataset into training set and test set in 80% vs. 20% fashion. Then we further divide the training set into feature selection training set (80%) and feature selection test set (20%). The best α is set as 0.75 after this experiment. As a result, there are 29 features selected from 45 features, as shown in Table 2. Clearly, the number of features chosen is reasonable. These 29 features are across four categories: social-demographic, trip information, Google Places API information, and Twitter information. Therefore, this feature subset completely captures traveler, trip, and place information. Moreover, this feature subset achieves the highest accuracy with the ANN. Afterwards, this feature subset is used for BNN model training, and the running time of the training process decreases with a 75% savings.

In the task of next purpose prediction, we employ similar features as a prediction of the current trip purpose. However, because the activity has not occurred, we do not have any location information about the next activity. Therefore, we utilize information about the current location in the model, including Google Places and Twitter information. In addition to the current location information, the purpose of the current activity that has just completed is included. Although the location of the next trip has not been decided, we still know the departure time, which is the same as the end time of previous activity. However, the trip duration and trip distance to the next activity location are unknown. We conduct feature selection by using the elastic et as well, and the result is shown in Table 3. In this case, $\alpha = 0.2$, and there are 24 out of 43 features.

It is found that the means and standard deviations for Twitter-related features are larger than the Google Place API-related features. This means that we can obtain more information from Twitter at one location. Moreover, even though the Google-related features and Twitter-related features do not cover all the categories of POIs, the combination of these two types of features covers the categories across all types of POIs. This means that Google and Twitter-related features are complementary.

5.2. Classification results – Prediction of the current trip purpose

In this section, we further compare the BNN model with several state-of-art algorithms, including support vector machine (SVM), ANN, K-nearest neighbors (KNN), and random forest (RF) (Harrington, 2012). SVM aims to map data into a high dimension and construct a set of hyperplanes to divide the data with maximum margins. We tune SVM with grid search method, and set cost equals 10 and gamma equals 0.1, where cost controls the cost of misclassification on the training dataset (a large cost gives you lower bias and high variance), and gamma defines how far the influence of a single training example research (small gamma gives you low bias and high variance as well). ANN was introduced in the previous section, and we apply the same structure as the BNN model in this

Table 2
Feature selection results for current purpose prediction.

Features from elastic net	Feature description	Mean	Std dev.
Social-demographic			
Employment	If the traveler is an employee	0.549	0.498
Student	If the traveler is a student	0.284	0.553
Age	The agebin of the traveler	4.299	1.954
Trip information			
depTime	The departure time of the trip	13.206	4.209
TripDistance	The trip distance to the destination	6.831	34.125
TripDuration	The trip duration to the destination	1.000	1.758
Google places API information[*]			
G_MONEY	How many money Google Places POIs are near the trip end location	1.968	1.953
G_FOOD	How many food Google Places POIs are near the trip end location	4.271	4.459
G_BAR	How many bar Google Places POIs are near the trip end location	0.936	1.996
G_CARE	How many care Google Places POIs are near the trip end location	1.557	1.999
G_TRANS	How many transit/transportation Google Places POIs are near the trip end location	0.902	1.495
G_AUTO	How many auto Google Places POIs are near the trip end location	0.910	1.633
G_CIVIC	How many civic Google Places POIs are near the trip end location	1.038	2.727
G_IMPROVE	How many improvement Google Places POIs are near the trip end location	4.075	3.832
G_EDU	How many education Google Places POIs are near the trip end location	1.144	1.560
G_RELIGION	How many religion Google Places POIs are near the trip end location	0.517	0.844
Twitter information			
T_FOOD	How many food-related tweets are posted near the trip end location	29.170	62.290
T_CARE	How many care-related tweets are posted the near trip end location	5.208	14.108
T_STORE	How many store-related tweets are posted near the trip end location	22.003	67.592
T_TRANS	How many transit/transportation-related tweets are posted near the trip end location	3.431	11.177
T_RELIGION	How many religion-related tweets are posted near the trip end location	1.410	5.252
T_IMPROVE	How many improvement-related tweets are posted near the trip end location	7.834	18.713
T_LODGE	How many lodge-related tweets are posted near the trip end location	5.660	31.196

* Please refer to Ermagun et al. (2017) for more details about Google Places.

Table 3
Feature selection results for next purpose prediction.

Features from elastic net	Feature description	Mean	Std dev.
Social-demographic			
Employment	If the traveler is an employee	0.617	0.486
Student	If the traveler is a student	0.233	0.551
Age	The age bin of traveler	4.674	1.767
Trip information			
depTime	The departure time of the trip	14.578	3.999
TripPurpose	The trip purpose for the current activity that has just finished	–[*]	–
Google places API information			
G_MONEY	How many money Google Places POIs are near the current location	1.274	1.848
G_FOOD	How many food Google Places POIs are near the trip end location	2.893	4.392
G_BAR	How many bar Google Places POIs are near the trip end location	0.723	1.874
G_CARE	How many care Google Places POIs are near the trip end location	0.969	1.768
G_STORE	How many store Google Places POIs are near the trip end location	3.160	4.939
G_TRANS	How many transit/transportation Google Places POIs are near the trip end location	0.544	1.177
G_RELIGION	How many religion Google Places POIs are near the trip end location	0.312	0.682
G_CIVIC	How many civic Google Places POIs are near the trip end location	0.808	2.552
G_HEALTH	How many health Google Places POIs are near the trip end location	6.321	10.702
G_IMPROVE	How many improvement Google Places POIs are near the trip end location	2.657	3.026
G_EDU	How many education Google Places POIs are near the trip end location	0.749	1.290
G_LODGE	How many lodge Google Places POIs are near the trip end location	0.334	1.281
Twitter information			
T_LEISURE	How many leisure-related tweets posted near the trip end location	6.511	18.513
T_FOOD	How many food-related tweets posted near the trip end location	24.211	62.932
T_STORE	How many store-related tweets posted near the trip end location	17.454	60.972
T_AUTO	How many auto-related tweets posted near the trip end location	0.734	4.877
T_CIVIC	How many civic-related tweets posted near the trip end location	3.230	14.735
T_IMPROVE	How many improvement-related tweets posted near the trip end location	6.459	17.544
T_LODGE	How many lodge-related tweets posted near the trip end location	4.320	23.289

* "–" means that there is no mean or standard deviation available since the trip purpose is a discrete variable, which is represented by characters.

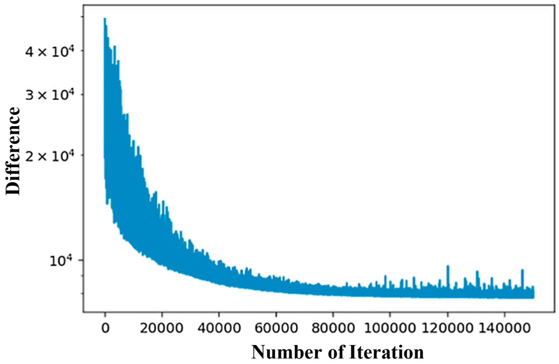

Fig. 5. The change in the difference between the estimated posterior and the true posterior.

experiment. KNN tends to classify objects to the class of the majority vote of its neighbors, and we set k equal to 6 after tuning. RF is an ensemble learning method that constructs numbers of decision trees together in the training stage, and the classification result is defined as the mode of classes. After tuning RF by using a grid search method, we set the number of variables randomly sampled as candidates at each split equal (mtry) to 6, and the number of trees to grow (ntree) equal to 4500. Finally, the accuracy and F1 score are employed to measure the performances, and the equation of the F1 score is shown below.

$$\text{Accuracy} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \tag{7}$$

$$\mathrm{F}_1 = 2 \times \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{8}$$

In the dataset, more than 40% of the trip purposes are home and work, and this observation complies with one's daily life pattern. However, the model trained with this dataset may lead to a large bias. As a result, we utilize data excluding home and work trip purposes (Ermagun et al., 2017). Moreover, for model validation, we implement 5-fold cross-validation and average the results.

A different number of hidden layers of BNN models is tested in this paper. However, the accuracy of the validation set will drop very fast when the number of hidden layers is larger than 2. One possible reason could be that when a multi-layer neural network consists of too many layers, it becomes harder and harder to find good weights and may cause the vanishing gradient problem. Therefore, we choose BNN with two hidden layers. For the number of neurons in hidden layers, we also examine different ones and select the best one which is 10 according to the accuracy of the validation set.

In the AVDI method, many iterations are conducted to minimize the KL divergence, which is equivalent to maximizing ELBO. Fig. 5 shows the change in the difference between the estimated posterior and the true posterior over the number of iterations. More iterations require longer computational times. However, the performance may not improve much after a certain number of iterations. As shown in Fig. 5, the difference does not decrease after 80,000 iterations. Therefore, we set the number of iterations as 80,000 in this paper.

One of the advantages of a BNN is that it can provide the probability for each category in as the trip purpose. We not only compare the top 1 result (the category with the highest prediction probability) of the BNN model named BNN-top-1 but also explore whether the top 2 and top 3 predictions are in the correct category. We denote them as BNN-top-2 and BNN-top-3, respectively.

The prediction confusion matrix of BNN-top-1 is illustrated in Fig. 6. The x-axis represents the predicted classes, and the y-axis represents the true classes. Each yellow line in the figure indicates a prediction result. For example, if a line is located in the diagonal region (A, A) where A represents a class, the predicted result is the same as the true class, and the prediction is correct. However, if a line is located in a nondiagonal region (A, B), it means that instead of the true class B, the prediction result is A. The density of the yellow lines in a diagonal box indicates the accuracy of that class or trip purpose.

In Table 4, one can observe that the BNN models outperform other algorithms on most trip purpose categories in accuracy and F1 score. The accuracy and F1 score of "Personal", which represent personal business activities, is lower compared to other methods. Additionally, the prediction performance of "EatOut" is relatively lower than "Recreation", "Education" and "Transportation". One reason is due to multiple activities in one place. Because one building or one plaza contains a number of merchants and offices, especially these three functions (shopping, personal business, and eating out), individuals can conduct several kinds of activities by just visiting one place. In this situation, the trip purpose is extremely difficult to define, and individuals may only report one trip purpose to represent multiple activities. Moreover, "Shopping", "Personal" and "EatOut" are difficult to distinguish from each other because they have similar activity durations. For instance, 13% of "Personal" activities are predicted to be "Shopping" activities, and 13% of "Personal" activities are predicted to be "EatOut" activities. Even the highest accuracy of "Personal" achieved by random forest is 59.24%, which is not high. This may be because personal business is the most complicated category. People are prone to categorize an activity as "Personal" when it is difficult to distinguish the activity as another category. Unlike "Education", "Shopping", or "EatOut", which only contain a specific activity or action, "Personal" consists of different activities or motions, including volunteer work or activities, servicing a private vehicle, visits to government offices, and civic or religious activities. To better

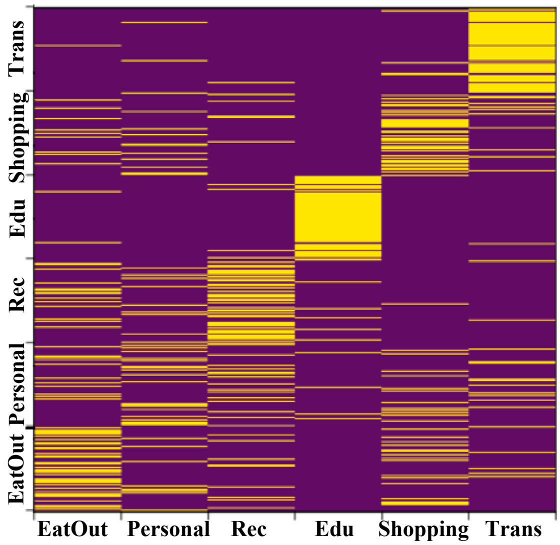

Fig. 6. Prediction confusion matrix of the BNN. A yellow line in the diagonal region means the predicted result is the same as the true class. Otherwise, the prediction is incorrect.

Table 4
Performance comparisons of the prediction of current trip purpose.

Precision	SVM	ANN	KNN	RF	BNN-top-1	BNN-top-2	BNN-top-3
EatOut	38.55%	37.20%	43.20%	52.78%	**64.26%**	**74.38%**	**85.70%**
Personal	41.05%	22.80%	38.00%	**59.24%**	48.17%	**60.94%**	**92.43%**
Recreation	46.15%	45.60%	44.80%	56.25%	**62.30%**	**80.17%**	**89.11%**
Education	64.44%	54.80%	48.00%	67.89%	**91.83%**	**93.70%**	**95.49%**
Shopping	51.48%	56.00%	53.20%	63.36%	**64.77%**	**76.26%**	**89.36%**
Transportation	57.77%	72.80%	43.60%	72.03%	**84.85%**	**88.00%**	**91.06%**
Average	49.91%	48.20%	45.13%	61.93%	**69.36%**	**78.91%**	**90.52%**
F1 Score	SVM	ANN	KNN	RF	BNN-top-1	BNN-top-2	BNN-top-3
EatOut	0.485	0.401	0.442	0.582	**0.636**	**0.742**	**0.864**
Personal	0.355	0.282	0.382	0.457	**0.575**	**0.704**	**0.927**
Recreation	0.389	0.468	0.414	0.591	**0.616**	**0.771**	**0.883**
Education	0.540	0.639	0.548	0.632	**0.905**	**0.951**	**0.971**
Shopping	0.512	0.426	0.491	0.626	**0.649**	**0.746**	**0.900**
Transportation	0.626	0.652	0.442	0.731	**0.752**	**0.812**	**0.887**
Average	0.484	0.478	0.453	0.603	**0.689**	**0.788**	**0.905**

Bold values represent the highest precision or F1 score among different machine learning algorithms for each category, respectively.

identify "Personal" activities, we can separate it into more specific categories. Additionally, one can put more effort into finding features that can better identify "Personal".

In general, a BNN, which estimates every parameter in the model as a distribution, provides a better description of data and leads to better accuracy. Therefore, the BNN models score the highest average accuracy. Moreover, they achieve extremely high performance for "Education" activities, whereas other algorithms only reach low accuracy. As a result, the BNN model is very powerful in the prediction of trip purpose.

Furthermore, we conduct experiments on specific features to determine how much the accuracy will decrease if such features are removed from the model. The results are shown in Fig. 7.

As shown in Fig. 7, if the model does not include features from Google Places and Twitter, the accuracy will drop the most. Then, the following are features from trip information and those from the trip duration and departure time. It is found that trip duration is one of the most important features of the model. Therefore, the trip duration is a significant factor to infer the trip purpose. The features from Google Places are also very significant. Furthermore, the Twitter-related features alone also improve prediction accuracy. If we remove Google Places and Twitter-related features, the accuracy decreases by 8% and 2%, respectively. However, if both are missing, the accuracy decreases by almost 16%. Therefore, Google Places and Twitter data are essential to improve the accuracy of predicting trip purpose.

In "EatOut" in Fig. 8(a), the Twitter data contribute as much as a 5% accuracy increase, and combining Twitter and Google Places information increases the accuracy by 30.9%. For "Education" activities, demographic-related features are the most important

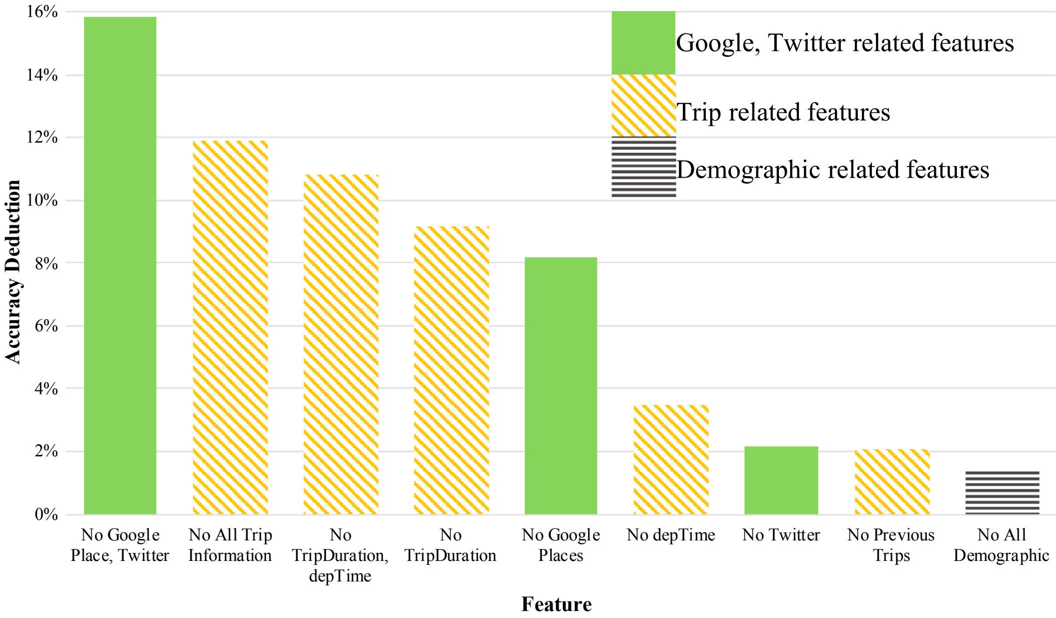

Fig. 7. Accuracy deduction for removing a different set of features.

because travelers who are students must go to school. Additionally, “Education” shows the most stable performance, in which the highest accuracy deduction is only 4.1%. Finally, for “Transportation” activities, trip-related features are the most important. It is also worth noting that accuracy decreases more if Google and Twitter features are removed together than the summation of removing them separately in most categories. Fig. 8 shows that Google and Twitter-related features are important for “EatOut”, “Personal”, “Recreation” and “Shopping”. It is reasonably concluded that people tend to tweet during these activities. The Google and Twitter information works in previous activities; however, it does not perform well in “Education” and “Transportation” activities. These have fewer tweets about these two types of activities. For “Education”, students do not post much when they study at school, and this activity is highly related to the social demographics of participants. In addition, people do not tweet frequently during “Transportation” activities, like transit, changing type of transportation, and pick-ups or drop-offs. First, it could be very dangerous to use cell phones while driving and walking. Second, there may be no signal when people are taking subways.

5.3. *Classification results – Prediction of the next trip purpose*

In this section, we also employ a BNN and traditional machine learning methods to complete this task. Parameters of SVM, gamma and cost, are configured as 1.1 and 100, respectively. For parameters in RF, mtry and ntree are set to be 8 and 2000, respectively. The performance of the BNN model is shown in Fig. 9 and Table 5. In this task, we do not include “Home” and “Work” as predicting both can achieve relatively high accuracy (above 90%). “Education” performs well as the previous task. Without the information on the location of the next activity, “Personal” and “Recreation” have relatively low performance. Because the survey data only contain one-day travel surveys, we cannot obtain much travel behavior information. If we analyze individuals’ multiday travel, the accustomed travel sequences will lead to higher accuracy. Overall, the BNN model outperforms other traditional machine learning algorithms.

Comparing Tables 4 and 5, the accuracy of the models for next purpose prediction is generally higher than the accuracy of the models for current purpose prediction. The reason is that they have different prediction timing. Current purpose prediction is conducted before an activity occurs, whereas next purpose prediction occurs after an activity completes. This results in different trip or activity information in the model. For next purpose prediction, the model contains the purpose of the current activity. This historical information is the key factor in achieving higher accuracy for next purpose prediction. However, it is worth noting that we do not use the previous activity purpose to predict the current trip purpose. We assume the model only knows what has just occurred. In the prediction of the current trip purpose, the features include departure time, trip distance and trip duration information, which are information about the previous trip. Moreover, for all second activities, most previous activities are “Home”, which makes the model inaccurate.

The analysis of the accuracy deduction by removing different sets of features is also conducted in this section, as shown in Fig. 10. If the model does not include features of Google Places and Twitter, the accuracy decreases the most. The runners-up are the model without Google Places and the model without Twitter information. Therefore, social media information is important in trip purpose prediction. Because the location of the next activity is unknown, the trip-related features are only the departure time and the previous trip purpose. Fig. 10 shows that the accuracy deduction for the model without all of the trip information is almost the same as the model without the previous trip purpose. This means that the previous trip purpose accounts for the most prediction power in all of the trip information. Additionally, demographic-related features are important to decide whether the next trip purpose is “School” or “Work”.

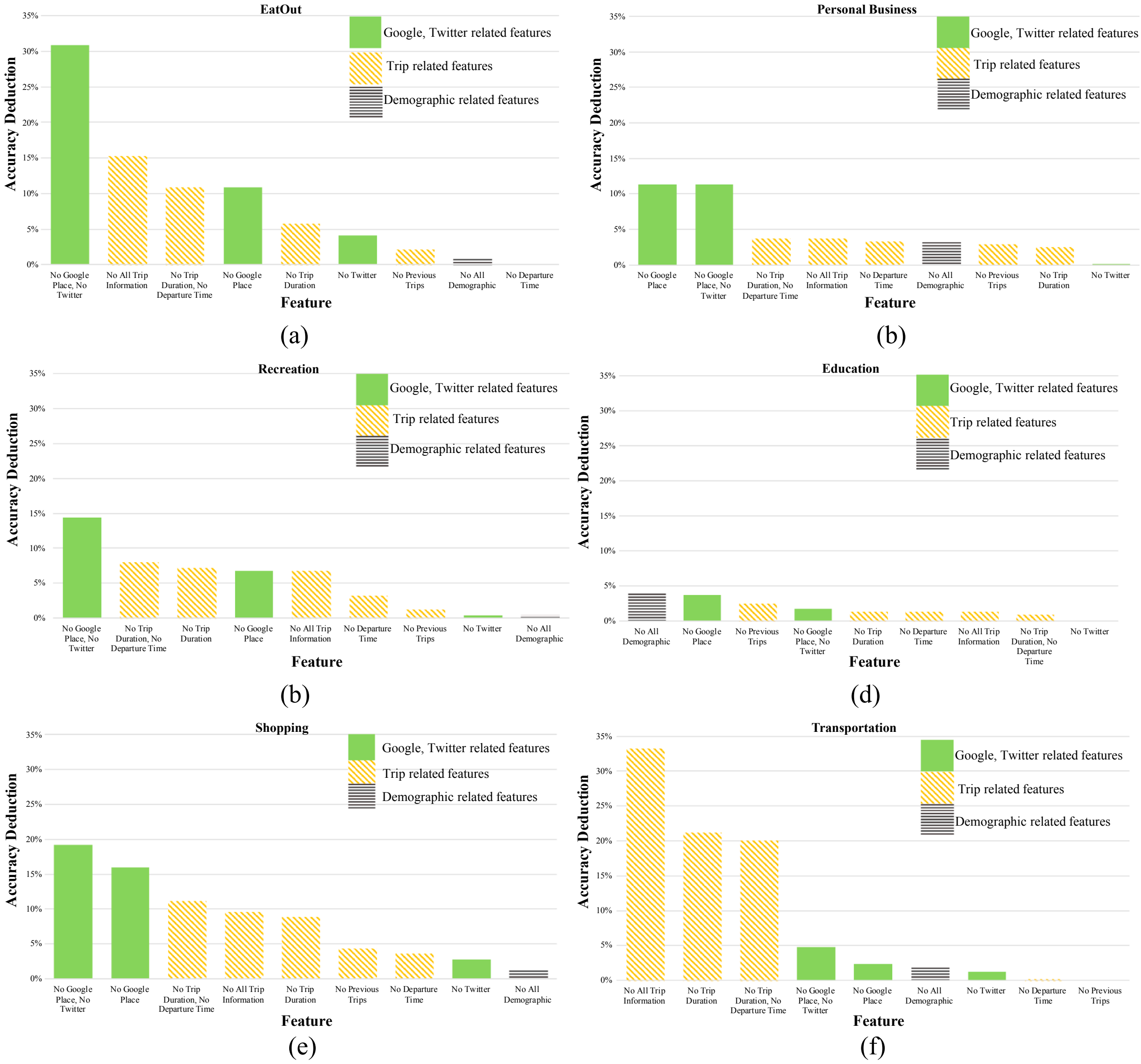

Fig. 8. Accuracy deduction for removing a different set of features for each category.

6. Potential applications

Our research has the following possible applications. First, this can be utilized in activity-based travel demand modeling. Our method provides better results when predicting the trip purpose given a location. It can further enhance the accuracy of demand forecasting. The second is survey labeling assistance. Whenever a user finishes an activity, the survey labeling assistance service can present three top predictions, ordered by their probability. Then, users can choose the correct one instead of filling out the survey. Our research can also be applied to online recommendations. Once the user inputs a destination in the online recommendation system, the proposed method can provide a prediction of what the user might do in that location. Based on the predicted activity, we can recommend shops or display corresponding advertisements to the user. In the future, researchers can develop better methods to mine more useful information from social media data, and this information can benefit the modeling and estimation of travel behavior. Moreover, with multiple-day survey data instead of one-day survey data, more user information can be obtained to promote the accuracy of the prediction.

7. Conclusions

This paper conducts two tasks in trip purpose prediction, including both current purpose prediction and next purpose prediction.

First, this paper implements a feature selection method with elastic net. The feature selection procedure is essential in that it remarkably reduces the running time of the BNN by 75% in the task of current purpose prediction.

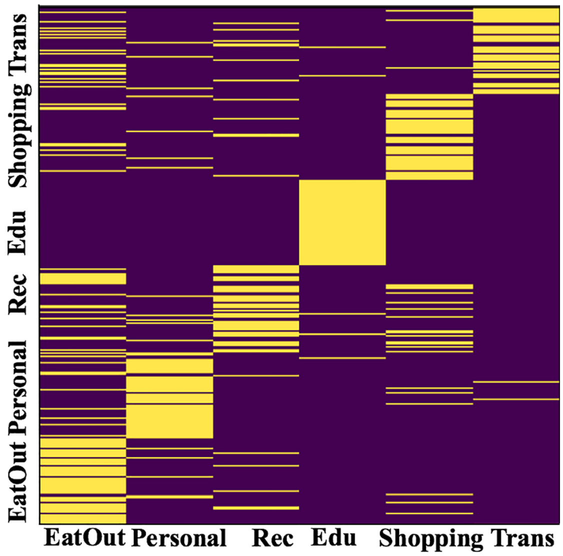

Fig. 9. Prediction confusion matrix of the BNN for prediction of next trip purpose. A yellow line in the diagonal region means that the predicted result is the same as the true class. Otherwise, the prediction is incorrect.

Table 5
Performance comparisons of the prediction of next trip purpose.

Precision	SVM	ANN	KNN	RF	BNN-top-1	BNN-top-2	BNN-top-3
EatOut	44.08%	48.56%	75.71%	58.42%	**79.27%**	**80.25%**	**91.26%**
Personal	29.29%	13.99%	33.37%	**82.38%**	67.11%	**83.24%**	**91.67%**
Recreation	45.11%	26.23%	43.58%	54.30%	**68.82%**	**74.33%**	**93.50%**
Education	94.60%	99.69%	98.72%	92.46%	**94.18%**	**99.90%**	**100.00%**
Shopping	42.56%	44.60%	48.16%	65.22%	**71.51%**	**77.47%**	**92.51%**
Transportation	77.76%	79.29%	47.15%	65.76%	**86.86%**	**87.77%**	**92.59%**
Average	55.57%	52.06%	57.78%	69.76%	**77.96%**	**83.83%**	**93.59%**
F1 Score	SVM	ANN	KNN	RF	BNN-top-1	BNN-top-2	BNN-top-3
EatOut	0.472	0.511	0.621	0.634	**0.731**	**0.778**	**0.823**
Personal	0.364	0.404	0.378	**0.783**	0.661	**0.784**	**0.902**
Recreation	0.462	0.182	0.424	0.572	**0.711**	**0.726**	**0.986**
Education	0.842	0.815	0.854	0.860	**0.936**	**1.000**	**1.000**
Shopping	0.548	0.297	0.470	0.652	**0.676**	**0.700**	**0.899**
Transportation	0.536	0.569	0.542	0.660	**0.877**	**0.925**	**0.926**
Average	0.538	0.463	0.549	0.694	**0.766**	**0.819**	**0.923**

Second, this study employs a Bayesian neural network to model the trip purpose. The BNN models outperform other prevailing algorithms. One major advantage of the Bayesian model is that it can return the possibility of each potential activity. The experiment shows a very high probability of correct prediction within the top 2 or top 3 ranked results.

Third, we observe that Google Places and Tweets greatly increase the accuracy compared with the model with other features, especially when predicting the activities “EatOut”, “Personal”, “Recreation” and “Shopping”.

Fourth, this research also conducts the prediction of the next trip purpose when the next activity location is unknown, and the model achieves high accuracy. Compared with the chosen features of predicting the current trip purpose and predicting the next trip purpose, we find that the model performance of predicting the next trip purpose highly relies on the information of the previous activity.

There still exist some limitations in this study. “Personal” activities need contemplation due to the complexity. Moreover, the data collection duration is relatively long because there are retrieval rate limits for both Twitter and the Google Places API within a time window. In addition, the population of social media users is different from the real-world population. Therefore, social media data may contain sampling bias. Additionally, individuals may post rumors or falsified information at activity places. This results in the issue of trustworthiness for the social media data.

BNN model generally outperforms some traditional machine learning models tested in this paper. However, Random Forest (RF) performs better in “Personal”. The underlying reason is that RF is such an ensemble learning method that it can achieve higher accuracy for complex classes, like “Personal”. Since RF performs especially well in this class, it can be employed for predicting “Personal” especially. In addition, a one vs all model for “Personal” also can be constructed in the further for better prediction results.

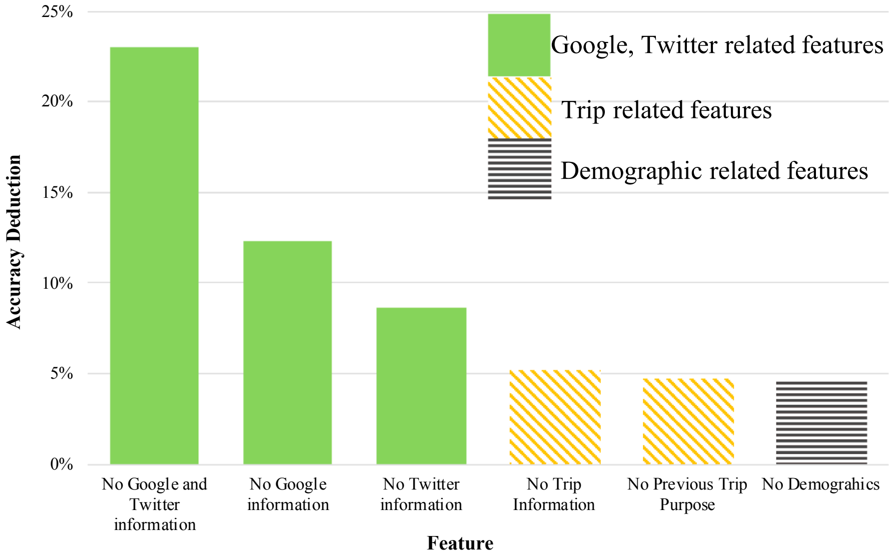

Fig. 10. Accuracy deduction for removing different sets of features.

In the future, one can extend this work with multiday GPS trajectories and incorporate travelers' trip histories. In addition, one can develop a social media-based travel survey and further implement this research in activity-based modeling and simulation.

Acknowledgements

This work was partially supported by National Science Foundation award CMMI-1637604 and Region 2 University Transportation Research Center. Authors also would like to thank Transportation Secure Data Center of National Renewable Energy (NREL) which provided us the dataset of 2010–2012 California Household Travel Survey.

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Courier Corporation.
Alexander, L., Jiang, S., Murga, M., González, M.C., 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. Transport. Res. Part C: Emerg. Technol. 58, 240–250.
Anderson, M., Perrin, A., 2017. Tech Adoption Climbs Among Older Adults. Pew Research Center.
Bhat, C.R., Lawton, T.K., 2000. Passenger Travel Demand Forecasting.
Blei, D.M., Kucukelbir, A., McAuliffe, J.D., 2017. Variational inference: a review for statisticians. J. Am. Stat. Assoc. 112, 859–877.
Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D., 2015. Weight Uncertainty in Neural Networks. Available from: arXiv preprint arXiv:1505.05424.
Bohte, W., Maat, K., 2008. Deriving and validating trip destinations and modes for multiday gps-based travel surveys: application in the netherlands. Transportation Research Board 87th Annual Meeting.
Broersma, M., Graham, T., 2012. Social media as beat: Tweets as a news source during the 2010 British and Dutch elections. J. Pract. 6, 403–419.
Castiglione, J., Bradley, M., Gliebe, J., 2015. Activity-Based Travel Demand Models: A Primer.
Chen, C., Gong, H., Lawson, C., Bialostozky, E., 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: lessons learned from the New York City case study. Transport. Res. Part A: Policy Pract. 44, 830–840.
Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transport. Res. Part C: Emerg. Technol. 68, 285–299.
Chen, P.-T., Chen, F., Qian, Z., 2014. Road traffic congestion monitoring in social media with hinge-loss Markov random fields. In: 2014 IEEE International Conference on Data Mining (ICDM), 2014. IEEE, pp. 80–89.
Cici, B., Markopoulou, A., Frias-Martinez, E., Laoutaris, N., 2014. Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, pp. 201–211.
Cui, Y., He, Q., Khani, A., 2018. Travel behavior classification: an approach with social network and deep learning. Transport. Res. Record 0361198118772723.
Culnan, M.J., McHugh, P.J., Zubillaga, J.I., 2010. How large US companies can use Twitter and other social media to gain business value. MIS Quart. Execut. 9.
D'Andrea, E., Ducange, P., Lazzerini, B., Marcelloni, F., 2015. Real-time detection of traffic from twitter stream analysis. IEEE Trans. Intell. Transport. Syst. 16, 2269–2283.
Daly, E.M., Lecue, F., Bicer, V., 2013. Westland row why so slow?: fusing social media and linked data sources for understanding real-time traffic conditions. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces. ACM, pp. 203–212.
Deng, Z., Ji, M., 2010. Deriving rules for trip purpose identification from GPS travel survey data and land use data: a machine learning approach. Traffic and Transportation Studies 2010.
Denker, J., Schwartz, D., Wittner, B., Solla, S., Howard, R., Jackel, L., Hopfield, J., 1987. Large automatic learning, rule extraction, and generalization. Complex Syst. 1, 877–922.
Dong, W., Li, J., Yao, R., Li, C., Yuan, T., Wang, L., 2016. Characterizing driving styles with deep learning. Available from: arXiv preprint arXiv:1607.03611.
Ermagun, A., Fan, Y., Wolfson, J., Adomavicius, G., Das, K., 2017. Real-time trip purpose prediction using online location-based search and discovery services. Transport. Res. Part C: Emerg. Technol. 77, 96–112.
Gal-Tzur, A., Grant-Muller, S.M., Kuflik, T., Minkov, E., Nocera, S., Shoor, I., 2014. The potential of social media in delivering transport policy goals. Transport Policy 32, 115–123.
Gers, F.A., Schraudolph, N.N., Schmidhuber, J., 2002. Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. 3, 115–143.

Grosenick, S., 2012. Real-Time Traffic Prediction Improvement Through Semantic Mining of Social Networks. University of Washington.
Gu, Y., Qian, Z.S., Chen, F., 2016. From Twitter to detector: real-time traffic incident detection using social media data. Transport. Res. Part C: Emerg. Technol. 67, 321–342.
Harrington, P., 2012. Machine learning in action. Manning, Greenwich CT.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Janocha, K., Czarnecki, W.M., 2017. On Loss Functions for Deep Neural Networks in Classification. Available from: arXiv preprint arXiv:1702.05659.
Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? Adv. Neural Inform. Process. Syst. 5580–5590.
Kingma, D.P., Welling, M., 2013. Auto-Encoding Variational Bayes. Available from: arXiv preprint arXiv:1312.6114.
Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., Blei, D.M., 2017. Automatic differentiation variational inference. J. Mach. Learn. Res. 18, 430–474.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.
Laboratory, N.R.E., 2017. Transportation Secure Data Center. In: Laboratory, N.R.E. (Ed.). < www.nrel.gov/tsdc > .
Li, R., Lei, K.H., Khadiwala, R., Chang, K.C.-C., 2012. Tedas: a twitter-based event detection and analysis system. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE). IEEE, pp. 1273–1276.
Liao, L., Fox, D., Kautz, H., 2007. Extracting places and activities from gps traces using hierarchical conditional random fields. Int. J. Robot. Res. 26, 119–134.
Lin, L., Ni, M., He, Q., Gao, J., Sadek, A.W., 2015. Modeling the impacts of inclement weather on freeway traffic speed: exploratory study with social media data. Transport. Res. Rec.: J. Transport. Res. Board 82–89.
Liu, Y., Yan, X., Wang, Y., Yang, Z., Wu, J., 2017. Grid mapping for spatial pattern analyses of recurrent urban traffic congestion based on taxi GPS sensing data. Sustainability 9, 533.
Lu, Y., Liu, Y., 2012. Pervasive location acquisition technologies: opportunities and challenges for geospatial studies. Comput. Environ. Urban Syst. 36, 105–108.
Lv, Y., Chen, Y., Zhang, X., Duan, Y., Li, N.L., 2017. Social media based transportation research: the state of the work and the networking. IEEE/CAA J. Autom. Sinica 4, 19–26.
Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.-Y., 2015. Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intell. Transport. Syst. 16, 865–873.
Mackay, D.J., 1992. A practical Bayesian framework for backpropagation networks. Neural Comput. 4, 448–472.
Mai, E., Hranac, R., 2013. Twitter interactions as a data source for transportation incidents. Proc. Transportation Research Board 92nd Ann. Meeting.
Meng, C., Cui, Y., He, Q., Su, L., Gao, J., 2017. Travel purpose inference with GPS trajectories, POIs, and geo-tagged social media data. 2017 IEEE International Conference on Big Data (Big Data) IEEE, 1319–1324.
Mitchell, T.M., 1997. Machine learning. WCB. McGraw-Hill, Boston, MA.
Ni, M., He, Q., Gao, J., 2014. Using social media to predict traffic flow under special event conditions. The 93rd Annual Meeting of Transportation Research Board.
Ni, M., He, Q., Gao, J., 2017. Forecasting the subway passenger flow under event occurrences with social media. IEEE Trans. Intell. Transport. Syst. 18, 1623–1632.
Oliveira, M., Vovsha, P., Wolf, J., Mitchell, M., 2014. Evaluation of two methods for identifying trip purpose in GPS-based household travel surveys. Transport. Res. Rec.: J. Transport. Res. Board 33–41.
Ponieman, N.B., Salles, A., Sarraute, C., 2013. Human mobility and predictability enriched by social phenomena information. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, pp. 1331–1336.
Saeedimoghaddam, M., Kim, C., 2017. Modeling a spatio-temporal individual travel behavior using geotagged social network data: a case study of greater cincinnati. In: ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, pp. 4.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Schneider, C.M., Rudloff, C., Bauer, D., González, M.C., 2013. Daily travel behavior: lessons from a week-long survey for the extraction of human mobility motifs related information. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. ACM, pp. 3.
Schulz, A., Ristoski, P., Paulheim, H., 2013. I see a car crash: Real-time detection of small scale incidents in microblogs. In: Extended Semantic Web Conference. Springer, pp. 22–33.
Su, X., Caceres, H., Tong, H., He, Q., 2016. Online travel mode identification using smartphones with battery saving considerations. IEEE Trans. Intell. Transport. Syst. 17, 2921–2934.
ThePyMCDevelopmentTeam, 2017. Inference (Accessed 2017). < http://docs.pymc.io/api/inference.html > .
Wang, F.-Y., 2014. Scanning the issue and beyond: Real-time social transportation with online social signals. IEEE Trans. Intell. Transport. Syst. 15, 909–914.
Wanichayapong, N., Pruthipunyaskul, W., Pattara-Atikom, W., Chaovalit, P., 2011. Social-based traffic information extraction and classification. In: 2011 11th International Conference on ITS Telecommunications (ITST). IEEE, pp. 107–112.
Wolf, J., Bricka, S., Ashby, T., Gorugantua, C., 2004. Advances in the application of GPS to household travel surveys. National Household Travel Survey Conference, Washington DC.
Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. Transport. Res. Rec.: J. Transport. Res. Board 125–134.
Xiao, G., Juan, Z., Zhang, C., 2016. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. Transport. Res. Part C: Emerg. Technol. 71, 447–463.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. Acc. Anal. Prevent. 39, 922–933.
Yang, X., 2017. Understanding the Variational Lower Bound.
Yates, D., Paquette, S., 2011. Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int. J. Inform. Manage. 31, 6–13.
Zhang, Z., He, Q., Gao, J., Ni, M., 2018. A deep learning approach for detecting traffic accidents from social media data. Transport. Res. Part C 86, 580–596.
Zhang, Z., He, Q., Zhu, S., 2017. Potentials of using social media to infer the longitudinal travel behavior: a sequential model-based clustering method. Transport. Res. Part C: Emerg. Technol. 85, 396–414.
Zhang, Z., Ni, M., He, Q., Gao, J., Gou, J., Li, X., 2016. An exploratory study on the correlation between twitter concentration and traffic surge. Transport. Res. Rec. 35, 36.
Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., Zhang, Q., Yang, L., 2016. Big data for social transportation. IEEE Trans. Intell. Transport. Syst. 17, 620–630.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Royal Stat. Soc.: Ser. B (Stat. Methodol.) 67, 301–320.