MoCha: Large-Scale Driving Pattern Characterization for Usage-based Insurance

Zhihan Fang
Rutgers University
Piscataway, NJ, USA
zhihan.fang@cs.rutgers.edu

Guang Yang
Rutgers University
Piscataway, NJ, USA
gy121@cs.rutgers.edu

Dian Zhang
Shenzhen University
Shenzhen, China
serena.dian@gmail.com

Xiaoyang Xie
Rutgers University
Piscataway, NJ, USA
xx88@cs.rutgers.edu

Guang Wang
Rutgers University
Piscataway, NJ, USA
gw255@cs.rutgers.edu

Yu Yang
Rutgers University
Piscataway, NJ, USA
yu.yang@rutgers.edu

Fan Zhang
SIAT, CAS
Shenzhen, USA
zhangfan@siat.ac.cn

Desheng Zhang
Rutgers University
Piscataway, NJ, USA
desheng@cs.rutgers.edu

ABSTRACT

Given widely adopted vehicle tracking technologies, usage-based insurance has been a rising market over the past few years. With potential discounts from insurance companies, customers voluntarily install sensing devices in their vehicles for insurance companies, which are utilized to analyze their historical driving patterns to derive the risks of future driving. However, it is challenging to characterize and predict driving patterns, especially for new users with limited data. To address this issue, we propose and evaluate a system called MoCha to accurately characterize driving patterns for usage-based insurance. The key question we aim to explore with MoCha is whether we can fully explore long-term driving patterns of new users with only limited historical data of themselves by leveraging abundant data of other users and contextual information. To answer this question, we design (i) a multi-level driving pattern modeling component to capture the spatial-temporal dependency on both individual and group level, and (ii) a multi-task learning method to utilize underlying relations of driving metrics and predict multiple driving metrics simultaneously. We implement and evaluate MoCha with real-world on-board diagnostics data from a large insurance company with more than 340,000 vehicles. Further, we validate the usefulness of MoCha by predicting driving risks based on real-world claim data in a Chinese city, Shenzhen.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*; • **Social and professional topics** → User characteristics; • **Information systems** → *Information systems applications.*

KEYWORDS

Usage-based Insurance, Driving Patterns, User Mobility

* Prof. Dian Zhang is the corresponding author of this paper

KDD '21, August 14–18, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8332-5/21/08…$15.00
https://doi.org/10.1145/3447548.3467114

ACM Reference Format:
Zhihan Fang, Guang Yang, Dian Zhang, Xiaoyang Xie, Guang Wang, Yu Yang, Fan Zhang, and Desheng Zhang. 2021. MoCha: Large-Scale Driving Pattern Characterization for Usage-based Insurance. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3447548.3467114

1 INTRODUCTION

Usage-Based Insurance (UBI) started in 2003 and has now become mainstream offered by most insurance companies especially across North America, Europe, and Asia [23]. Currently, several U.S. insurance companies offer usage-based insurance such as Progressive, StateFarm, Metromile, and Allstate. In general, these companies provide UBI based on users' potential driving risk, which is modeled by many factors including distance, speed, time, and detailed contexts (e.g., traffic congestion) [22]. With UBI, an auto insurance company tracks how users use their vehicles under users' consent and then quantifies the risks of their future driving [22].

There are a few approaches to implement UBI based on Black-Box devices [16], OBD-II devices with Smartphone apps [8], and a hybrid approach of them [9]. All these approaches log speeds, locations, or both, when a vehicle is driven and uploads these data through a smartphone app, which makes these data user-specific. Based on this logged information, insurance companies analyze driving patterns in terms of three key metrics, i.e., distance, time, and speed [21]. Previous studies [22] [26]show that these three vehicle usage metrics are the main factors to quantify potential future risks through the multi-factor fitting, along with other metrics, e.g., car and road condition, traffic density, user experience. We verified this assumption in our motivation section. Thus, how to predict future driving metrics (e.g., distance, time, and speed) and resultant risks based on historical data is essential for UBI.

Currently, based on the interactions with a major insurance company from which we obtain user-specific On-Board Diagnostics (OBD) data (uploaded by smartphone apps), the driving pattern analysis they are using is mostly statistical on the individual user level, which works fine for users with sufficient data and stable driving patterns. However, there are two kinds of users introducing significant challenges for a UBI company to continuously characterize their mobility patterns by three driving metrics (i.e., distance, time, and speed): (i) newly insured users with limited historical

data (e.g., transferred from another company); (ii) some existing users with evolving driving behaviors (e.g., new job, new home, new grocer, new school for kids).

To address this challenge, we explore two fundamental aspects of vehicular mobility. (1) It has been shown that the role of a user (e.g., daily commuters, Uber drivers) has an important impact on mobility patterns [25][14], which inspires us to group users by sophisticated mobility features (e.g., home, work, frequent routes, etc) to compensate for newly insured users with limited data or existing established users with bias historical patterns due to new driving behaviors. Due to the dynamic roles of drivers and evolving driving patterns, we periodically update our user groups with update-to-date features to compensate for new driving behaviors. Technically, we design a multi-modal learning component to capture the driving pattern based on **individual user level (one modal)** and **user group level (the other modal)** features. (2) The key driving metrics including distance, time, and speed variance are highly correlated with each other, so prediction results of one metric can improve the prediction results of others. It motivates us to design a multi-task learning component to predict the three target metrics simultaneously where the prediction of each metric is a task.

Based on these two components, we design and test a system called MoCha for **Mo**bility **Cha**racterization. The key novelty of MoCha is to jointly consider (1) individual-level and dynamic group-level mobility with multi-modal learning; (2) correlation of driving metrics with multi-task learning to accurately predict three driving metrics (i.e., distance, time and speed variance) and driving risks for both new UBI users with limited data and established UBI users with involving mobility patterns, in contrast to existing methods for UBI mostly, if not all, focusing statistical methods with individual or static user groups. Our key contributions are as follows:

- To our knowledge, we design and test MoCha as the first system of large-scale driving pattern prediction for usage-based insurance. MoCha considers both (i) correlation between different UBI users for driving pattern grouping and (ii) correlation between different driving metrics to address two practical challenges regarding new users with limited data and established users with evolving patterns. The design insight of MoCha is based on real-world vehicular data with more than 340 thousand vehicles. Under the permission of the UBI company, we will share sample data of 1,000 anonymous vehicles for reproducibility to encourage researchers to work in this direction.
- We design a multi-modal learning component to integrate individual driving patterns with group driving patterns, where each of them serves as a concrete modality to improve each other. Specifically, we periodically cluster users into groups based on seven mobility features regarding essential mobility patterns including home/work locations, driving time/distance, etc. We treat individual driving patterns and group driving patterns as two separate modalities and integrate them with a multi-modal LSTM model. Further, we found the driving metrics to be predicted (i.e., distance, time, and speed variance) are highly correlated since they share spatial and temporal contextual factors such as traveled road types. It motivates us to design a multitask learning component to learn these metrics simultaneously to improve their prediction accuracy. We implement and evaluate MoCha with On-Board Diagnostics data from a national-scale insurance company with 340 thousand personal and commercial vehicles.
- More importantly, we deploy MoCha and validate its usefulness by predicting driving risks through predicted metrics based on real-world claim data as the ground truth. We found that the future driving metrics predicted by MoCha can be utilized by two learning models to predict their accident risks with an error rate of 25%.

2 MOTIVATIONS

2.1 Prediction Justification

Some insurance companies (including the one we are working with) provide a dynamic rate for users with a combination of driving distance and driving risk. The insurance rate consists of two parts, i.e., the base rate and per-mile rate, as shown in Figure 1.

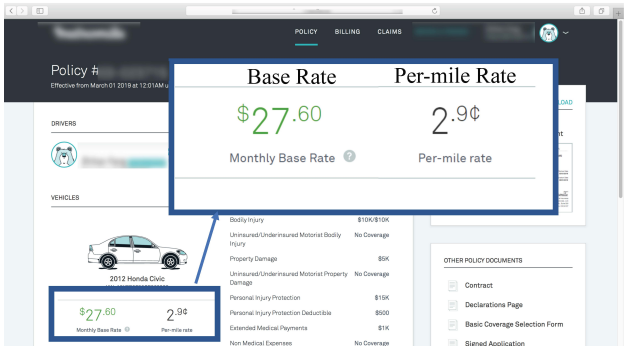

Fig. 1: UBI (metromile.com); the base rate and per-mile rate are determined by the potential future risk quantified by future predicted metrics.

At the beginning of every month, insurance companies offer users the rates of the upcoming month. The base rate and per-mile rate are two dynamic values and determined by the potential risk of a user, which is quantified through multi-metric analyses. The dynamic metrics include travel distance [11], travel time [22] [5], speed variance [24] [22], and static metrics such as gender, coverage, years of driving and previous claims [11], e.g., exceed speed limits, traffic signal violation, traffic accidents. At the end of each month, the total premium is calculated by the summation of two rates with a formula *base rate (quantified by predicted driving behaviors) + per mile rate (quantified by predicted driving behaviors) × driving miles.* Therefore, the total premium needs the prediction on future driving behaviors to determine the rates at the beginning of the month, even though it is paid by the end of the month based on the actual mileage.

2.2 Metric Justification

Based on the current practice of the UBI insurance we are working with, we explore a set of dynamic metrics as factors related to driving risk by utilizing real-world claim data and OBD data. We found various metrics captured by OBD data are correlated to driving risks. However, some detailed metrics, e.g., home/work locations, trip origin and destination, and travel time, are potential privacy issues since their information was not included in the UBI agreement. So we utilize these features as features to predict the

three metrics in almost all UBI agreements without privacy issues. Based on our dataset, as shown in Figure 2, we found that users with accidents have longer distances and time, along with higher standard deviations on speed, i.e., higher variance.

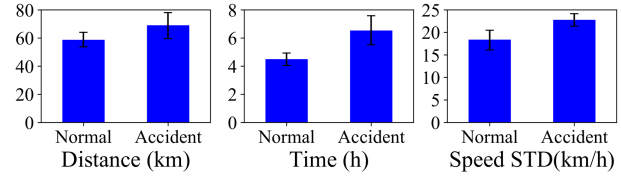

Fig. 2: Normal users w/o Acc. v.s. users w/ Acc.

2.3 Technical Challenges

New UBI Users with Limited Data: In the business of insurance companies, there are lots of incoming users with limited data. They are (i) experienced drivers starting to use UBI, (ii) new drivers getting their first cars, and (iii) a company with existing commercial vehicles switching to a new insurance model. Based on our analysis, we have around 0.1% - 0.5% new users without historical data coming into the systems every day, which makes their driving pattern prediction challenging. Based on interactions with our UBI collaborator, the company has to delay the intensive evaluation of a new customer until enough data are collected, which would potentially increase both the premiums for users and the risk of the UBI company. To provide some quantitative results, we study both new UBI users (i.e., the users with fewer than one-week data, which is suggested according to the domain experts from insurance companies and the fact that human mobility presents repeated weekly patterns [6]) and established users (i.e., the users with more than one-week data), for both personal and commercial vehicles. Given new UBI users, a naive method to predict their mobility patterns is to use the average value of existing observations from other users within the same category, but it leads to large prediction errors. As in Figure 3, we compare the predictability of two groups by calculating the standard deviation (STD) of travel distance, travel time, and travel speed in trips on the individual level. The formula for STD is given by $STD(X) = \sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2/(N-1)}$ where X is a collection of user data of a specific driving metric, such as travel distance, travel time, or travel speed. We use STD instead of Mean because we try to understand the variance of new users' driving behaviors since the variances are tied to the predictability of driving patterns. New users present lower predictability (larger STD) compared with existing users, which is caused by a limited sample size, e.g., the number of historical records.

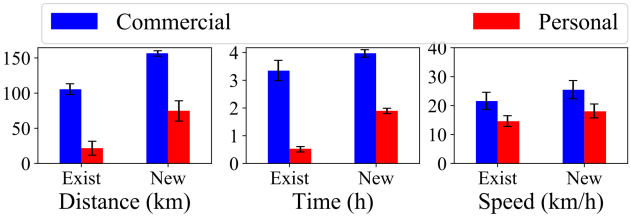

Fig. 3: STD of Three Driving Metrics

Established Users with Evolving Pattern: With data collected by OBD, the current solution of our collaborator is straightforward, i.e., to use historical vehicle usage directly obtained by OBD data as future vehicle usage. However, based on our analyses, we found that users' driving patterns have been evolving a lot given more collected data, and the existing solution has a large prediction error due to lower predictability (higher STD) in this group of users compared with routine users without much evolving on driving patterns as in Figure 4. This phenomenon is mainly caused by (i) new infrastructure (e.g., new roads, new shopping mall, long-term constructions such as new subway lines), (ii) new personal routines (e.g., moving to new apartments, picking up and dropping off kids to new schools), (iii) new business routines (e.g., new delivery areas for logistic trucks). The OBD data are uploaded in real-time based on a smartphone app and tied to individual users and vehicles. The details of the collected OBD data are given in Section 3. As a result, our collaborator plans to predict future driving patterns of both existing and new customers by building customer-specific models given limited yet constantly-accumulated OBD data.

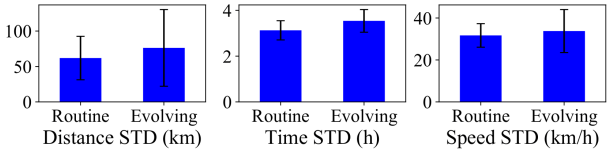

Fig. 4: Established Users with Evolving Patterns

3 DATASET AND PREPROCESSING

The details of 5-year the OBD system data are listed in Figure 5, where the direction field is given between 0 to 360 degrees to north.

	Personal Vehicles	Commercial Vehicles
Daily Data Size	11 GB	10 GB
# of Vehicles	295,001	60,773
# of Daily Records	240 million	85 million
Format	Device ID Date&Time Direction GPS&Speed	Device ID Date&Time Direction GPS&Speed

Fig. 5: Datasets

These two types of vehicles are spatiotemporally complementary to each other due to their purposes. Since we focus on an evolving scenario, we present the driving pattern evolving of UBI vehicles in Table 1. The average travel speed, travel distance, and travel time have been increasing in the recent 4 years starting 2017.

Table 1: 4-Year Evolving Pattern

Year	First Year	Second Year	Third Year	Forth Year
Ratio of New Vehicles	10.1 %	36.5%	47.3%	55.1%
Daily Distance (km)	30.43	33.02	35.20	36.81
Daily Time (h)	1.25	1.28	1.33	1.34
Average Speed (km/h)	22.02	23.77	24.74	25.81

Spatial Partition: To reduce the computational cost and improve prediction accuracy, we divide a geographic area into prefixed grids. A grid partition is given by (i) the maximum and minimum coordinates of the area; (ii) the length of each grid; (iii) individual grids with their coordinates. Based on the grid partition, a trajectory is modeled as continuous changes of grids on the spatial dimension.
Temporal Partition: Based on the previous study [6], human mobility shows a regular pattern given temporal contexts due to the periodicity of trips. Thus, we utilize a temporal partition including *Time of Day* (ToD) and *Day of Week* (DoW). ToD is indicated by different time slots within a day, e.g., a 10 min or one hour slot;

DoW is indicated by *Weekdays* and *Weekends* or *Monday* to *Sunday* because of different patterns in these two categories.

Trajectory vs. Trips: Based on the spatial and temporal partition settings, a vehicular trajectory captured by OBD devices is defined as a sequence of spatial-temporal changes, i.e.,

$$\text{traj.trace} = \{g_1, g_2, \cdots, g_n\};\ \text{traj.time} = \{t_1, t_2, \cdots, t_n\},$$

where g_i is a spatiotemporal record at timestamp t_i. In our system setting where onboard devices periodically collect GPS data, a trajectory is represented as continuous spatial changes after cleaning. Based on the above spatial-temporal partition, we divide a continuous physical trajectory of a user into several logical trips based on a temporal interval between OBD data. In OBD data uploading, the temporal interval is fixed at 10 seconds as long as the engine of a vehicle is on. We define a trip as a set of records from an engine-on and engine-off event.

4 MOCHA DESIGN

4.1 System Overview

We present an overview of Mocha in Figure 6 with three modules: (i) an *internal information feeder* to feed mobility data of individual users and user groups introduced in Sec. 4.2; (ii) an *external information feeder* to align contextual factors such as population distribution and road type distribution with users' mobility data introduced in Sec. 4.3; (iii) a *multi-modal multitask learning module* to predict future usage based on both internal mobility and external factors in Sec. 4.4.

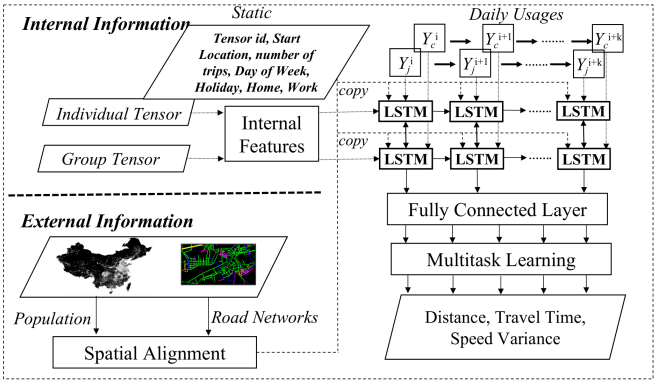

Fig. 6: Mobility Prediction Framework

4.2 Internal Information Feeder

The internal information feeder constructs two kinds of tensors, i.e., individual tensors and group tensors as output, and then feeds the tensor attributes to the learning module. An individual tensor describes attributes of single users and a group tensor describes attributes of a user group. The attributes in tensors are categorized into two types: (1) The static attributes are metrics describing general information, e.g., start location of one day, day of the week, etc. (2) The time series attributes are dynamic metrics used to describe detailed driving patterns, e.g., distance, travel time, speed variance.

4.2.1 Individual Tensors. We organize trips belonging to the same user (uploaded by the same smartphone) with a few mobility tensors $\mathcal{A} \in \mathbb{R}^{N \times N \times M}$ with the same dimensions.

- A temporal dimension indicates a specific time of day on a day of week when a trip starts (e.g., a slot from 0:00 AM to 0:05 AM on Monday): $[t_1, ..., t_M]$.
- A spatial dimension indicates specific spatial units as the origins of trips: $[g_1, ..., g_N]$.
- A spatial dimension indicates specific spatial units as the destinations of trips: $[g_1, ..., g_N]$.

For a vehicle, we use 5 datasets related to a trip with a fixed start time, origin, and destination to obtain five tensors: (1) Frequency: an entry is the frequency of a trip; (2) an entry is the average trip distance; (3) Duration: an entry is the average trip time; (4) Speed: an entry is the speed variance of a trip; (5) Route: an entry is a detailed trajectory on a grid level, which is represented by an additional matrix with features related to road types and population density related to a route. Without loss of generality, current driving pattern modeling focuses on factors including driving distance, time, and speed. Other factors such as acceleration (e.g., braking) and turning can be embedded in our model through tensor extension. The left part of Figure 7 gives an example of the individual mobility tensors.

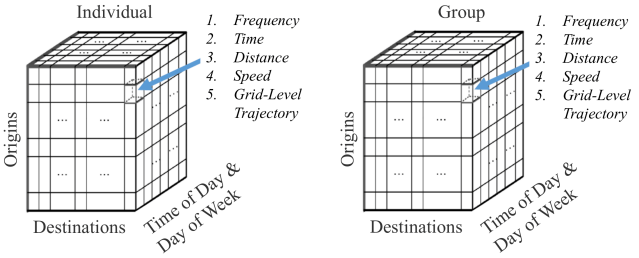

Fig. 7: Tensor Construction

4.2.2 Group Tensors. We utilize the individual mobility tensors (e.g., frequency, distance, time, speed, and route tensors) to represent the driving patterns of individual users. However, individual tensors limit the ability of our model to capture new mobility patterns because of new users with limited data. If we combine individual tensors, we have global tensors showing the driving patterns in the same city. Global tensors contain overall driving patterns, giving the ability to capture driving patterns for new users with limited data. But the key drawback of a city-wide global tensor is too generic due to a large number of users.

As a result, we aim to find some group tensors, which contain mobility patterns in a group of users with similar mobility patterns. Based on individual tensors, the users in the same group have very similar mobility patterns; whereas the users in different groups have very different mobility patterns. Therefore, we design a clustering algorithm to group users into different groups. For clustering, a feature vector is created for all groups and is dynamically updated when new OBD data are fed into the system. A vehicle's feature vector contains a set of advanced features, which are obtained by direct tensor operations, e.g, projection, and aggregation. We cluster the vehicles into different groups by three steps.

Step (i): Creating a Feature Vector for Each User. Based on the individual mobility tensors, the feature vector we used in MoCha is given in Table 2, in which (1) the home grid and work grid are inferred from spatial-temporal features of individual tensors based on existing work [10] (for commercial vehicles, instead of home/work locations, we use the top two frequent locations); (2) the average

daily driving time, average daily driving distance, and average standard deviation of daily speeds are daily mobility patterns on weekdays and weekends; (3) the number of daily distinct ODs is the distinct origin-destination pairs a user traveled in one day at the grid level on weekdays and weekends; (4) we use the number of daily trips on different days of the week as a weekly travel pattern. These features capture a permanent geometric distribution of a user during a long time period, which enables an effective clustering.

Table 2: User Features

Features	Values
Home Grid	g_i
Work Grid	g_j
Daily Driving Time	t_1,t_2
Daily Driving Distance	d_1,d_2
Daily Speed STD	s_1, s_2
# of Daily Distinct OD	n_1^{od},n_2^{od}
# of Daily Trips in DoW	n_1,..,n_7

Step (ii): Clustering Users into Groups based on Feature Vectors. Based on feature vectors, we cluster users into groups by a Gaussian Mixture Model (GMM) [2] as in Equation 1.

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \quad (1)$$

where x is a feature vector in our model, $\mathcal{N}$ is a Gaussian distribution with μ_k as the mean and Σ_k as the covariance matrix. We apply a standard Expect Maximization algorithm to maximize the likelihood iteratively. The output of the clustering gives the centroid μ of each cluster and the corresponding probability of x being in a cluster. We apply a Gaussian-based clustering method since Gaussian distributions are fit into mobility metrics in many scenarios [3].

Step (iii): Periodically Optimizing Clusters based on Davies-Bouldin Index. We use Davies-Bouldin index to tune the optimum number of user groups. Davies-Bouldin index measures both the separation of clusters and cohesion within clusters, which mathematically guarantees good clustering results. We found the optimum number of clusters is 135. The results are corresponding to 27 major pairs of home-work areas and 5 real-world driving groups, i.e., daily commuters, weekend users, weekday commuters, for-hire vehicle users, and others. Since users' behavior and roles change over time, e.g., a daily commuter can change to a Uber user, we apply the clustering method periodically to update the users' group information dynamically based on their most recent driving data.

The right part of Figure 7 gives an example of group tensors. With group tensors, we use the mobility patterns from similar users to predict future usage for existing users with limited data.

4.3 External Information Feeder

The external information feeder collects external information on trips, e.g., road networks and population density, as external features and then are fed to the learning component. External features include road type distribution and population distribution. We study road types and population density as external features since they have a significant influence on driving patterns based on our analysis in the following analysis. We incorporate these external features in our system given how often a user travels on a route with different road types and population density.

(a) Road Types: We divide all roads into 5 major types (i.e., highway, road, link road, path, and special road) based on road types provided by OpenStreetMap road networks [1]. First, we run map matching on personal and commercial vehicle datasets. Then, we calculate the average speed and driving distance on-road segments. We found a significant difference in the speed and distance distribution among different road types, which is considered as a context in our model. We omit the results due to space limitation.

(b) Population Density: We further investigate the impact of population density distribution on driving behaviors by calculating the correlation between speed and population in grids based on Worldpop dataset [7], which includes population distribution at night. We aggregate the average speed from 4 pm to 11 pm in grid partition to study the correlation. We found that over 80% grids show a negative correlation between population and speed, which motivates us to consider population density in our model.

4.4 Multi-Modal Multitask Learning

With both internal information and external information, we adopt a multi-modal LSTM (Long-Short-Term-Memory) in the recurrent neural network layer. This is because (i) both of the input and output of our usage can be seen as time-series data; (ii) LSTM is one of the most effective models to deal with time-series data prediction and is insensitive to temporal gaps[12]. We did not choose a more complicated model due to a practical deployment.

Multi-modal LSTM: A multi-modal Long Short Term Memory LSTM is designed to integrate multiple data sources with different weights [13] due to its insensitivity to temporal gaps. The previous work [15] has shown that the multi-modal LSTM model outperforms other models in time series prediction problems. The memory cell unit of Multi-modal LSTM is shown in Figure 8. The model can be described by Equations in Figure 8 where k indicates the modality (e.g., individual tensors and group tensors as two modalities) and $|K| = 2$ is the total number of modalities. Instead of merging heterogeneous data in the preprocessing step, a multimodal model shares weights across different types of modalities during the forward pass in the training process but does not share memory units [13]. W_{gh}, W_{ih}, W_{fh}, and W_{oh} are hidden layers' weights in the forward pass, which gives the features of sharing weights across modalities. The weights are initialized as random and are updated during the training. Instead, every modality keeps the memory unit $h_{t_1}^k$ in the forward pass. Therefore, it has some features to share weights but not memory units in the forward pass. In this work, as shown in the

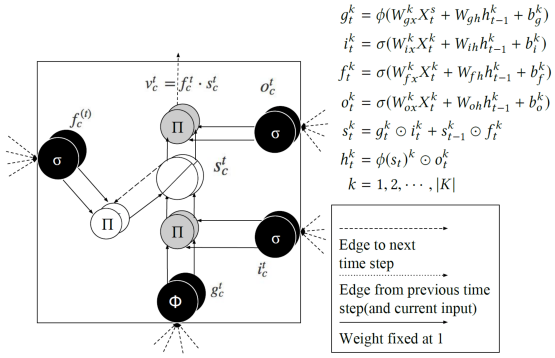

Fig. 8: Multi-modal LSTM Memory Cell

model overview of Figure 6, individual tensors and group tensors

are fed to two separate LSTM modules since we treat the individual tensors and group tensors as different modalities. Motivated by this idea, we integrate the individual mobility patterns with its corresponding group mobility patterns, but we keep the memory cell of individual mobility and group mobility. We separate the training data transferred from the internal information feeder into two categories of features, i.e., static features and usage features. Relying on external information, we extract contextual features, such as road type distribution and population density distribution. Finally, we feed usage features, static features, and contextual features in the learning component. The rest of the process is standardized after we format the problem this way, and the details of LSTM can be found at [12].

Input and Output: Given m days of previous usage data of vehicle i, i.e., travel distance, travel time, and speed STD, along with external contextual information, our target is to predict i's usage over next n days, the output is denoted as $\hat{\mathbf{Y}}^i = (\hat{Y}^i_{\tau+1}, \hat{Y}^i_{\tau+2}, \cdots, \hat{Y}^i_{\tau+n})$ where $\hat{Y}^i_\tau = (\hat{d}^i_\tau, \hat{t}^i_\tau, \hat{v}^i_\tau)$, τ is the current day; d^i_j, t^i_j, s^i_j is the daily distance, daily travel time and daily speed standard deviation to present the variance of the speed for vehicle i at day τ. In our design we make a daily prediction for fine-grained insurance policies. It is straightforward to obtain a n day distribution with our model with an adaptive method.

Loss Function: Since the dependency existing in the three predicted metrics, we apply a multi-task learning component with a loss function of the average of the three metrics to show the performance. We use MAPE which takes the average absolute error between the estimated value $\hat{p}$ and ground truth $\bar{p}$. $\epsilon(p) = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{p}_i - \bar{p}_i|}{\bar{p}_i}$. For each user, the travel distance, travel time and speed variance are closed correlated since all of them are derived from trips of users. Therefore, we apply a multitask learning model in the prediction to capture the underlying correlation among them. A joint loss function is normally defined in multitask learning models [28]. We define a joint loss function as the weighted MAPE of the three metrics in Equation 2, where d is the daily travel distance, t is the daily travel time, and v is the daily travel speed STD.

$$\begin{aligned} L &= \alpha \cdot \epsilon(d) + \beta \cdot \epsilon(t) + \gamma \cdot \epsilon(v) \\ s.t.\ &\alpha + \beta + \gamma = 1 \end{aligned} \tag{2}$$

We tune α, β and γ in our training process to achieve the best overall performance. The joint loss function is commonly used in multitask learning since training of one feature can benefit the learning of other two metrics and prevent overfitting in a single metrics learning [19]. To justify our design choice, we compare MoCha with a single task learning model where separate loss functions are defined for each metrics in our evaluation of Section 5.

5 EVALUATION

5.1 Methodology

(i) Setting: We utilize two kinds of vehicle data as shown in Figure 5, which contains nationwide long-term personal and commercial vehicle OBD data for the evaluation. The personal vehicle dataset has OBD data from 295 thousand vehicles; the commercial vehicle dataset has OBD data from 60 thousand vehicles. Both commercial data and personal data contain the exact time, location, and speed of the vehicles and an uploading device ID to identify each user. We train our model with 10-fold cross-validation (i.e.,90% days of data for training and 10% of days of data for testing) with both internal features and external features. We perform temporal cross-validation to gradually increase the number of continuous days as training data. The details are given in the evaluation results.

(ii) Metrics: We compare our predicted results $\hat{p}$ with real metrics $\bar{p}$ in terms of travel distance, time, and speed variance by Mean Absolute Percent Error (MAPE).

(iii) Baseline Approaches:

- *ARIMA:* An AutoRegressive Integrated Moving Average (ARIMA) auto-regression model is proposed in [18] to predict human behaviors with a set of auto-regression models.
- *DeepTransport (DT):* DeepTransport [17] is a state-of-art model to predict human mobility. It applies recurrent neural networks to predict human trajectories and travel time on the individual level. We adopt the model for distance, travel time, and speed STD prediction. In particular, we apply DT model on historical records and incorporate external information in the model.
- *MoCha-:* We implement multitask learning with a joint loss function to learn the underlying correlations among driving metrics and prevent overfitting in a single task learning model. In contract, we use a single task learning model with three individual loss functions to predict specific metrics, i.e., we drop multi-task learning from MoCha in this baseline model.

(v) Impacts of Factors: We evaluate four factors and their impacts on our system. (a) New Users vs Established Users: We separate prediction results into two groups. The first group is new users with historical records less than one week, i.e., belonging to evolving patterns. For all incoming UBI users, we use their first one-week data to predict their future metrics. After more data was collected about these users, they became established users. The prediction accuracy is compared in these two groups to show the ability of the model to capture patterns of new users. (b) Impact of Training Data Length: (c) Impact of Predicted Period: We use an adaptive way to predict future user metrics and evaluate the performance in a varied number of predicted days. (d) Impact of External Information.

(vi) Implementation: Our model and baseline models are implemented with Keras and Tensorflow libraries. We train and evaluate our work on 8 Nvidia K40C GPU servers. We set the learning rate as 0.001 and train the model with 100 epoch with cross-validations. We train our model with a previous one-week driving pattern as inputs and predict future daily metrics. We apply an adaptive learning method to predict long-term driving patterns, i.e., we use the predicted values as input to predict the further three metrics.

5.2 Evaluation Results

(i) Established Users vs. New Users: As the number of users in UBI is increasing, one of the most important challenges for insurance companies is usage prediction for new users. Therefore, before giving the general performance, we evaluate our models on the new users and established users. We report the overall performance, which is the average prediction error of travel distance, travel time, and travel speed STD in Figure 9. Our model shows the best performance compared with baselines, especially for new users. In particular, for new users, *MoCha* outperforms the existing model

ARIMA and DT by around 25.4 % by reducing the average MAPE from 30% to 22.6%. In detail, MoCha reduces MAPE of distance prediction from 26.1% to 18.8%, travel time prediction from 36.4% to 25.2%, and speed variance from 32.3% to 23.8%. Moreover, the period to update user groups is critical for the prediction for new users and evolving users. In the evaluation, we found we can reduce the prediction error by 12.4% by updating user groups daily compared with one-time clustering for new and evolving users.

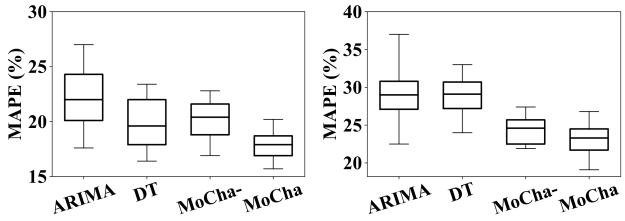

Fig. 9: Overall Performance

(ii) Travel Distance Prediction: We evaluate the prediction accuracy of the three investigated metrics, i.e., travel distance, travel time, and speed STD. First, we evaluate *MoCha* on distance prediction in Figure 10. *MoCha* presents a better and more stable performance compared with baselines in both personal and commercial vehicles. We found a higher performance gain in personal vehicles when comparing LSTM-based models (MoCha, DT, MoCha-) with ARIMA. One possible reason is that driving patterns of personal vehicles are more relevant to recent mobility since LSTM assigns a higher weight to short-term memories.

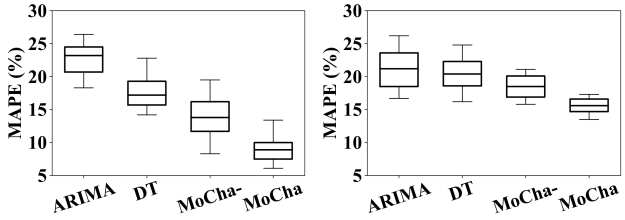

Fig. 10: Travel Distance

(iii) Travel Time Prediction: Compared with predicting travel distance, predicting travel time is more challenging since it is affected by external factors such as road traffic and travel start time while daily travel distance is mostly determined by OD (origin and destination) pairs and travel routes. As a result, we found a higher MAPE in Figure 11 compared with the travel distance prediction. We found a lower MAPE in MoCha, DT, and MoCha- compared with ARIMA in both users of personal vehicles and commercial vehicles. MoCha has the best performance, contributed by the prediction improvement on new drivers, and integration of external features. Besides, MoCha achieves a lower performance variance in personal vehicle users. Based on the historical data, we found over 30% of commercial vehicle users operate more than 20 hours per day. It may because some commercial vehicles are shared by more than one user in a rotation to maximize profits. In contrast, since personal vehicles have constant mobility patterns and fewer origins or destinations than commercial vehicles, personal vehicle users have a small variance on the performance compared with commercial vehicle users. We found the multitask learning component (MoCha-) and the multimodal training component (DT) have larger impacts on personal vehicle users. The possible reason is personal vehicle users have more constant mobility patterns. As a result, it is easier for MoCha to find similar drivers, prevent overfitting, and learn underlying correlations among metrics.

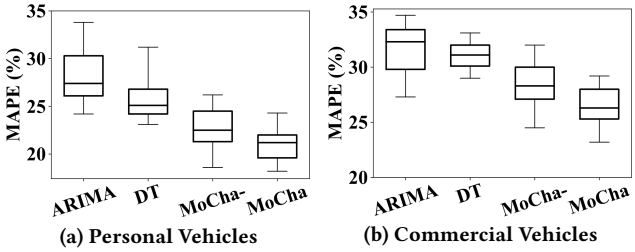

Fig. 11: Travel Time

(iv) Speed STD Prediction: Different from travel time and distance, speed is an instant value. We compare the estimated speed STD with the speed STD calculated based on the raw data from the OBD reader to evaluate the prediction performance. We found a different prediction performance during days of one week in Figure 12. Similar to travel distance and travel time prediction, MoCha- has a better performance than DT. In all four models, personal vehicles show a higher prediction error on weekends while commercial shows a lower error. The reason is that personal vehicle users have regular mobility patterns during weekdays and more random patterns on weekends. For commercial vehicle users, mobility patterns are not affected significantly by weekday patterns. On weekends, better traffic conditions lead to lower randomness on speeds.

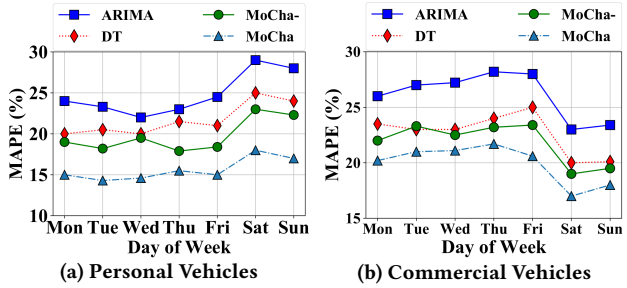

Fig. 12: Speed Variance

(v) Impact of Training Data Volume: We further study the impact of training data on the performance. We apply an $N + 1$ validation and evaluate the model by average MAPE of three predicted features. In the $N + 1$ validation, N continuous days of data are trained and 1 following day is tested. We evaluate N from 1 to 28 on both personal and commercial vehicle users. We fill null values for missing features. We study the impact of training data on overall performance in Figure 13. MoCha has the best performance in the average prediction errors. The elbows of the performance changes locate at around 14 days since it covered the day-of-week pattern.

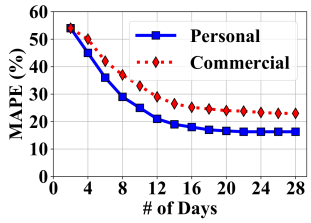

Fig. 13: Training Data

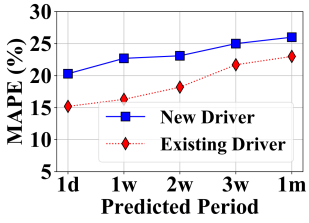

Fig. 14: Predicted Period

(v) Impact of Predicted Period: Since long-term driving pattern characterization is of great importance to quantify individual driving risks, we apply an adaptive method to predict individual driving patterns. Specifically, to predict daily driving distance, travel time, and speed variance in weeks or months, we use the predicted values as the input for new prediction. We predict different days of metrics ranging from one day to one month. We found the overall performance drops as the number of predicted days increases in Figure 14. This is caused by the accumulation of error in the prediction since we use the predicted data as the input of users.

(vii) Impact of External Features: We compare the impact of external information with hypothesis tests. We found (i) road information improves the average model performance by reducing MAPE from 18.93% to 17.86% for personal vehicle users and 25.88% to 24.67% for commercial vehicle users; (ii) Worldpop [7] population improves the average performance by reducing MAPE from 17.86% to 17.42% in personal vehicles and 24.67% to 24.05% for commercial vehicles. We omit the detailed results due to space limitations.

6 DEPLOYMENT: RISK PREDICTION

The insurance company we are working with is interested to know if our driving metric prediction has impacts on future driving risk prediction of drivers. It is straightforward to quantify driving risks by the probability of accidents happened to a user based on insurance claim data. We deploy our system for a pilot study to conduct a case study on 196 UBI users' claim data to predict their probability of accidents. As incentives, these users received additional discounts for further analyses so they consent their claim data sharing. We incorporate the results of MoCha in the prediction task to study how MoCha contributes to this real-world application.

Claim Data for Validation: We have access to claim data of 196 drivers from Shenzhen and their detailed OBD traces records for one year as ground truth for validation. The dataset contains four types of accidents based on their causalities, i.e., *collision*, *wading*, *collapse*, and *others*. According to claims, 93.14% of accidents are caused by collisions, 4.90% of accidents are caused by wading, 0.98% of accidents are *collapse*, which is caused by falling objects and 0.98% accidents are caused by other reasons.

Setup: We build two versions of a learning model X with (1) a logistic regression (i.e., X=LR), and (2) a neural network (i.e., X=NN) with sigmoid activation functions, to study how *MoCha* helps improve risk prediction in the two models. The input is the three metrics of a user (e.g., historical average or future prediction based on *MoCha*) with static factors such as gender and age along with if he/she has accidents before. The output is a value between 0 and 1, i.e., a higher value means a higher potential risk of the user. *Metrics*: Both loss function and evaluation metrics are defined by Mean Absolute Error (MAE) $\frac{\sum_{i=1}^{n}|\hat{p}_i - \bar{p}_i|}{n}$ between real risk $\bar{p}_i$ as ground truth (i.e., 0 without accident or 1 with accident obtained by ground truth) and predicted risk $\hat{p}_i$. We define the *MoCha* based method as *MoCha + X* and compare its performance with two baselines: *Hist+X* and *DT+X*. All *MoCha* and baselines use the same learning model X to learn the relationship between drivers' behavioral factors to their potential risk, but their behavioral inputs are different. *(i) Hist + X*: It uses the historical average of distance, time, and speed variance as the input without prediction; *(ii) DT + X* and *(iii) MoCha + X*: We use *DT* [17] and *MoCha* to predict the future driving distance, time, and speed variance as input for *X*, respectively.

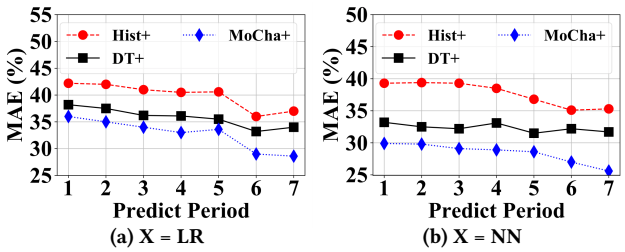

Fig. 15: Quantifying Future Driving Risk

Evaluation: Figure 15 shows the prediction of *MoCha* improves the performance in quantifying future risks of drivers in both LR and NN. The X-axis is the *MRE* and Y-axis is the length of the future days. As shown in Figure 15a, *MoCha* improves the performance by reducing the error from 38% to 29% in the logistic regression model. In Figure 15b, *MoCha* improves the performance by reducing the error from 36% to 26% in the neural network model. We found both prediction based methods have better performance than historical data based method Hist+. The reason is that historical data are biased when used to describe future driving patterns.

7 LESSONS LEARNED AND DISCUSSION

Key Lesson Learned: The most fundamental lesson learned in this we can predict mobility metrics for new UBI users with limited data and existing users with new patterns with high accuracy due to accurate driver group clustering design. The key insight is that the similarity between new users and existing users can be found by carefully designing a mobility feature set to quantify their similarity by periodically clustering. However, using group driving patterns alone cannot achieve the best performance due to overfitting, so we need to consider both the individual-level driving pattern and group driving pattern as two modalities and integrate them in a multi-modal learning model where these two modals interact with each other to improve the prediction accuracy. This insight provided some guidance on the cold start and user pattern evolving problem for current or future UBI companies.

Deployment Obstacles: Based on our result, the key obstacle for a large-scale deployment is that for the brand new users without any historical data, MoCha has the limited ability to predict their future mobility patterns. In general, their premium plan is determined based on their demographic information, e.g., gender and age, at the beginning of using UBI insurance. It makes challenging to convince the UBI company for a full-scale deployment. Our next step would be collecting enough historical data for the brand new users and then adjusting their premium according to the prediction of MoCha.

Privacy Protections and Consent: While modeling and predicting vehicle usage is important for insurance companies and we have UBI users' consent, we protected the privacy of involved users by using the aggregated metrics to model the driving patterns in *MoCha*. Therefore, we minimize the exposure risk for individual locations collected by the on-board GPS devices.

8 RELATED WORK

We study related work via two features, i.e., spatial scale and vehicle modality.

Aggregate Mobility vs. Individual Mobility We divide existing works on vehicle mobility into two categories based on the mobility level: (i) On the Aggregate Level, vehicle mobility is estimated by aggregating historical records without considering individual behaviors. Zhou *et al.* compare the speed estimation from either explicit or implicit sensing data [**?**]. Further, a few model calibration techniques have also been proposed to model travel speeds, e.g., offline calibrating based on sensitivity analyses [4]. (ii) On the individual level, personal behaviors and mobility patterns are taken into account for mobility modeling. Fang *et al.* propose a system called Mac to infer fine-grained travel time [5]. Song *et al.* propose a multi-task learning model based on historical trajectories to predict individual mobility such as transportation mode [17].

Single vs. Multiple Vehicle Fleets Due to the separation and isolation among vehicular fleets and transportation systems, most existing works have been focused on mobility on a single fleet such as taxi travel time estimation in Beijing [20]. Those works are well-designed for a single-vehicle fleet for vehicle mobility. However, due to the diversity of driving patterns in different fleets, the generalizability of such modes is not tested on other vehicular fleets. A few works were conducted on nationwide data for vehicle mobility. For instance, Zhang *et al.* propose a model to estimate the traffic volumes on major highways of China [27], but did not focus on predicting mobility behaviors.

Summary Technically, MoCha is different from the above works from two perspectives. (i) We focus on evolving issues in a UBI setting where new UBI users with limited data and established UBI users with long-term records, i.e., three years; whereas the existing works are mostly based on short-term data, e.g., a few days or months. (ii) we focus on modeling and prediction on multi-modality vehicle patterns, e.g., both commercial and personal vehicles; whereas the existing work is mostly focused on one modality.

9 CONCLUSIONS

In this work, we design, implement and evaluate a driving pattern characterization system called *MoCha* which models and predicts individual vehicle usage in the setting of usage-based insurance. We study driving patterns on three metrics, i.e., travel distance, travel time, speed variance. To solve the problem of data limitation of new drivers, we cluster existing drivers into groups based on their similarity of mobility patterns, and then combine individual driving patterns with group driving patterns based on a multi-modal multitask LSTM model. Our evaluation results show both good prediction results on driving behavior metrics and effectiveness on driving risk prediction based on real-world GPS and claim data.

10 ACKNOWLEDGEMENT

This work is partially supported by NSF 1849238, 1932223, 1951890, 1952096, and 2003874.

REFERENCES

[1] Open Street Map. In *http://www.openstreetmap.org/*.
[2] Chen, Y., and Krumm, J. Probabilistic modeling of traffic lanes from gps traces. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2010), GIS '10, ACM, pp. 81–88.
[3] Cho, E., Myers, S. A., and Leskovec, J. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 1082–1090.
[4] Ciuffo, B., and Lima Azevedo, C. A sensitivity-analysis-based approach for the calibration of traffic simulation models. *Intelligent Transportation Systems, IEEE Transactions on 15*, 3 (2014).
[5] Fang, Z., Yang, Y., Wang, S., Fu, B., Song, Z., Zhang, F., and Zhang, D. Mac: Measuring the impacts of anomalies on travel time of multiple transportation systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3*, 2 (2019), 1–24.
[6] Feng, J., Yang, Z., Xu, F., Yu, H., Wang, M., and Li, Y. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020), pp. 3426–3433.
[7] Golding, N., Burstein, R., Longbottom, J., Browne, A. J., Fullman, N., Osgood-Zimmerman, A., Earl, L., Bhatt, S., Cameron, E., Casey, D. C., et al. Mapping under-5 and neonatal mortality in africa, 2000–15: a baseline analysis for the sustainable development goals. *The Lancet 390*, 10108 (2017), 2171–2182.
[8] Händel, P., Ohlsson, J., Ohlsson, M., Skog, I., and Nygren, E. Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance. *IEEE Systems Journal 8*, 4 (2014), 1238–1248.
[9] Hosseinioun, S. V., Al-Osman, H., and El Saddik, A. Employing sensors and services fusion to detect and assess driving events. In *Multimedia (ISM), 2015 IEEE International Symposium on* (2015), IEEE, pp. 395–398.
[10] Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., and Willinger, W. Human mobility modeling at metropolitan scales. MobiSys '12.
[11] Jin, W., Deng, Y., Jiang, H., Xie, Q., Shen, W., and Han, W. Latent class analysis of accident risks in usage-based insurance: Evidence from beijing. *Accident Analysis & Prevention 115* (2018), 79–88.
[12] Lipton, Z. C., Berkowitz, J., and Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019* (2015).
[13] Papaioannou, S., Markham, A., and Trigoni, N. Tracking people in highly dynamic industrial environments. *IEEE Transactions on Mobile Computing (Issue: 99)* (2016).
[14] Ren, H., Pan, M., Li, Y., Zhou, X., and Luo, J. St-siamesenet: Spatio-temporal siamese networks for human mobility signature identification. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020), pp. 1306–1315.
[15] Ren, J. S., Hu, Y., Tai, Y.-W., Wang, C., Xu, L., Sun, W., and Yan, Q. Look, listen and learn-a multimodal lstm for speaker identification. In *AAAI* (2016), pp. 3581–3587.
[16] Skog, I., and Handel, P. In-car positioning and navigation technologies—a survey. *IEEE Transactions on Intelligent Transportation Systems 10*, 1 (2009), 4–21.
[17] Song, X., Kanasugi, H., and Shibasaki, R. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In *IJCAI* (2016), pp. 2618–2624.
[18] Song, X., Zhang, Q., Sekimoto, Y., and Shibasaki, R. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), KDD '14, ACM, pp. 5–14.
[19] Suresh, H., Gong, J. J., and Guttag, J. V. Learning tasks for multitask learning: Heterogenous patient populations in the icu. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), pp. 802–810.
[20] Wang, Y., Zheng, Y., and Xue, Y. Travel time estimation of a path using sparse trajectories. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 25–34.
[21] Warren, G., and Greenlee, M. Calculation of driver score based on vehicle operation for forward looking insurance premiums, Apr. 21 2006. US Patent App. 11/409,493.
[22] Weiss, J., and Smollik, J. Beginner's roadmap to working with driving behavior data. In *www.casact.org/pubs/forum/12wforumpt2/Weiss-Smollik.pdf* (2012).
[23] Wikipedia. Usage-based insurance — wikipedia, the free encyclopedia, 2017. [Online; accessed 7-December-2017].
[24] Xie, X., Yang, Y., Fang, Z., Wang, G., Zhang, F., Zhang, F., Liu, Y., and Zhang, D. cosense: Collaborative urban-scale vehicle sensing based on heterogeneous fleets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2*, 4 (2018), 1–25.
[25] Yang, Y., Xie, X., Fang, Z., Zhang, F., Wang, Y., and Zhang, D. Vemo: Enabling transparent vehicular mobility modeling at individual levels with full penetration. *arXiv preprint arXiv:1812.02780* (2018).
[26] Yuan, Z., Zhou, X., and Yang, T. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), pp. 984–992.
[27] Zhang, D., Zhang, F., and He, T. Multicalib: national-scale traffic model calibration in real time with multi-source incomplete data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2016), ACM, p. 19.
[28] Zhang, Y., and Yang, Q. A survey on multi-task learning, 2017.