1 Research Objectives

1.1 Introduction

The United States is undergoing massive transformation not only in terms of the transportation infrastructure and mobility offerings but also in terms of the travel behaviors. Thanks to the significant technology breakthroughs, Americans today have more travel options than ever before. With the widely adopted smartphones, travelers are able to easily access to real-time traffic information, up-to-date bus schedules as well as available rides through shared mobility options (i.e., Uber, Lyft) [1]. The technologies are not only reshaping how we travel but also are changing our transportation infrastructure as well as urban context. For example, the concept of smart cities have been receiving more and more attentions in recent years. The essence of smart city is to use information and communication to improve quality of life, energy usage and urban operation and service. More specifically, a smart city utilizes sensors, actuators and technology to connect various components across the city (i.e. building, street lights, and vehicles). Therefore, it generates and collects significant amounts of data in their daily operations [2] [3]. Given the improvement in big data technologies such as machine learning, cloud computing and databases, big data visualization, the large amount of data, if managed well, can be used by planners to better understand people's needs and thus improve efficiency of services. Besides of fully connected cities and transportation, autonomous vehicles are also been considered as a vital component in a smart city. For example, the Sidewalk Labs, an urban-tech focused subsidiary of Google parent company Alphabet, launched a project to transform a neighborhood, namely Quayside, located at the southeast of Downtown Toronto to the world's most innovative city neighborhoods. It is envisioned by the company that the neighborhood will be a place where pedestrian centered and the only vehicles are shared and self-driving; where buildings have no static use; where connectivity will be ubiquitous; where large scale data play a vital role for people's daily life and for operations and improvement of city services [4].

In addition to technology breakthroughs, America is also experiencing demographic transformation caused by a rapid increase in the elderly population. It is estimated that the aging boomers will lead to a double of the elderly populations (65 years old and above) in 2050 [5]. Meanwhile, millennials have already overtaken the baby boomers and have become the largest generation by proportion in the U.S. They also comprise the largest proportion of the labor force today [6]. Many studies have found that activity-travel behavior various across generations. For example, the aging boomers are found to have higher trip rates retiring later from the workforce and spending more active life styles, compared to their prior counterpart [5]. Millennials are also found to exhibit different activity-travel behaviors with making fewer trips and owning fewer vehicles [7]. These demographic shifts are of particular interest to transportation professionals because of the differences in the activity-travel behaviors of these different generations, and the unique needs they have in terms of the transportation system and the modal offerings. Their travel behaviors need to be carefully analyzed in order to provide appropriate services for their unique mobility needs.

These technological and demographic changes are altering the way people pursue activities and engage in travel. This in turn will have an impact on the transportation infrastructure through new and changing mobility needs. Advanced modeling and forecasting approaches are needed to better understand the activity travel behavior of the Americans under alternate scenarios. Subsequently, these insights can be used to quantitatively evaluate solutions and promote efficient transportation solutions and regulations. In this dissertation, three significant technological and demographic changes we are facing today were explored.

1. *Emergence of big data*
   Over the last few years, as with many other fields, the transportation discipline has also been swept by the big data revolution. The growth in data is from a variety of sources including traffic detectors, remote sensors, mobile devices, smart card data, global positioning system (GPS), and survey datasets among others [8]. This revolution has not only brought about tremendous opportunities for conducting interesting data driven analysis, it has also highlighted challenges associated with using traditional analytical methods to analyze these large datasets. These datasets can often be too large to be stored and analyzed using traditionally used approaches owing to memory limitations, computational runtimes, and scalability of methods among others. As a result, analyzing such datasets can be a significant challenge.

2. *Aging in population and demographic shift*
   The United States has experienced a dramatic demographic shift in recent years. The total number of the elderly population has increased from 35.1 million (12.4% of the total population) in 2000 to 52.5 million (14.5% of the total population) in 2018. It is projected that the number of the elderly population will reach 98 million by 2060 [9]. A key contributor to this trend is aging baby boomers – it is anticipated that nearly 10,000 baby boomers will turn 65 years old each day over the next 15 years [10]. The dramatic change in demographics will bring tremendous challenges to transportation planning and policy making.

3. *Emergence of new technology - autonomous vehicles*
   Autonomous vehicles (AV), also referred as driverless vehicles, are being touted as the future of mobility. AV technologies and solutions have been receiving increasing attention owing to their wide reaching potential for transforming movement of goods and services while also significantly improving safety. AVs are now being prototyped and deployed by many car manufacturers (i.e., Toyota, Tesla, Benz and Audi) and information technology companies (i.e., Google, Huawei) in recent years. Real-world tests and public pilots of AVs are being conducted across the US and elsewhere. For example, Waymo, a subsidiary of Google’s parent company Alphabet, announced that their self-driving fleet has already driven themselves for more than 10 million miles and tested on the real-world road conditions across varies locations of the U.S. including California, Arizona, Texas etc. They have also launched a public trial of fleet in Phoenix, Arizona, where the riders can use the self-driving car to go places where they frequently go everyday such as work, school [11]. Uber resumed their test on Pittsburgh on July 2018 after stopped all self-driving tests due to a fatal accident happened in Tempe, Arizona. Although the vehicles for testing are only operated in manual mode for now, the company says that it is the “first step” for resuming autonomous car tests in Pittsburgh [12]. nuTonomy, a local startup in Boston, are now allowed to conduct test-run of their self-driving cars on all of the roads in Boston [13]. nuTonomy also corporates with Lyft to offer self-driving ridership to passengers in Boston’s Seaport district [14]. More and more automakers, startups and tech companies are participating in the development of AVs and have announced that the AVs

will be commercially available within next a few years [15]. While the direct impacts of AVs are undisputed, the indirect and induced transportation implications of AVs are less understood.

1.2 Research Objectives

In an effort to address these research gaps, following are the specific objectives of this dissertation.

1. The first objective of this research is to develop a novel method for analyzing big data. More specifically, a Divide and Combine based approach to estimating Mixture Markov model so as to analyze large categorical time series data is developed. The validity of the proposed approach is demonstrated using a simulation study. The feasibility and applicability of the approach for analyzing large categorical time series data is highlighted by analyzing activity-travel patterns from multiple national household travel survey datasets. Individual's daily activity-travel behavior is characterized as categorical time series in order to incorporate multiple aspects of travel simultaneously. Subsequently, cluster compositions are explored to understand within and between cluster differences and their associations with generational cohort factors, socioeconomic attributes, and demographic variables.
2. The second objective of this research is to analyze activity-travel behavior of the elderly by incorporating multiple aspects of travel. Daily activity-travel pattern of the elderly will be represented as categorical time series. A time varying mixture Markov model will be applied to characterize the various types of travel behavior of the elderly by incorporating unobserved heterogeneity. Elderly population will be segmented into several subgroups based on their daily activity-travel pattern. Socioeconomic (i.e., income, employment status, car ownership), demographic (i.e., age, gender, generational cohorts) and land use (living in urban area, accessibility of public transit) attributes that associated with each type of traveler profile will also be explored by a multinomial regression analysis. Implications of transportation service provision and policy formulation will be discussed to better cater the various mobility needs of the elderly.
3. The third objective of this research is to enhance the activity pattern generation (APG) stage of existing tour-based activity-based (TABM) systems. The enhancement of the APG stage will be achieved by applying a set of multiple discrete continuous (MDC) models which allows to represent time as a continuous entity and also imposes more rigorous temporal constraints for activity pattern generation. The improved framework will be used to study AV future under different mobility configurations.

The rest of the prospectus is organized as follows: In next chapter, the first study focusing on a novel Divide and Combine based approach for analyzing large categorical time series data will be discussed. The application of the proposed approach for analysis of activity-travel pattern using multiyear travel survey data will also be presented. In the third chapter, the second study exploring activity-travel behavior of the elderly population using a time varying mixture Markov model will be presented. The fourth chapter demonstrates the exploration of MDC modeling approach in activity pattern generation and its applicability in TABM systems for studying impacts of AVs. The expected contributions and work plan of this research will be summarized in the final section.

2 A Novel Divide and Combine Based Approach to Estimating Mixture Markov Model for Large Categorical Time Series Data: An Application to Study of Clusters Using Multiyear Travel Survey Data

2.1 Background

Transportation, along with other fields, have been experiencing big data revolution in recent years. The growth of data is mainly due to the increasing in passively collected data (i.e., GPS data and mobile phone monitoring data), nevertheless, actively collected data are also increasing in size significantly. For example, the sample size of National Household Travel Survey (NHTS), which collects detailed travel behavior data of American residents has increased from 6,438 households and 17,382 persons in 1983 to 129,696 households and 264,234 persons in 2017 [16] [17]. In the next iteration of the NHTS, it is proposed that the survey data be augmented with passively collected data to carry out additional analysis – this will only add to the scale and complexity. Such improvement in travel survey data provides opportunity for uncovering new insights about patterns of activity-travel behaviors. In most big data analysis in the transportation domain, the focus has been on passively collected data [18] [19], to the authors' best knowledge, very limited studies have explored large actively collected survey data. The research presented in this study will contribute to this latter body of work attempting to analyze large actively collected survey data.

When analyzing large travel survey data, most analysis have limited their exploration to univariate or bivariate analysis of activity and travel engagement (e.g. trip rate, mode choice, activity type). Also, in many of these studies, the temporal dimension of activity-travel behavior (i.e. timing, duration, sequencing of trips and activities) are ignored. However, it is important to study multiple dimensions together (e.g. daily activity-travel engagement) while also considering the timing and schedule of engagement choices. One approach is to characterize individual's activity-travel behavior as a categorical time series. The commonly used attributes include number of activities, activity type, locations visited and average duration of activities and travel [20] [21] [22] [23] [24] [25] [26]. Categorical time series data structure can also be used for describing data other than travel-activity behavior such as driver behavior [27] and origin-destination matrices of transit route [28].

2.2 Literature Review

In the transportation literature, two methods are commonly used for analyzing categorical time series data: sequence alignment method [20] [21] [22] [23] [24] [25] and Markov model [26]. Sequence alignment method was originally developed for analyzing protein sequences in the biology field. The method compares pairs of sequences on the basis of their composition and sequencing orders, and assigns a score representing the degree of similarity or dissimilarity. Subsequently, the score is used to identify clusters. Despite the very intuitive approach, there are a number of limitations of this approach for analyzing large datasets. First, this approach is usually computationally very intensive because it requires comparing each pair of sequence. The computational runtime increases quadratically with sample size. Second, it requires assigning a distance measure for each pair of categories which is often arbitrary and impacts the results. In addition, most readily available software for sequence alignment are specifically designed for DNA or protein sequences. They require significant time and effort to be used with data from other fields.

Markov model is another widely used approach for analyzing categorical time series data. It has been applied in a number of fields such as speech recognition [29] [30] and life course trajectory analysis in social sciences [31]. Compared to sequence alignment, Markov models have some clear advantages. First, for a given datasets, the computational run times are much less compared to sequence alignment. Second, it doesn't require any arbitrary inputs such as the distance metric in sequence alignment. Third, there are number of modeling packages that are general enough that they can be readily used with any categorical time series data. Because of these advantages, in this analysis, Markov modeling approach was adopted. In particular, the Mixture Markov modeling approach was adopted so that in addition to characterizing the categorical time series, the clustering can be carried out using a single model system.

Most software implementing Markov models have been developed with small datasets in mind. To the authors' best knowledge, there are no model implementations that are able to handle large categorical time series data. One approach to overcome this limitation is to combine the traditional modeling technique with an approach for handling large datasets. Divide and Combine is one such strategy known for its effectiveness when analyzing large data. The basic idea of Divide and Combine strategy is to divide the large data into several manageable subsets. For each subset, traditional data analysis methods can then be applied. Subsequently, results from each subset are combined to obtain single model estimates for the full data. This approach has found applications in a number of disciplines [32] [33] [34] [35]. For example, in computer science field, the divide and combine strategy is used for developing efficient sorting [32] and information retrieval [33] algorithms.

2.3 Objective

The objective of this study is to develop a Divide and Combine based approach to estimating Mixture Markov model so as to analyze large categorical time series data. The validity of the proposed approach is demonstrated using a simulation study. The feasibility and applicability of the approach for analyzing large categorical time series data is highlighted by analyzing activity-travel patterns from multiple national household travel survey datasets. Individual's daily activity-travel behavior is characterized as categorical time series in order to incorporate multiple aspects of travel simultaneously. Subsequently, cluster compositions are explored to understand within and between cluster differences and their associations with generational cohort factors, socioeconomic attributes, and demographic variables.

2.4 Methodology

2.4.1 Markov Model

A Markov model is a stochastic model used to model randomly changing systems [36]. In a first-order Markov model, it is assumed that future state at time $t+1$ depends only on the current state at time t and not on any other events that occurred before it. Assume the categorical time series for observation i is represented by $X_{i,t}$ where i is an index for observations and can take values from $1, \ldots, N$ and t is an index for time and assumes values from $1, \ldots, T$. $X_{i,t}$ can take any value from a state space $S = \{1, \ldots, m\}$. A first-order Markov model can be described by two parameters $\theta = (\delta, q)$. δ is an m -dimensional vector of initial probabilities, and an element δ_j represents the probability that the categorical time series will be in a state j at the beginning i.e., $t = 1$. q is an $m \times m$ matrix of the transition probabilities, where an element $q_{j,h}$ indicates the probability of transitioning from state j to state h and is assumed to be the same for any time t. That is, a time homogenous Markov model is assumed. Equation (1) shows the probability expression for $q_{j,h}$.

$$q_{j,h} = P\left(X_{i,t} = h \middle| X_{i,t-1} = j\right). \tag{1}$$

The two parameters are usually estimated via maximum likelihood. The likelihood and log likelihood function are shown in Equations 2(a) and 2(b) respectively:

$$\mathrm{L}(\theta) = \prod_{i=1}^{N} P\left(X_{i,1} = x_{i,1} | \theta\right) \prod_{t=2}^{T} P\left(X_{i,t} = x_{i,t} \middle| X_{i,t-1} = x_{i,t-1}, \theta\right) \tag{2a}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \log\left(P\left(X_{i,1} = x_{i,1} | \theta\right)\right) + \sum_{i=1}^{N} \sum_{t=2}^{T} \log\left(P\left(X_{i,t} = x_{i,t} \middle| X_{i,t-1} = x_{i,t-1}, \theta\right)\right) \tag{2b}$$

where $x_{i,t}$ represents the observed state for observation i at time period t. The parameters thus estimated have neat closed form solutions. The estimator of initial probability is the same as the observed rate of the initial state:

$$\hat{\delta}_j = \frac{n_{j,1}}{N} \tag{3}$$

where $n_{j,1}$is the count of the observed series starting in state j at time $t = 1$. The estimator of the transition probability is the same as the observed transition rate:

$$\hat{q}_{j,h} = \frac{n_{j,h}}{\sum_{h=1}^{m} n_{j,h}} \tag{4}$$

where $n_{j,h}$is the count of observed occurrence of transitioning from state j to state h.

2.4.2 Mixture Markov Model

Clustering the categorical time series is an important analysis objective. One way to achieve this is by first fitting a Markov model to each categorical time series observation. The estimated parameters of the Markov models can then be used as features in the clustering algorithms (e.g., k-means, hierarchical clustering). However, this method can be time consuming since it involves fitting a large number of models, one for each of the N observations in the data set. An alternative approach is to use a Mixture Markov model wherein both Markov models characterizing the time series and clusters identifying similar time series, are both generated as part of the modeling results.

A Mixture Markov model incorporates a predefined number (K) of Markov models. Each model has its own set of parameters. It assumes that the time series of each observation is best represented by one of the Markov models. Assume we have a Mixture Markov model with K submodels M^k where $k = 1, \ldots, K$. Each M^k is a first-order Markov model with parameters $\theta^k = (\delta^k, q^k)$. For each time series $X_{i,t}$, $\alpha(k)$ denotes the priori probability that the observed time series follow the submodel M^k. The log likelihood function of the Mixture Markov model is written as:

$$\mathcal{L}(\theta, \alpha) = \sum_{i=1}^{N} \log\left[\sum_{k=1}^{K} \alpha(k)\, P\left(X_{i,t} = x_{i,t} \middle| M^k\right)\right] \tag{5}$$

where $P(X_{it}|M^k)$ is the probability that the time series x_{it} is generated by the submodel M^k. The probability is similar to the likelihood expression in Equation 2(b). Compared to the usual Markov model, the log-likelihood expression involves an additional summation over the K submodels weighted by the prior probability. The Markovian parameters for each of the submodels can be estimated using the expectation maximization algorithm. Additional details regarding the Mixture Markov modeling approach and estimation can be found in reference [37] [38]. The membership of an observed time series to a submodel is determined by the estimated $\alpha(k)$. Greater the value of $\alpha(k)$ higher is the likelihood that the observed time series follows the submodel M^k. In terms of the clustering discussion, the K submodels can be thought of as K clusters characterized by the corresponding Markov model and the observed time series belongs to the cluster with the highest $\alpha(k)$ value. Despite the appeal of providing both time series characterization and clustering results within a single model, most implementations of Mixture Markov models suffer from computational intractability when faced with a large dataset. Therefore, alternate estimation approaches are needed.

2.4.3 Divide and Combine

As noted above, when faced with large data, estimation of a Mixture Markov model is computationally intractable. An approach to overcome this limitation is to adopt the Divide and Combine approach [33]. The basic idea of Divide and Combine strategy is to divide the large data into several subsets each with a manageable number of observations. For each subset, a Mixture Markov model is fit. Subsequently, results from each subset will be combined to obtain single model estimates for the full data. The main contribution of this research is in development of a new Divide and Combine based approach for estimating a Mixture Markov model. The proposed approach is as follows:

1. Divide the large data set for N observations into a number of subsets with size N_s (where s is an index for the subset). N_s should be large enough to ensure that the properties of the estimators resulting from the EM algorithm are not compromised. Further, they may also be impacted by the limits of the Mixture Markov model estimation software and hardware configuration.
2. Fit a Mixture Markov model to each subset. Mixture Markov model estimation requires K to be predefined. However, the optimal K for a given subset is unknown. Therefore, multiple model estimates need to be carried out with different values of K to determine the optimal K. Model selection criterion such as Bayesian information criterion (BIC) can be used to select the optimal K. For the Mixture Markov model, the BIC is as shown in Equation (6).

$$BIC = -2\mathcal{L}(\theta, \alpha) + Dlog(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{i,t}) \tag{6}$$

   where $\mathcal{L}(\theta, \alpha)$ is computed using Equation (5), D is the number of estimated parameters.
3. Combine the estimation results from each subset to generate the model estimates for the entire data. Since the results from each Mixture Markov model include count of clusters and associated Markov model parameter estimates, a traditional clustering algorithm can be used to combine the results from each subset. The results from the subset can be thought of as "local" results based on a subset of data. The local results from all subsets will be

combined to obtain "global" results. As noted earlier, there are two parts to Mixture Markov model results namely: the clusters, and the Markov model parameters (i.e. initial probability and transition probability) associated with the clusters. First, to identify the clusters, the Markov Model parameters for each local cluster can be used as features to identify global clusters. In addition, the global cluster's identification can be traced back to the time series observations through the mapping to local clusters. Second, for each global cluster, Mixture Markov model parameters can be obtained by taking the average of Markov Model parameters of all local clusters that belong to it.

Any clustering method can be used at this step such as k-means and hierarchical agglomerative clustering. In the research presented in this paper, hierarchical clustering was chosen because of the sample size requirements associated with k-means. The hierarchical clustering approach does not have any sample size requirement whereas the k-means approach requires the sample size to be no less than 2^g(g=number of features for clustering) [39]. In this research, Euclidean distance was used when applying hierarchical clustering. The final output of the hierarchical clustering method is a tree diagram called as a dendrogram. The height of the dendrogram indicates the dissimilarity between clusters. A larger value indicates higher dissimilarity between clusters. Typically the number of clusters is determined by cutting the tree diagram at a certain height. There is no systematic process for identifying the threshold height. One can obtain different clustering solutions based on the choice of threshold height. In this research, the three criterions were utilized to determine the threshold height: 1) each cluster should show a distinct pattern, 2) each cluster should include a reasonable number of observations, and 3) splitting the cluster further does not reveal any new patterns [40].

The proposed Divide and Combine approach was implemented in R. The 'seqHMM' package [37] was used for fitting Mixture Markov model on each subset and 'cluster' package was used for hierarchical agglomerative clustering [41].

2.5 Results and Contributions

A new Divide and Combine based approach to estimating Mixture Markov models is presented which can overcome computational issues associated with large categorical time series data. The approach can be used to analyze categorical time series data to not only characterize the time series process but also to identify clusters in the data. The proposed approach was validated using a simulation study. Further, the feasibility and applicability of approach was demonstrated using a case study aimed at analyzing activity-travel patterns using multiyear household travel survey data.

In simulation study, three groups of categorical time series were generated by three Markov models. The first Markov model has distinct initial and transition probabilities while the second and the third models having similarly initial and transition probabilities. In other words, the categorical time series generated by first Markov model are more distinct compared to that generated by the other two models. The proposed approach were able to successfully classify the time series generated by the first Markov model with no misclassification. For the time series generated by the second and third Markov models, the proposed method were able to distinguish most of them with a small misclassification rate of 3%.

In the case study, the proposed approach was applied to identify clusters of activity-travel patterns. The composition of clusters and their evolution over time were analyzed to understand changes in travel patterns during the last thirty years. Three types of travelers namely night discretionary travelers (ND), work and home travelers (WH), more in-home travelers (MH) were identified. The results suggested that changes in activity-travel patterns of millennials are different from those of boomers and generation X. For boomers and generation X, the share of individuals who fall into ND type has increased and the percentage of them fall into MH has decreased. The opposite effect was true for millennials. In an effort to disentangle potential confounding factors results were analyzed as a function of age, employment status and gender. It was found that in addition to generation effects, other factors may also be at play in explaining the changes in travel patterns.

3 Analyzing Travel Behavior of the Elderly Using Time Varying Mixture Markov Model

3.1 Background

Demographic changes, with respect to the great increase in the elderly population, have a tremendous implication in transportation planning and policy making, especially because of that the elderly are found to exhibit more complex and diverse travel behavior and mobility needs [42] [43] [44] [45]. Specific transportation facilities and services are required to serve the mobility needs of the growing elderly population. There are several transportation issues need to be considered. First, the mobility needs of the elderly population show more diversities compared with the working age population. The elderly have more free time for making various types of out-of-home activities (i.e., social activities, health care) [42]. Second, the travel behavior of the elderly are not monolithic, instead it can be differ in socioeconomic characteristics (i.e., income, health condition, household structure, gender) and psychographic factors (i.e., mobility-related attitudes) [5] [42] [43] [44] [45] [46] [47] [48]. Third, the socioeconomic characteristics, lifestyle and mobility-related attitudes of the elderly population are changing. The aging baby boomers, who will form the majority of the elderly population in the near future, have been found to have higher educational level and higher income compared to their counterparts. They also exhibit distinct travel behavior. For example, the aging boomers are found to have higher trip rates retiring later from the workforce and spending more active life styles, compared to their prior counterpart [47]. Such changes in the elderly population imply that the historical assumptions and traditional solutions might not be able to well serve mobility needs of the aging population. A close look at the activity travel behavior of the elderly population is needed in order to prepare efficient transportation solutions.

3.2 Literature Review

Studies that examine travel activity behavior of the elderly often conclude that the travel behavior varies significantly within the elderly population. Some studies focus on the age effects on travel behavior of the elderly population. The elderly population can be split into younger elderly (aged 65 to 75 years) and the older elderly (over 75 years). It was found that both younger and older elderly travelers prefer to use automobile for traveling either as a driver or passenger. Though declining as the increase in age, the overall trip rates of the younger elderly are not significantly lower than younger population (age under 65 years old). However, the older elderly population were found to have a lower trip rate mainly due to physical limitations [45] [46] [47]. Besides of physical condition, Georggi and Pendyala [46] also pointed that the socioeconomic characteristics of the younger and older elderly differ significantly. In addition, the elderly women were found to have the lowest trip rate compared to the elderly men [47]. The low travel demand of the elderly women is related to income effect instead of a reduced desire for travel. Their travel demand is expected to increase since the currently aging women are more resemble as men in terms of income level, driving experience, education and professional accomplishment etc.

There are some studies focusing on the effect of the socioeconomic characteristics on travel behavior of the elderly [42] [43] [48]. Hildebrand [43] categorized the elderly into six lifestyle groups based on socioeconomic factors and found that travel behavior of the elderly differs significantly across the six lifestyle groups. For example, it was found that the elder females who live alone with valid driver license tend to spend more time for shopping activities, while the elder workers tend to participate the least in maintenance and social activities. Kim and Ulfarsson [42] also addressed the difference in travel behavior of working elderly and non-working elderly. They conducted their analysis focusing on the effects of socioeconomic factors and trip characteristics on mode choice of the non-working elderly. It was found that the elderly are more likely to use transit if living within five blocks of a bus stop. They also prefer walking for recreational and personal trips.

Existing studies have yielded limited understanding for the overall travel profile of the elder population. First, most of them have focused on a single aspect of travel such as mode choice, trip distance and trip rate. To the author's best knowledge, there is no study examining the travel patterns of the elderly by analyzing multiple attributes of activity-travel behavior, especially the sequence order and timing of activities and travels. Second, although heterogeneities in travel behavior of the elderly have been address in most of the studies, the segmentation of heterogeneous subgroup of elderly population is largely based on their socio-demographic characteristics (age, employment status etc.). A few studies have proposed to characterize the elderly population into different groups based on their mobility-related attitude (i.e., perceived car stress, perceived mobility necessities) [44]. None of them directly use the activity-travel patterns of the elderly for segmentation. A differentiation of subgroups of elderly is required in order to develop new services and improve the existing system to cater varies mobility needs. As described in section 2.1 and 2.2, characterizing activity-travel behavior as a categorical time series allows us to incorporate multiple aspects of activity-travel simultaneously, such characterization will also be applied for this study. Moreover, a more advanced Markovian type model will be explored for analyzing activity-travel behavior of the elderly.

3.3 Objective

The objective of this study is to analyze activity-travel behavior of the elderly by incorporating multiple attributes of travel. Daily activity-travel pattern of the elders will be represented as categorical time series. A time varying mixture Markov model will be applied to characterize the travel behavior. The model can incorporate the unobserved heterogeneity in travel behavior and time effect into consideration. Elderly population will be characterized into several subgroups based on their daily activity-travel pattern. Socioeconomic and demographical attributes that associated with each type of activity-travel behavior will be explored by a multinomial regression analysis. Implications of transportation service provision and policy formulation will be discussed in order to better service the various mobility needs of the elderly.

3.4 Methodology

3.4.1 Time Varying Mixture Markov Model

As described in section 2.3.1, a first-order Markov model can be described by two parameters $\theta = (\delta, q)$, where δ is the initial probability and q is the transition probability. For a time homogenous Markov model, it is assumed that both δ, q is constant over time. It might not be appropriate for analyzing activity-travel behavior since the probabilities of switching between different types of activities are often sensitive to the time of the day. In order to overcome this issue, a finite mixture time varying Markov model proposed by Marutti and Rocci [49] will be adopted for this study. Assume the categorical time series for observation i is represented by $Y_{i,t}$ where i is an index for observations and can take values from $1, \dots, N$ and t is an index for time and assumes values from $0, \dots, T$. $Y_{i,t}$ can take any value from a state space $S = \{1, \dots, M\}$. Marutti and Rocci specified the transition probabilities of the Markov model as a logit-type model shown as equation (7), where the unobserved heterogeneity is introduced as a random effect $\boldsymbol{c_{ijk}}$ and time effects are introduced as a set of time-related exogenous variables $\boldsymbol{x_{it}}$. The basic idea is to assume that among the N categorical time series, G latent groups or clusters are present, whereby all the subjects within each group are characterized by a non-homogenous Markov model with both fixed and group-specific parameters [49].

$$\boldsymbol{q_{itjk}} = \frac{\exp(x_{it}{}'\gamma_{jk}+c_{ijk})}{1+\sum_{h:h\neq j}^{m}\exp(x_{it}{}'\gamma_{jh}+c_{ijh})} \quad \forall \boldsymbol{j},\ \boldsymbol{k} \in \boldsymbol{M} \tag{7}$$

The initial probabilities are specified as equation (8)

$$\boldsymbol{\delta_{ij}} = \frac{\exp(c_{ij0})}{1+\sum_{h=1}^{m-1}\exp(c_{ih0})} \quad \forall \boldsymbol{j} \in \boldsymbol{M} \tag{8}$$

Where

- q_{itjk} is the transition probability of subject i visits state k at time t when s/he was in state j at time $t-1$
- δ_{ij} is the initial probability of being in state j at time 0 for person i
- $x_{it} = (x_{it1}, \dots, x_{itp})$ is a vector of exogenous variables for individual i at time t
- $\gamma_{jk} = (\gamma_{jk1}, \dots, \gamma_{jkp})$ is a vector of fixed regression coefficients; $j, k \in (1, \dots, m)\ and\ j \neq k$
- c_{ijk} is the group-specific random effects; $j, k \in (1, \dots, m)\ and\ j \neq k$
- For all $i = 1, \dots, n$, constrains $c_{im0} = 0$, $\gamma_{jk} = 0\ and\ c_{ijk} = 0$ if $\mathrm{j} = k$ for $\mathrm{j}, k = 1, \dots, m$
- The distribution of $c_{ijk} = \left(c_{i10}, \dots, c_{i(m-1)0}, c_{i21}, \dots, c_{i2(m-1)}, \dots, c_{im1}, \dots, c_{im(m-1)}\right) \in \mathbb{R}^{m^2-1}$ as a realization of multivariate discrete random variable C takes value of $C = \{c_1, \dots, c_g, \dots, c_G\}$ with probabilities $\Pi = \{\pi_1, \dots, \pi_g, \dots, \pi_G\}$. c_g is a vector of intercepts with dimension of $m^2 - 1$

The likelihood for N independent subjects can be expressed as equation (9):

$$\log\left(\mathrm{L}(\Theta|S)\right) = \log\left\{\prod_{i=1}^{N}\sum_{g=1}^{G}\pi_g\sum_{s_{i0}\in M}\cdots\sum_{s_{iT}\in M}\left[\Pr\left(s_{i0}|c_g\right)\times\prod_{t=1}^{T}\Pr(s_{it}|s_{it-1}, c_g, x_{it})\right]\right\}$$

$$= \sum_{i=1}^{N}\log\left(\sum_{g=1}^{G}\pi_g\sum_{s_{i0}\in M}\cdots\sum_{s_{iT}\in M}\left[\Pr\left(s_{i0}|c_g\right)\times\prod_{t=1}^{T}\Pr(s_{it}|s_{it-1}, c_g, x_{it})\right]\right) \tag{9}$$

Here, the initial and transition probabilities, $\Pr\left(s_{i0}|c_g\right)$ and $\Pr\left(s_{it}|s_{it-1}, c_g, x_{it}\right)$, denote the Markov parameters for the gth component of the finite mixture. The log-likelihood function (equation (10)) can be considered as the marginalization of the complete log-likelihood:

$$\log L_c(\Theta) = \sum_{i=1}^{n}\sum_{g=1}^{G}\eta_{ig}\,log\pi_g + \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j\in M}\eta_{ig}v_{i0j}\,log\delta_{jg}$$

$$+\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{j\in M}\sum_{k\in M}\eta_{ig}\mu_{itjk}\,logq_{itjkg} \tag{10}$$

Where

- $\eta_{ig} = I(C_i = c_g)$: 1 if and only if the subject is generated by the gth component of the finite mixture
- $\mu_{itjk} = \mathrm{I}(S_{it} = k, S_{it-1} = \mathrm{j})$: 1 if and only if moved from j state at t-1 to k state at t
- $v_{i0j} = \mathrm{I}(S_{i0} = j)$: 1 if and only if initial state is j

The expectation maximization (EM) algorithm will be applied for model estimation. The EM algorithm is based on the complete log-likelihood (equation (10)). The E step consist of computing the conditional expectation of those functions of the missing data ($\eta_{ig}, \mu_{itjk}, v_{i0j}$) that appear in the complete-data log-likelihood given the observations and given the current estimation of Θ. In the M step the log-likelihood is maximized with respect to the parameters.

Once identified the subgroups of the elderly population based on their differences in activity-travel behavior, a multinomial logistic regression model will be estimated to explore the attributes that contribute to each type of travel profile. Three types of variables will be examined including 1) demographical variables (i.e., age, gender, generational cohorts); 2) socioeconomic variables (i.e., income, employment status, education level); 3) land use (i.e., living in urban area, accessibility of public transportation).

3.5 Expected Results and Contributions

It is expected that this study will demonstrate a new approach to analyzing activity-travel pattern of the elderly travelers by considering multiple attributes simultaneously. By incorporating unobserved heterogeneity as random effects and introducing time effects as exogenous variables, the time-varying mixture Markov model will help to identify homogenous subgroups of the elderly travelers with respect to their activity travel behavior. Demographical, socioeconomic and land use characteristics of each subgroups will be explored. The results of this study will help decision makers have better understanding of various travel needs for the elderly population.

4 Development of Continuous Time, Temporally Constrained and Behaviorally Consistent Tour Pattern Generation System for Modeling the Impacts of Autonomous Vehicle Future

4.1 Background

Over the last a few years, the AV technology and solutions have made significant progress in bringing the futuristic fantasy to nearby reality. As many major players announced that the AVs will be available on the market within next a few years, there are still a lot of uncertainties associated with AVs remaining for exploration. Optimists of AVs claim that AVs can be beneficial in various fields. For mobility aspect, AVs hold the promise to improve mobility for many groups such as the non-drivers, the elderly and people with travel-restrictive medical condition [50] [51]. With AVs, these groups of people can be well served either though shared autonomous vehicles (SAV) or privately owned AVs. In addition to the underserved populations, the mobility of other groups of people could also be increased due to the ease of travel and reduced travel cost by AV usage [52]. For the safety aspect, AVs are expected to improve traffic safety since they require much less reaction time when encountered with accident compared with human drivers. More importantly, it was found that most of the fatal traffic accident happened in the U.S. are caused by human errors such as alcohol impairment, speeding and drug usage [53], AVs are expected to significantly reduce crashes by avoiding human errors [54] [55] [56]. For congestion and traffic operations, AVs are expected to efficiently reduce road congestion due to their traffic-flow-smoothing capabilities, vehicle to vehicle communication technology (V2V) and smarter routing strategies [57] [58]. For land use, AVs are expected to free up a large scale of urban areas where are currently been used as parking space. AVs are be able to relocate themselves either to find a free parking space outside the urban core or proceed to serve others if offered as a shared mode [59]. On the other hand, skeptics and opponents of AVs expressed concerns to the claimed benefits. AVs can potentially lead to a dramatic increase in overall vehicle miles traveled (VMT) either due to new demand generated by the underserved population or by the existing drivers. With AVs, driving is no longer a tedious task, instead, it is a journey where people can perform various types of activities such as working, entertainment and sleeping [60]. Nevertheless, though AVs can park themselves far away from dense urban area and free up urban space, it may also result in a significant increase in zero-occupancy vehicle miles traveled. With that, any congestion and mobility benefits may be substantially or entirely offset [55]. Moreover, it was also found that most of the benefits of AVs are claimed under the assumption that AVs are operated as a shared mode instead of privately owned. Although there are a lot of uncertainties associated with AV implementation, their impacts on travel behavior are undeniable. Analytical tools or approaches are in need to evaluate the impacts of AVs under different policies and service types.

4.2 Literature Review

Several methods have been applied by researchers to study the potential implications of AV technologies. In general, the studies can be classified into four categories including: 1) speculative studies [61] [62]; 2) analytical and simulation based studies [51] [63] [64] [65] [66]; 3) survey based studies [67] as well as 4) virtual reality or simulator based studies [68] [69].

Speculative studies tend to convey possibilities of AVs implications based on information and data from existing modes. For example, studies have found that the emergency of shared ride, such as Uber and Lyft, changes car ownership patterns. They speculate such change will also be observed caused by implementation of SAVs [62]. Analytical and simulation based studies use the pre-defined assumptions to simulate the individual's travel behavior and network operations under the different AV implementation scenarios [51] [63] [64] [65] [66]. Meyer et al., [66] simulated impact of AVs on accessibility of Swiss municipalities use Swiss national transport model. They assumed 80% to 270% increase in highway capacity and 40% increase in urban road capacity due to AV implementation. They also assumed that new vehicle demands are generated because of new users substitute other modes to AVs. Based on their simulation result, they pointed that overall AVs can provide increase in accessibilities. However, when demand increased drastically, the accessibility decreases. Survey studies design questionnaires for collecting data from individuals to analyze different aspect of AVs. The commonly asked questions include willing to pay, mode choice, and ownership etc., In order to study the adoption of AVs with distinction between shared and owned vehicles, Krueger et al., [67]distributed an online state preference mode choice survey to 435 residents in Australia, they found that young individuals and individuals with multimodal travel patterns may be more likely to adopt SAVs. Virtual reality or simulator based approach overcomes issue of lack of realism of the stated preference survey approach by developing highly realistic environment of the AV future. Farooq et al. [69] developed a virtual immersive reality environment platform for conducting a range of stated preference experiments in a highly realistic, immersive, interactive environment. They used this platform to explore pedestrian acceptance of autonomous vehicles and associate infrastructure changes in urban setting. By comparing the result with the result from a text-only and a visual animation survey, they found that the experiment tools have significant impact on the result. The virtual immersive reality environment has a positive impact on the respondent's perceptions of autonomous vehicles.

The proposed method of this study contributes to the analytical and simulation approach by enhancing existing tour-based activity-based model (TABM) systems to allow the modeling of individual's time use and travel behavior under different AV service offerings. Activity-based model (ABM) systems have been gaining significant research attention in recent years due to their behaviorally accurate representation of individual activity-travel pattern under various policies and planning applications. Unlike traditional trip-based travel demand model systems that predict aggregate-level (i.e., TAZ level) travel demand for long-term socio-economic scenarios, the ABM systems focus on modeling various aspects of disaggregate-level (i.e., individual-level, household-level) activity-travel pattern impacted by short-term demand management policies such as congestion pricing and singe occupant vehicle regulation [70].

Tour based ABM systems use tour as analysis unit. The systems subdivide individual's daily activity-travel schedule into a set of tours. TABM represents the most advanced state of the practice of ABMs. These systems have been widely adopted by the transportation agencies and authorities (i.e., Sacramento Area Council of Governments (SACOG), Denver Regional COG (DRCOG), Metropolitan Transportation Commission at San Francisco, CA (MTC))[71]. A tour is defined as a sequence of trips starting and ending at home or work anchor location. If a tour is anchored at home it is defined as a home-based tour, while if it is anchored at work location it is defined as a work-based subtour. Individual activity-travel pattern can be characterized as a set of tours with each tour consists of a primary destination and a series of intermediate stops either before or after the primary destination. In the state-of-the-art tour-based modeling approaches, daily activity-travel patterns of the decision makers are formed in two stages, namely, the activity pattern generation (APG) and the activity scheduling (AS). The APG is the identification of characteristics of all tours including tour purpose, number of stops within a tour, purpose, mode and destination for each stop and time allocated to all tours and stops among other decisions. The AS stage models the timing and placement of tours stops within a day. The advantage of TABM systems compared to traditional trip based modeling systems is that they treat time as an all-encompassing continuous entity while the latter treat time as a simply "cost" for making trips [72]. However, due to the ease of applicability, almost all of the TABM systems in practice today represent time in discrete units [73]. Second, most tour-based model systems do not explicitly acknowledge the temporal constraints when modeling tours, or when modeling stops within a tour. Temporal constraints are often accommodated afterwards using heuristics and logical checks at the activity scheduling stage. To this end, the proposed method of this study contributes to the enhancement of the APG stage of the existing TABM framework by addressing the above mentioned two limitations.

4.3 Objective

The primary objective of this study is to enhance the APG stage of existing TABM. Multiple Discrete-Continuous (MDC) modelling approach formulated by Bhat [72] will be applied for modeling daily tour engagement and time allocation by treating time as a continuous entity. The models will be implemented to an existing TABM framework named San Francisco Chained Activity Model Process (SF-CHAMP). The enhanced framework and model formulation are demonstrated using data from the 2010-2012 California Household Travel Survey. Results from the enhanced framework and original SF-CHAMP will be compared to indicate the feasibility of the enhanced framework. In evaluating AVs, a scenario based study will be conducted using the same data. The different AV technologies and solutions will be evaluated by modeling scenarios representing different types of AVs (i.e., SAVs, privately owned AVs).

4.4 Methodology

4.4.1 Enhanced APG Framework

The APG stage of SF-CHAMP is achieved by DaySim which is an activity-based travel simulator developed by Bowman and Ben-Akiva [74]. The APG stage of DaySim first determines the main pattern of travel for each individual. A multinomial logistic regression (MNL) is applied to jointly estimate the participation of home-based tour a person undertakes during a day for seven purposes, and the occurrence of additional stops during the day for the same seven purposes. After this, another MNL is applied to determine the exact number of tours that an individual pursuits for a given purpose. For each tour, the exact number of intermediate stops and their purposes are estimated through a MNL. The outcomes of the model is strongly conditional on main pattern of travel predicted at the first step. For the last modeled tour, the model is constrained to accomplish all intermediate stops purposes predicted at the first step. The timing of tour and stop are estimated through time-of-day models which determines tour primary destination arrival and departure time, and intermediate stop arrival or departure time. The time-of-day models are also a set of MNLs which determine the combination of arrival and departure time by dividing a day into 48 time slots with 30-minute interval. It can be seen that the existing APG stage is based on a large number of independent MNLs and discrete time unit. The proposed study will focus on improvement on four dimensions of the existing APG stage, namely, 1) the choice of participation in different types of home based tours, 2) time allocation to each tour, 3) the choice of participation in different intermediate stops within tour and 4) time allocation to each stop. Other dimensions of individual's APG such as destination choice of stops, mode choice of trips and sequencing of stops within a tour etc. will be modeled using existing approaches.

4.4.2 MDCEV

MDCEV is a utility theory-based model for discrete/continuous choice that derived and formulated by Bhat [72]. The major difference of the MDCEV model framework in activity participation, in contrast to the standard discrete choice model, is that it assumes that the alternatives imperfectly substitutable for each other. This assumption leads a multiple discreteness model which allows to the simultaneously choose of multiple activities and allocate time to each chosen activity given a resource constraint. The activity generation stage of SF-CHAMP will be replaced by MDCEVs. Instead of sequentially modeling number of tours a person undertakes and the time allocated to them, the MDCEV directly determines a series of home based tours that an individual pursuits as well as the time allocation to each tour. Moreover, it can also determines time allocation for in-home activities (i.e., time allocation at home before the first trip of the before, time allocation at home after returned from the last trip of the day, time allocation at home between out-of-home activities). After generating tour pattern for each individual, the intermediate stop participation and time allocation within each tour are also modeled by MDCEV.

Following Bhat, The MDCEV model assumes a translated non-linear additive specification for the utility function. Assume there are K different home based tour purposes that an individual can choose to allocate time to. Let t_j be the time allocated to tour purpose j $(j = 1,2,3, \dots, K)$. The utility accrued to the individual is specified as the sum of the utilities obtained from investing time in each tour purpose, which can be expressed as equation (11):

$$U = \sum_{j=1}^{K} \psi(x_j)(t_j + \gamma_j)^{\alpha_j} \quad ; \quad (0 < \alpha_j \le 1) \tag{11}$$

$$\psi(x_j, \varepsilon_j) = \exp(\beta' x_j + \varepsilon_j) \tag{12}$$

Where $\psi(x_j)$ is the baseline utility for time allocated to activity purpose j; To ensure that utility is greater than 0, ψ is defined as an exponential function of observed characteristics x_j and unobserved characteristics ε_j associated with activity purpose j, where ε_j is assumed to be independent and identically type I extreme value distributed across tour purposes and individuals. The utility function belongs to the family of translated utility functions, where γ_j determines the translation and α_j influences the rate of diminishing in marginal utility of allocating time in activity purpose j. Based on utility maximization theory, individual is seeking to maximize the sum of the utilities subject to the time budget constraint T:

$$\max U = \sum_{j=1}^{K} [\exp(\beta' x_j + \varepsilon_j)(t_j + \gamma_j)^{\alpha_j}] \ ; \ (0 < \alpha_j \le 1) \tag{13}$$

Subject to:

$$\sum_{j=1}^{K} t_j = T$$

The optimal time allocations can be found by applying the Lagrangian function. The same model framework will also be applied for intermediate stop participation and time allocation within a tour where the time budget is the allocated time for that tour. In this framework, time is treated as a continuous entity thus allocations of time to tours and stops are in continuous time units. The temporal constraints are incorporated in the model framework. First, total time allocation across all tours and in-home activities is constraint to the total time available in a day (i.e., 1440 minutes). Second, total time allocation across all stops within a tour is equal to the time allocation for the said tour [72].

4.5 Expected Results and Contributions

The study will help enhance the APG stage of existing TABM. A memo describing the approach used to integrate the enhancements into an existing TABM will be prepared. The enhanced TABM will be made available under open-source licensing agreements. The feasibility and applicability of the enhanced TABM will be demonstrated by comparing the travel outcomes and varies computational measures from the enhanced TABM against that from the existing TABM. The enhanced TABM will be applied to study the impacts of different AV service offers. Travel behavior impacts of such offers will be analyzed. The findings from the research will help both public and private industry chart the way for developing solutions and policies that meet mobility, livability, and sustainability needs of urban cities and regions around the US.

5 Summary and Work Plan

This section summarizes expected contributions from this study. Current research progress is presented, and also the work plan for the remaining tasks are described.

5.1 Anticipated Contributions

This research contributes substantively to the exploration of modeling and forecasting methodologies for activity-travel behavior analysis. Three statistical approaches are proposed to address the impacts on travel behavior and activity engagement by the tremendous technological and demographical changes in recent years, namely big data revolution, aging in population and AV future. Three contributions can be expected in this research:

1. The first study of the research is expected to offer a feasible and applicable method for analyzing the emerging big data in transportation domain. With respect to the methodological contribution, the proposed method provides an efficient way for analyzing large categorical time series data by a clustering analysis. Validity of the method is demonstrated by a simulation study and the applicability is demonstrated by a case study for analyzing daily activity-travel pattern using multiyear NHTS data. With respect to activity-travel behavior analysis, the proposed method contributes to identify three traveler types based on their unique activity-travel behavior. By exploring the generational cohort, socioeconomic and demographic composition of each traveler type, results provide evidence in support of recent claims about different generational cohorts and their activity-travel behaviors. It is hoped this method can be applied to other large transportation related dataset to help answer interesting research questions.
2. The second study is anticipated to demonstrate an approach for analyzing activity-travel behavior of the elderly by considering multiple aspects of traveling simultaneously. On the methodological front, the study will apply an advanced time-varying mixture Markov model to incorporate time effects and unobserved heterogeneities in activity-travel behavior. This model framework eliminates the strong homogeneity assumption of Markov model thus offers more accurate results. With respect to activity-travel behavior insights, results from the research are expected to provide complete profiles of the elder travelers. Heterogeneities in activity-travel behavior of the elderly can be used to segment the elderly into subgroups. Subgroup compositions are explored to understand within and between group differences and their associations with generational cohort factors, socioeconomic attributes, and demographic variables. Findings of this study is expected to help the decision makers to provide efficient solutions to service varies mobility needs.
3. The third study of this research is expected to develop a rigorous modeling system to simulate the potential changes on individual's travel behavior and time use in the AV future. On the methodological front, MDC modeling approach will be applied wherein activity engagement and time allocation are both generated as part of the modeling results. The new framework will be developed to address two main limitations of existing TABM frameworks, namely, characterizing time as a continuous entity and accommodating temporal constraints within which an individual schedules their activity engagement. With respect to activity-travel behavior analysis, the impacts on travel behavior under different AV service offerings scenarios will be examined. Multiple travel outcomes will be compared to evaluate different AV service offering. Subsequently, these insights can be used to quantitatively promote efficient transportation solutions and regulations that meet mobility, livability, and sustainability needs of urban cities and regions around the U.S.

5.2 Study Status and Work Plan

The study is still in progress. It is anticipated to be finished by June 2019. The following part describes the status of the three studies included in this prospectus and remaining tasks.

1. The first study of the prospectus has been completed and submitted to 2019 Transportation Research Board (TRB) annual meeting. It is been accepted for presentation and is also under consideration for publication [75].
2. The second study of the prospectus is still in progress. A derivative of this work have been invited for publication as a book chapter [76]. Literature review has been completed. Model estimation has been implemented using R. There are two remaining tasks for this study:
   a. Task 1.1: A simulation study will be carried out to validate the performance of time-varying mixture Markov model.
   b. Task 1.2: The model will be applied to study activity-travel behavior of the elderly using 2017 National Household Travel Survey data. A MNL will be built to identify sociodemographic attributes that related to different types of activity-travel behavior of the elderly.
3. The third study of the prospectus is also in progress. Data for model estimation and validation have been processed. A brief descriptive analysis has been conducted in order to identify alternatives for MDC models. There are four remaining tasks for this study:
   a. Task 2.1: A set of MDC models will be estimated and validated using 2010-2012 California Household Travel Survey data.
   b. Task 2.2: MDC models will be integrated to the DaySim framework by replacing a large number of MNL models for activity-pattern generation.
   c. Task 2.3: To indicate the feasibility of the enhanced framework, a comparison of the results from enhanced framework and from the original framework will be conducted
   d. Task 2.4: The enhanced framework will be used for analyzing the impacts of AVs on daily travel behavior. A number of scenarios will be analyzed to explore the implication of AVs under different service offerings.

Table 1 below provides an overview of the schedule for the remaining tasks. A duration of seven months (from December 1, 2018 through June 30, 2019) is anticipated. The timing and the duration of the various tasks are highlighted in the table below. The due date for each task is identified with "X".

Table 1 Work Schedule

Task	Dec	Jan	Feb	Mar	Apr	May	June
Task 1.1	X						
Task 1.2		X					
Task 2.1			X				
Task 2.2					X		
Task 2.3						X	
Task 2.4							X

The results of the proposed research will be presented at different professional conferences like Transportation Research Board annual meeting and International Association for Travel Behaviour Research (IATBR) conference. The final research outputs will be targeted for publications in academic journals and books, such as *Transportation Research Part B: Methodological* and *Mapping the Travel Behavior Genome.*

Reference

[1] "Examining the Future of Mobiity and Its Energy Impacts," Office of Energy Eficiency & Renewable Energy, 13 January 2017. [Online]. Available: https://www.energy.gov/eere/articles/examining-future-mobility-and-its-energy-impacts. [Accessed 25 October 2018].

[2] "Smart cities: acceleration, technology, cases and evolutions in the smart city.," i-scoop, [Online]. Available: https://www.i-scoop.eu/smart-cities-smart-city/. [Accessed 25 October 2018].

[3] T. Maddox, "Smart cities: 6 essential technologies.," TechRepulic, 1 August 2016. [Online]. Available: https://www.techrepublic.com/article/smart-cities-6-essential-technologies/. [Accessed 25 October 2018].

[4] "Project Vision," [Online]. Available: http://www.passivehousecanada.com/wp-content/uploads/2017/12/TO-Sidewalk-Labs-Vision-Sections-of-RFP-Submission-sm.pdf. [Accessed 25 October 2018].

[5] S. Bush, "Forecasting 65+ travel: An integration of cohort analysis and travel demand," in *84th Transportation Research Board Annual Meeting*, Washington, DC, 2005.

[6] A. Caumont and C. D'vera, 31 March 2016. [Online]. Available: http://www.pewresearch.org/fact-tank/2016/03/31/10-demographic-trends-that-are-shaping-the-u-s-and-the-world/. [Accessed 25 October 2018].

[7] N. C. McDonald, "Are millennials really the “go-nowhere” generation?," *Journal of the American Planning Association,* vol. 81, no. 2, pp. 91-103, 2015.

[8] M. Clarke, "Big Data in Transportation," Institution of Engineering and Technology Transport Sector Insight, [Online]. Available: file:///C:/Users/jiz13007/Downloads/IET%20Sector%20Insight%20Transport%20(1).pdf. [Accessed 20 October 2018].

[9] "US census factfinder," [Online]. Available: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_DP05&src=pt. [Accessed 20 October 2018].

[10] "The Growth of the U.S. Aging Population," SeniorCare, [Online]. Available: https://www.seniorcare.com/featured/aging-america/. [Accessed 20 October 2018].

[11] "On the Road," Waymo, [Online]. Available: https://waymo.com/ontheroad/. [Accessed 10 November 2018].

[12] N. Ingraham, "Uber resumes self-driving car tests, but only in manual mode:Tests have resumed in Pittsburgh, but there's no timeline for when they'll resume autonomous driving," Engadget, 24 July 2018. [Online]. Available: https://www.engadget.com/2018/07/24/uber-self-driving-tests-pittsburgh-manual-mode/. [Accessed 10 November 2018].

[13] Z. Enwemeka, "NuTonomy Can Now Move Its Self-Driving Cars On Any Road In Boston," Bostonomix, 20 June 2018. [Online]. Available: http://www.wbur.org/bostonomix/2018/06/20/nutonomy-self-driving-car-boston-expansion. [Accessed 20 November 2018].

[14] "nuTonomy and Lyft Launch Boston Self-Driving Pilot," nuTonomy, [Online]. Available: https://www.nutonomy.com/company-news/nutonomy-and-lyft-launch-boston-self-driving-pilot/. [Accessed 10 November 2018].

[15] L. Ciferri, "BMW unveils futuristic self-driving concept," Automotive News Europe, 7 March 2016. [Online]. Available: http://europe.autonews.com/article/20160307/ANE/160309896/bmw-unveils-futuristic-self-driving-concept?cciid=email-ane-daily. [Accessed 10 November 2018].

[16] "1983 Nationwide Personal Transportation Survey.," [Online]. Available: https://www.fhwa.dot.gov/ohim/1983/1983page.htm.. [Accessed 27 July 2018].

[17] "2017 National Household Travel Survey," [Online]. Available: https://nhts.ornl.gov. [Accessed 27 July 2018].

[18] J. G. M. C. W. P. S. T. M. G. a. B. A. L. Candia, "Uncovering individual and collective human dynamics from mobile phone records," *Journal of physics A: mathematical and theoretical,* vol. 41, no. 22, p. 224015, 2008.

[19] M.-p. Kwan, "Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems: a methodological exploration with a large data set," *Transportation Research Part C: Emerging Technologies,* vol. 8, no. 1-6, pp. 185-203, 2000.

[20] C.-H. Joh, T. Arentze and H. Timmermans, "Pattern recognition in complex activity travel patterns: comparison of Euclidean distance, signal-processing theoretical, and multidimensional sequence alignment methods," *Transportation Research Record: Journal of the Transportation Research Board* , no. 1752, pp. 16-22, 2001.

[21] C. Wilson, "Activity patterns of Canadian women: Application of ClustalG sequence alignment software," *Transportation Research Record: Journal of the Transportation Research Board,* no. 1777, pp. 55-67, 2001.

[22] W. W. Recker, M. G. Mcnally and G. S. Root, "Travel/Activity Analysis: Pattern Recognition, Classification, and Interpretation," *Transportation Research A,* vol. 19, no. 4, p. 279–296, 1985.

[23] E. I. Pas, "Weekly Travel-Activity Behavior," *Transportation,* vol. 15, no. 1-2, pp. 89-109, 1988.

[24] N. Shoval and M. Isaascon, "Sequence alignment as a method for human activity analysis in space and time," *Annals of the Association of American geographers,* vol. 97, no. 2, pp. 282-297, 2007.

[25] A. Zhang, Kang, J. E. Kang and K. K. C. Axhausen, "Multi-day activity-travel pattern sampling based on single-day data," *Transportation Research Part C: Emerging Technologies,* vol. 89, pp. 96-112, 2018.

[26] K. G. Goulias, "Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models," *Transportation Research Part B: Methodological,* vol. 33, no. 8, pp. 535-558, 1999.

[27] A. Sathyanarayana, P. Boyraz and J. H. Hansen, "Driver behavior analysis and route recognition by hidden Markov models," in *2008 IEEE International Conference on Vehicular Electronics and Safety*, Columbus, 2008.

[28] B. Li, "Markov models for Bayesian analysis about transit route origin–destination matrices," *Transportation Research Part B: Methodological,,* vol. 43, no. 3, pp. 301-310, 2009.

[29] B. Schuller, G. Rigoll and M. Lang, "Hidden Markov model-based speech emotion recognition," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hongkong, China, 2003.

[30] M. Gales and S. Young, "The application of hidden Markov models in speech recognition," *Foundations and Trends® in Signal Processing,* vol. 1, no. 3, pp. 195-304, 2008.

[31] N. Barban and F. C. Billari, "Classifying life course trajectories: a comparison of latent class and sequence analysis," *Journal of the Royal Statistical Society: Series C (Applied Statistics),* vol. 61, no. 5, pp. 765-784, 2012.

[32] N. Satish, M. Harris and M. Garland, "Designing efficient sorting algorithms for manycore GPUs," in *2009 IEEE International Symposium on Parallel & Distributed Processing*, Rome, Italy, 2009.

[33] J. Chen and Y. Saad, "Divide and conquer strategies for effective information retrieval," in *SIAM International Conference on Data Mining.Society for Industrial and Applied Mathematics*, 2009.

[34] M. Khalilian, F. Z. Boroujeni, N. Mustapha and N. M. Sulaiman, "K-means divide and conquer clustering," in *International Conference on Computer and Automation Engineering*, Bangkok, Thailand, 2009.

[35] X. Chen and M.-g. Xie, "A split-and-conquer approach for analysis of extraordinarily large data," *Statistica Sinica,* vol. 24, no. 4, pp. 1655-1684, 2014.

[36] P. A. Gagniuc, Markov Chains: From Theory to Implementation and Experimentation. John Wiley & Sons., John Wiley & Sons, 2017.

[37] S. Helske and J. Helske, "Mixture Hidden Markov Models for Sequence Data: The seqHMM Package in R," 3 April 2017. [Online]. Available: https://arxiv.org/abs/1704.00543. [Accessed 20 July 2018].

[38] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *International Computer Science Institute,* vol. 4, no. 510, p. 126, 1998.

[39] S. Dolnicar, "A review of unquestioned standards in using cluster analysis for data-driven market segmentation," *Faculty of Commerce Papers. University of Wollongong, Australia,* p. 273, 2002.

[40] G. M. Downs and J. M. Barnard, "Hierarchical and non-Hierarchical Clustering," 1 December 1995. [Online]. Available: http://www.daylight.com/meetings/mug96/barnard/E-MUG95.html. [Accessed 10 July 2018].

[41] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, J. Gonzalez and K. Kozlowski, "Package ‘cluster’," 9 April 2018. [Online]. Available: https://cran.r-project.org/web/packages/cluster/cluster.pdf. [Accessed 10 July 2018].

[42] S. Kim and G. Ulfarsson, "Travel mode choice of the elderly: effects of personal, household, neighborhood, and trip characteristics.," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1894, pp. 117-126, 2004.

[43] E. D. Hildebrand, "Dimensions in elderly travel behaviour: A simplified activity-based model using lifestyle clusters," *Transportation,* vol. 30, no. 3, p. 285–306, 2003.

[44] S. Haustein, "Mobility behavior of the elderly: an attitude-based segmentation approach for a heterogeneous target group," *Transportation,* vol. 39, no. 6, p. 1079–1103, 2012.

[45] R. Alsnih and D. A. Hensher, "The mobility and accessibility expectations of seniors in an aging population," *Transportation Research Part A: Policy and Practice,* vol. 37, no. 10, pp. 903-916, 2003.

[46] N. L. Georggi and R. M. Pendyala, "Analysis of Long-Distance Travel Behavior of the Elderly and Low Income," in *Transportation Research Board Conference, Personal Travel: The Long and Short of It*, Washington, DC, 1999.

[47] S. Rosenbloom, "Sustainability and automobility among the elderly: An international assessment," *Transportation,* vol. 28, no. 4, pp. 375-408, 2001.

[48] J. C. Cornman, D. Glei, G. Rodríguez, N. Goldman, B.-S. Hurng and M. Weinstein, "Demographic and socioeconomic status differences in perceptions of difficulty with mobility in late life," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* vol. 66, no. 2, pp. 237-248, 2010.

[49] A. Maruotti and R. Roberto, "A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption," *Statistics in Medicine,* vol. 31, no. 9, pp. 871-886, 2012.

[50] C. D. Harper, C. T. Hendrickson, S. Mangones and C. Samarasa, "Estimating potential increases in travel with autonomous vehicles for the non-driving, elderly and people with travel-restrictive medical conditions," *Transportation research part C: emerging technologies,* vol. 72, pp. 1-9, 2016.

[51] D. J. Fagnant and K. M. Kockelman , "Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas," *Transportation,* vol. 45, pp. 1-16, 2016.

[52] J. M. Anderson, N. Kalra, K. D. Stanley, P. Sorensen, C. Samaras and O. A. Oluwatola, Autonomous vehicle technology: A guide for policymakers. Rand Corporation., Santa Monica: RAND Corporation, 2014.

[53] N. H. S. Administration, "2017 motor vehicle crashes: overview.," U.S. Department of Transportation, Washington, DC, 2018.

[54] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice,* vol. 77, pp. 167-181, 2015.

[55] E. Guerra, "Planning for Cars That Drive Themselves: Metropolitan Planning Organizations, Regional Transportation Plans, and Autonomous Vehicles," *Journal of Planning Education and Research,* vol. 36, no. 2, pp. 210-224, 2015.

[56] G. Silberg, "Self-Driving Cars: The Next Revolution," KPMG LLP, Delaware, 2012.

[57] S. E. Shladover, D. Su and X.-Y. Lu, "Impacts of Cooperative Adaptive Cruise Control on Freeway Traffic Flow," *Transportation Research Record,* vol. 2324, no. 1, pp. 63-70, 2012.

[58] P. Tientrakool, Y.-C. Ho and N. F. Maxemchuk, "Highway Capacity Benefits from Using Vehicle-to-Vehicle Communication and Sensors for Collision Avoidance," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*, San Francisco, 2011.

[59] T. Snyder, "How the Self-Driving Car Could Spell the End of Parking Craters," Streetsblog USA, 26 March 2014. [Online]. Available: https://usa.streetsblog.org/2014/03/26/how-the-self-driving-car-could-spell-the-end-of-parking-craters/. [Accessed 13 June 2018].

[60] C. Mims, "Driverless Cars to Fuel Suburban Sprawl: research suggests advances in transportation technology contribute to bigger and more sprawling cities.," The Wall Street Journal, 20 June 2016. [Online]. Available: https://www.wsj.com/articles/driverless-cars-to-fuel-suburban-sprawl-1466395201. [Accessed 2018 13 June].

[61] T. Litman, "Autonomous Vehicle Implementation Predictions: Implications for Transport Planning," Victoria Transport Policy Institute, 2013. [Online]. Available: https://www.vtpi.org/avip.pdf. [Accessed 2018 10 June].

[62] Y. Freemark, "Will autonomous cars change the role and value of public transportation," The Transportation Politic, 23 June 2015. [Online]. Available: https://www.thetransportpolitic.com/2015/06/23/will-autonomous-cars-change-the-role-and-value-of-public-transportation/. [Accessed 13 June 2018].

[63] J. Zachariah, J. Gao, A. Kornhauser and T. Mufti, "Uncongested Mobility for All: A Proposal for an Area Wide Autonomous Taxi System in New Jersey," in *Transportation Research Board 93rd Annual Meeting*, Washington DC, 2014.

[64] M. W. Levin and S. D. Boyles, "Effects of Autonomous Vehicle Ownership on Trip, Mode, and Route Choice," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2493, pp. 29-38, 2015.

[65] S. L. Bozorg and S. M. Ali, *Potential Implication of Automated Vehicle Technologies on Travel Behavior and System Modeling*, 2016.

[66] J. Meyer, H. Becker, P. M. Bösch and K. W. Axhausen, "Autonomous vehicles: The next jump in accessibilities?," *Research in Transportation Economics,* vol. 62, pp. 80-91, 2017.

[67] R. Krueger, T. H. Rashidi and J. M. Rose, "Preferences for shared autonomous vehicles," *Transportation Research Part C: Emerging Technologies,* vol. 69, pp. 343-355, 2016.

[68] T. Al-Shihabi and R. R. Mourant, "Toward more realistic driving behavior models for autonomous vehicles in driving simulators," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1843, pp. 41-49, 2003.

[69] B. Farooq, E. Cherchi and A. Sobhani, "Virtual immersive reality for stated preference travel behavior experiments: a case study of autonomous vehicles on urban roads," *Transportation Research Record: Jornal of the Transportation Research Board,* 2018.

[70] A. R. Pinjari and C. R. Bhat, "Activity-based Travel Demand Analysis," in *A Handbook of Transport Economics*, Edward Elgar Publishing, 2011.

[71] J. Evans, "Short-Term Trip-Based Model Strategy Implementation Plan," 11 November 2015. [Online]. Available: http://www1.mwcog.org/uploads/committee-documents/a1xfVlle20151125084532.pdf. [Accessed 10 November 2018].

[72] C. R. Bhat, "A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions," *Transportation Research Part B: Methodological,* vol. 39, no. 8, pp. 679-707, 2005.

[73] C. R. Bhat and F. S. Koppelman, "A retrospective and prospective survey of time-use research," *Transportation,* vol. 26, no. 2, pp. 119-139, 1999.

[74] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transportation Research Part A: Policy and Practice,* vol. 35, no. 1, pp. 1-28, 2001.

[75] J. Zhang and K. C. Konduri, "A Novel Divide and Combine based Approach to Estimating Mixture Markov Model for Large Categorical Time Series Data: A Application to Study of Clusters Using Multiyear Travel Survey Data," in *Transportation Research Board Annual Meeting*, Washington, D.C., 2019.

[76] J. Zhang and K. C. Konduri, "Explore Activity-Travel Behavior of the Elderly Population," in *Mapping the Travel Behavior Genome*, Elsevier, 2019.