
*Present*

# 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

Over the last few years, as with many other fields, the transportation discipline has also been swept by the big data revolution. 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. To this end, this paper proposes a new Divide and Combine based approach to estimating Mixture Markov models for analyzing large categorical time series data. The validity of this approach is demonstrated using a simulation study. Further, the feasibility and applicability is highlighted by conducting a clustering analysis of large activity-travel sequences using multiyear travel survey datasets. In the case study, each individual's daily activity-travel behavior is characterized as a categorical time series that attempts to capture multiple aspects of travel and activity engagement simultaneously while also incorporating the timing and the schedule of different episodes. The proposed Divide and Combine based Mixture Markov models are then used to cluster the large data. Subsequently, cluster compositions are explored to understand within and between cluster differences and their associations with generational cohort factors, socioeconomic attributes, and demographic variables. As a preliminary exploration, the results suggest that travel patterns of individuals over the last three decades can be categorized into three types of travel patterns. Results also provide evidence in support of recent claims about different generational cohorts and their activity-travel behaviors.

**Speaker:** Jingyue Zhang, Transportation Graduate Student

AND

# SYNTHESIZING HOUSEHOLD AND PERSON-LEVEL ATTRIBUTES JOINTLY FOR INDIVIDUAL GEOGRAPHIES USING HIDDEN MARKOV MODEL

In order to apply microsimulation based models of land use and travel demand, socio-economic and demographic attributes about all individuals in a region is required. This disaggregate level information is not readily available and people resort to population synthesis procedures. These procedures combine readily available information in the form of sample data and marginal distributions to generate the required inputs. In this study, a simulation-based technique for population synthesis using a Hidden Markov Model (HMM) framework is presented. An important feature of the proposed approach is the ability to generate more heterogeneous synthetic households and persons. The proposed simulation-based approach is demonstrated using a case study for Connecticut. Synthetic population is generated for two block groups in Connecticut under alternate configurations. A comparative analysis is carried out to highlight the feasibility and applicability of the proposed approach in generating consistent multilevel agents while adhering to geography-based heterogeneity. The current work is similar in spirit to other recent simulation-based generators, however, there are two important contributions. First, a hierarchical transition structure is proposed in the HMM-based model, to capture the dependencies across household and person-level attributes. Thus the procedure ensures that both household and person level attributes are controlled simultaneously. Second, the transition matrices are estimated at the geography level incorporating the sample as well as marginal information available. This helps synthesize populations that are more accurate and consistent with available information.

**Speaker:** Amit Mondal, Transportation Graduate Student

## Monday, November 05, 2018
## 12:20 – 1:10 PM
## ITE Room 322