Activity Chain Generation with Dynamic object-oriented Bayesian Networks

J. Urata[a,*], Y. Mochizuki[b] and E. Hato[b]

[a] University of Tsukuba, Tsukuba, Japan
urata.junji.gf@u.tsukuba.ac.jp
[b] The University of Tokyo, Tokyo, Japan
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 27, 2024

Keywords: Activity simulation, Bayesian Network, Transport Demand

1 INTRODUCTION

With the rise in computing power, the potential for using activity models to replicate individual movements has grown. Recent progress in studying activity simulators using Bayesian Networks, referred to as BN-ACT, has been reported(Joubert & De Waal (2020), Sallard & Balać (2023), De Waal & Joubert (2022)). BN-ACT enables data-driven assessments of variable impacts and correlations, offering the notable advantage of constructing the model structure without analyst input.

Conversely, as the number of variables increases, the potential network structures multiply, and the cost of searching for the optimal structure escalates. Furthermore, when generating activity chains, the sequence of activities must be determined. Including the sequence of activities as output targets of the Bayesian Network (BN) complicates the search for the network structure. This study introduces dynamic objects into BN-ACT to reduce structural search costs and represent the sequence of activities. Additionally, BN-ACT faces issues such as overlapping times between multiple activities and trips. To address this, we propose algorithms for generating validated trip chains. We test the effects of these proposed dynamic objects and trip chain generation on improving prediction accuracy using data from the Person-Trip survey in the Tokyo metropolitan area.

2 METHODOLOGY

2.1 Dynamic object-oriented Bayesian Networks

The Object-Oriented Bayesian Network (OOBN) is an extension of the BN that incorporates objects (Koller (1997)). An object in this context is a collection of variables that can store multiple values. This arrangement helps restrict the network structure among variables from different objects, thus reducing the number of candidates in a search for the optimal Bayesian Network structure. Weber & Jouffe (2006) introduced Dynamic OOBN to manage temporally dependent variables, allowing connections only between consecutive time-related variables. In attempts to replicate activities, such as in the work by De Waal & Joubert (2022), this time-sliced approach is used to establish dependencies within similar time series. However, they assume that the network structure remains constant across all time periods and is predetermined externally.

This study endogenously considers dependencies among variables within dynamic objects for structural BN estimation. Variables related to a single activity are grouped into one dynamic object, while correlations between different dynamic objects are only made with variables from the subsequent activity. Additionally, it's assumed that variables related to individuals can affect all activity-related variables; therefore, these agent-related variables are not included in the dynamic objects. These methods allow for more flexible dependencies among variables and improve predictability. At the same time, they simplify the search for optimal BNs and support the creation of explainable BNs.

2.2 Validated Activity Chain Generation

We propose a generation algorithm to ensure the activity chains produced by the BN are valid. In our BN structure estimation process, each activity is handled using a dynamic object. The total number of activities is determined as one of the BN's agent variables. Our generation algorithm then creates activities to meet this specified total. Additionally, activity chains generated by BN-ACT may have overlapping activity and travel times, leading to implausible sequences. To resolve this, we use activity resampling to address the issue of overlapping times. If the sum of the start time, duration time of the i-th activity, and travel time to the next activity exceeds the start time of the subsequent activity, it is identified as a time constraint violation. When this happens, we resample all activities from the $i{+}1$-th activity onwards. Both algorithms proposed for generating validated activity chains have also been implemented in existing agent-based activity simulators (ex. Roorda *et al.* (2008)).

3 RESULTS

3.1 Data and Settings

To validate our proposed approach, we use real data from the 6th Tokyo Person Trip Survey (referred to as PT-Data), conducted in the Greater Tokyo Area in 2018. This dataset represents approximately 1% of the population in the region, providing detailed information on individual movements over the course of a day, collected through survey forms. The dataset includes a total of 32 variables: 14 related to household and personal attributes assigned to each individual, and 18 related to trip characteristics assigned to each movement.

The case study utilizes 13 variables in constructing the BN. Dynamic objects, which represent individual trips and activities, are defined by seven variables: trip purpose, destination type, number of facilities in the destination zone, trip start time, travel time, activity duration, and primary mode of transportation. Each of these variables is divided into four to five categories for use in the BN. Six agent variables that influence all trips and activities are used: age, gender, employment type, driver's license possession, vehicle availability, and number of activities. The number of activities is capped at five, while the other variables are divided into two to three categories for use. According to PT-Data, less than 1% of individuals in the sample engage in more than five activities per day. By setting an upper limit of five activities and incorporating dynamic objects, the Bayesian Network is built using a total of 41 variables, calculated as (7 variables per activity $\times$ 5 activities) + 6 agent variables. Approximately 300,000 individuals exhibited these behaviors, with 80% of the data used for training and 20% used for validation in testing. To demonstrate the effectiveness of the proposed model, we use an existing activity model (Joubert & De Waal (2020)), a type of Bayesian Network (BN) without object introduction, as a comparative model.

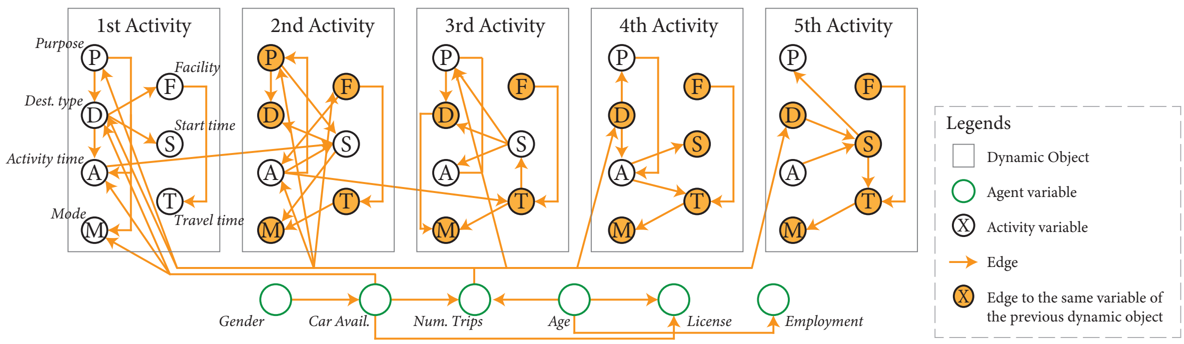
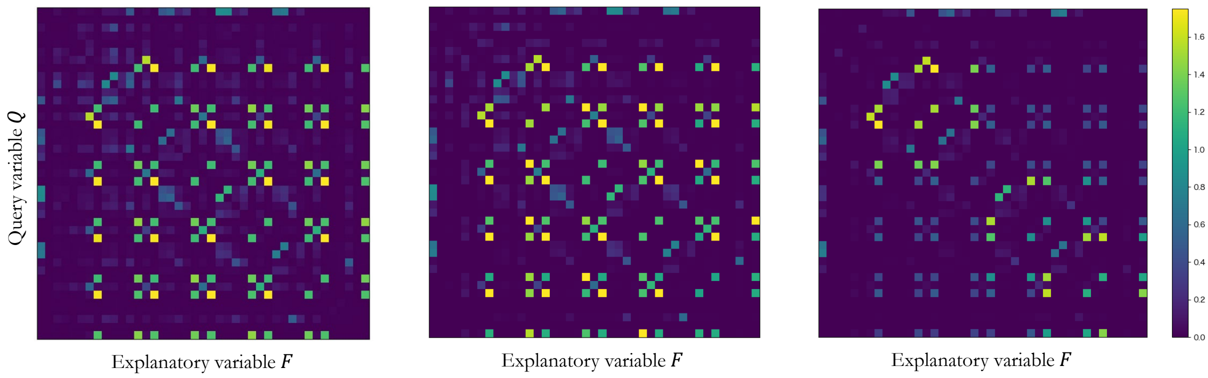


Figure 1 – *Estimated BN graph by the proposed approach*



(a) *Test data* (b) *Proposed BN* (c) *Existing BN*

Figure 2 – *Comparison of mutual information*

3.2 Validation Results

First, the graph structure obtained from the proposed BN-ACT model is illustrated in Figure 1. In this graph, variables are depicted as nodes, and influences between variables are depicted as edges. Additionally, connections among the same variables within dynamic objects representing prior activities are indicated by orange-filled nodes. The large number of filled nodes demonstrates that the influences among dynamic objects are effectively captured. Furthermore, the distinct graph structures within each dynamic object highlight the importance of estimating the structures for each dynamic object separately. Additionally, the rate of infeasible schedules generated by the existing BN was 5.44%, while the proposed model significantly improved this rate to 1.72%. Furthermore, without applying the resampling algorithm outlined in Section 2.2, the rate was 4.45%, clearly demonstrating the effectiveness of the proposed resampling algorithm.

Second, we assess how new information from one node affects the expected changes in the posterior probability distribution of a target node. Mutual Information (Marcot (2012)) calculates the reduction in entropy when new data is added from one variable in the network. Let Q represent the query variable, and F represent the explanatory variable, as shown in Figure 2. Cells with higher values, shown in yellow, indicate a greater degree of mutual dependency. The proposed BN model (Figure 2b) successfully identifies inter-variable dependencies. Our model accurately reproduces both strong and weak dependency pairs. Conversely, the existing BN model (Figure 2c) fails to replicate pairs with weak dependencies in the test data.

Third, the generation rates for activity chains by type are evaluated and presented. Figure 3 shows only the major activity chains. The proposed BN closely matches the distribution of activity chains in the test data. However, the reproducibility of the existing BN is poor, particularly in underestimating less common activity chains with frequencies around 0.05. The

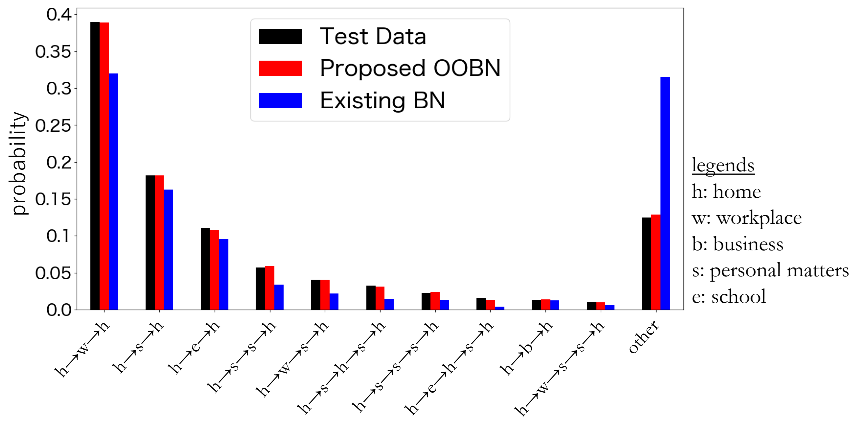


Figure 3 – *Composition ratio of generated activity chain*

Mean Absolute Percentage Error (MAPE) for the proposed BN was 4.0%, while the existing BN had a much higher error rate of 38.1%. The proposed BN effectively captures the relationships between variables across different activities, leading to improved reproducibility of activity chains that involve multiple activities.

4 DISCUSSION

This study proposes a Bayesian Network (BN) activity model that can flexibly build model structures in a data-driven way, and validates the model using real data from the Tokyo metropolitan area. The model effectively estimated the differences in dependencies among variables caused by the sequence of activities using a dynamic object-oriented BN. By incorporating resampling that accounts for time constraints, it successfully generated plausible activity chains.

References

De Waal, Alta, & Joubert, Johan W. 2022. Explainable Bayesian networks applied to transport vulnerability. *Expert Systems with Applications*, **209**, 118348.

Joubert, Johan W, & De Waal, Alta. 2020. Activity-based travel demand generation using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, **120**, 102804.

Koller, D. 1997. Object-oriented Bayesian networks. *In: Proc. of 13th Conf. on Uncertainty in Artificial Intelligence, 1997.*

Marcot, Bruce G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological modelling*, **230**, 50–62.

Roorda, Matthew J, Miller, Eric J, & Habib, Khandker MN. 2008. Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A: Policy and Practice*, **42**(2), 360–375.

Sallard, Aurore, & Balać, Miloš. 2023. Travel demand generation using Bayesian Networks: an application to Switzerland. *Procedia Computer Science*, **220**, 267–274.

Weber, Philippe, & Jouffe, Lionel. 2006. Complex system reliability modelling with dynamic object oriented Bayesian networks (DOOBN). *Reliability Engineering & System Safety*, **91**(2), 149–162.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers 23H01527 and the Fusion Oriented Research for Disruptive Science and Technology Program Number JPMJFR225Q by the Japan Science and Technology Agency (JST).