Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
2023, Vol. 2677(8) 1–23

DOI: 10.1177/03611981231155426
journals.sagepub.com/home/trr
SAGE

Activity Schedule Modeling Using Machine Learning

Anil Koushik[1], M Manoj[1], N Nezamuddin[1], and AP Prathosh[2]

Abstract

A novel data-driven approach for activity schedule modeling is presented in this paper. The paper's contribution is twofold. First, the activity schedule is modeled as a time series to facilitate simultaneous prediction of activity participation, start times, and duration. Simultaneous prediction helps avoid assuming a predefined decision structure and allows all possible interdependencies among these choice facets to be modeled. The time series representation also ensures time budget constraints are automatically satisfied. Second, a machine learning tool called long short-term memory (LSTM) network is used to model the time series. The LSTM's ability to model long-term dependencies ensures that activity patterns are generated considering the influence of distant and recent past. A bidirectional LSTM is used to capture the effect of (planned) future activities on the present activity participation. The model derives all the relations from the data without requiring assumptions by the modeler on the decision-making behavior. Further, the problems arising from class imbalance in the schedule caused due to less frequently performed activities are also explored and addressed. The models are calibrated and validated using the activity-travel diary data from the OViN 2016 dataset. To evaluate the robustness of the model, it is also tested on a time budget dataset with 23 different activity types. The results indicate that the proposed method can predict the distributions of activity start times and duration with reasonable accuracy. The results demonstrate that the proposed method can efficiently model activity schedules and can be a useful tool for travel demand modeling.

Keywords

Long Short-Term Memory (LSTM) networks, activity-based modeling, travel demand modeling, neural networks, imbalanced datasets

In recent years there has been a surge in the use of activity-based travel demand models. The strength of these agent-based models lies in treating travel as a demand derived from the need and desire to participate in activities. Typically, activity-based models aim to predict the following facets of an individual's activity schedule: set of activities in which the individual participates, the timing and duration of the activities, the location and mode of travel (if traveling is involved), and details of the accompanying parties (*1*). These facets are modeled as functions of individual and household sociodemographics, and land-use and environmental characteristics subject to inter-personal, spatial, and temporal constraints. Among the existing approaches for activity-based modeling two major paradigms can be seen: a utility maximizing econometric approach and a computational process modeling (CPM) approach (*2*). Each of these approaches has its strengths and limitations and modelers are constantly exploring alternative methods to address some of the shortcomings in the existing methods (*3*). Among several such methods, data-driven methods have recently emerged as effective alternatives to traditional approaches.

Data-driven methods, also referred to as machine learning (ML), are tools that automatically discover patterns and relationships in data. Recent advances in this field, particularly in deep learning, have revolutionized

[1]Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India
[2]Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India

Corresponding Author:
M Manoj, manojm@civil.iitd.ac.in

image and natural language processing. They have also seen a wide variety of applications in transportation research (*4*). Their capacity to model complex nonlinear relationships with no assumptions on data distributions or on the functional form of the relationship is an advantageous characteristic for transportation research (*5*). Additionally, the ability to handle a variety of data types, as well as relative simplicity for field implementation make them attractive alternatives to the conventional methods. Further, the increased availability of data thanks to recent advances in GPS and social media-based data collection methods is expected to enhance their applicability in transportation research (*6*). Despite numerous applications in transportation research, machine learning has had limited applications in activity schedule modeling.

In this paper, we present a proof of concept of a data-driven approach for activity schedule modeling which simultaneously predicts activity participation (activity type) and the timing and duration of participation. This is achieved by representing the schedule as a time series and employing a machine leaning (ML) technique called a long short-term memory (LSTM) network to model the time series. The time series representation facilitates simultaneous prediction of activity participation, start times, and duration eliminating the need for a predefined decision structure and allowing all possible interdependencies among the choice facets to be modeled. Additionally, problems resulting from activity overlap during prediction are avoided and time budget constraints are inherently satisfied. This kind of representation does present a challenge that the sequence gets quite long and incorporating long-term dependencies becomes difficult. To address this, this study uses LSTMs which can model long- and short-term dependencies and ensure that activity patterns are generated considering the influence of distant and recent past. Further, a bidirectional LSTM is adopted to account for the influences of activities planned for future on the present activity participation. The model is calibrated and validated using the activity-travel diary data of a sample from the Netherlands captured in the ODiN 2016 data set. To address the underrepresentation of the less frequent activities in the schedule, a weighted loss function is used. In addition, comparison with models using unidirectional LSTM and other methods of handling class imbalance in data are also presented.

Background

Based on modeling approaches, activity-based models can be viewed under two major paradigms: econometric method-based models and CPMs (*2*). Econometric method-based model systems employ econometric methods such as multiple linear regression, hazard-based models, and discrete choice methods. Some of the prominent model systems in this category include PCATS (*7*), the BB system (*8*), and CEMDAP (*9*). The BB system is a tour-based system that uses a nested logit structure to compute schedule probabilities where the primary tour characteristics are evaluated first, and the secondary tour characteristics are computed conditioned on the primary tour features (*8*). PCATS also uses a nested logit structure to develop a sequential activity schedule constrained within the Hägerstrand's space-time framework (*7*). CEMDAP is another tour-based system that employs a series of different econometric methods such as linear regression, hazard-based models, and utility-based methods to model schedule features (*9*). More recently, CUSTOM (*10*) used a random utility maximization approach by jointly modeling the type of activity, its location, timing, and duration. Most of these modeling systems use the utility maximization approach at some point whose treatment of individuals as rational, optimizing agents was deemed unrealistic (*11*).

To relax these assumptions, computational process models were suggested which were intended to imitate the decision-making process of individuals (*12*). Among computational process models, SCHEDULER (*13*) and AMOS (*14*) are some of the earliest frameworks. ALBATROSS, the first fully operational CPM, is a heuristic-based multi-agent model (*1*). It uses a set of 'if-then' rules represented in a CHAID decision tree to model schedule-related decisions. The activity attributes such as location, duration, companion, and travel mode choice are decided successively in a predefined order of priority. TASHA uses the concept of a project and activity schedules are generated from a project agenda in a fixed order of priority (*15*). ADAPTS has a dynamic framework that allows activity attributes (time of day, destination, travel mode, company) to be decided at different time-steps based on their flexibility (*16*). The main criticism of these computational process models (except ADAPTS) is that the assumed decision-making process is based mostly on experience and heuristics rather than empirical evidence (*17*). Empirical studies have often reported results contradicting their assumptions (*18*).

More recently, data-driven methods are being explored as alternatives to existing approaches. Data-driven approaches employ ML algorithms that learn relationships from data themselves and reduce the need for assumptions on decision-making process or on the underlying choice behavior. Their ability to model highly nonlinear relationships with sufficient accuracy makes them suitable candidates to model complex activity schedules. Moreover, recent advances in GPS and social media-based data collection methods and the resulting availability of large-scale data sets has drawn more attention toward these data-driven methods (*6*). They are easier to implement and transfer to a new region only needs retraining on the new data which reduces implementation costs. While the black-box nature of these methods remains a concern, recent progress in Explainable AI (XAI) (*19*, *20*) shows significant promise in addressing that issue.

Many studies have identified these advantages and have explored their applicability in travel behavior research. While some studies use ML to model only certain facets of a schedule, others use a series of submodules to model more complex patterns or the entire schedules. ALBATROSS, for example, employs decision trees to generate entire activity schedules for a day. Vanhulsel et al. (*21*) use reinforcement learning to generate activity schedules using activities as states and similarity to the actual schedule as a reward. They use q-learning to simulate activity sequences and improve the q-learning by using CART decision trees to approximate q-values. They report a shorter computation time and better results given this improvement. Allahviranloo and Recker (*22*) deploy a series of support vector machines (SVMs) to predict activity sequences and compare the model performance with multinomial logit (MNL) models. They report a superior performance of SVM over MNL models. Further, they conclude that models appearing later in the series benefit from the earlier models' results. Anda and Ordóñez Medina. (*23*) propose an extension of hidden Markovian models based on input-output hidden Markovian models (HMM) presented in Bengio and Frasconi (*24*) to model activity schedules and then generate schedules for synthetic population. More recently, Drchal et al. (*6*) developed a data-driven activity scheduler (DDAS) with four models sequentially predicting activity participation decisions, activity duration and location, and travel mode using decision trees. Outputs from previous models are input to subsequent models. The model progresses in time, conditioning future decisions on past decisions.

Recently another line of research has evolved which focuses on extracting activity travel patterns from nonconventional sources such as smart card data and GPS trajectories. For example, Ordóñez Medina (*25*) uses smart card data to extract weekly primary activity patterns by detecting activity types with the help of household travel survey data. The study then clusters patterns using the DBscan algorithm to derive most common primary activity patterns. Anda et al. (*26*) use dynamic Bayesian networks to develop several HMMs to synthesize disaggregate level activity travel patterns from the aggregate-level mobile phone data. These studies rely less on socio-demographic data and generate patterns based on observed trajectories directly.

A study of the literature on data-driven methods in travel behavior research suggests that the two most frequently employed ML algorithms have been neural networks (NN) and decision trees (DT) (*4*). DTs offer the advantage of interpretability, whereas NNs have the advantage of flexibility which allows new architectures to be designed to suit the problem and also facilitate modifications and improvements at a later stage. Their ability to model complex nonlinear relationships, handle a variety of data types, and their robustness to outliers can be very useful in handling activity travel data.

While DTs have been used frequently for activity schedule modeling NNs see their application limited to mode choice prediction, and have been scarcely used in activity generation and scheduling models (*4*). Kato et al. (*27*) use a sequence of NNs to model commuters' tours and the intermediate stops during home-to-work and work-to-home trips. The study reports that error in this series of neural network models is progressive; that is, networks later in the sequence show inferior prediction abilities. This could be extended to most of the decision tree based methods reported above which also use a sequence of models and could be prone to error progression. That is, models appearing later in the sequence are affected by the prediction errors in the models appearing before. This error progression happens as a result of the lack of a feedback mechanism, and the networks that are earlier in the sequence are uninformed about the poor performance of the networks placed later in the sequence. Recurrent NNs (RNN) allow for a feedback mechanism that enables them to maintain excellent performance through all the stages of the modeling (*28*).

The contribution of this study is twofold: (i) it uses time series representation to model activity schedules; and (ii) it employs a NN-based ML tool called LSTM network to model the time series. Representing schedule as a time series enables simultaneous prediction of activity type, activity duration, and time of participation, eliminating the need for assuming a fixed order of prediction. This brings flexibility to the model and allows all possible interactions among activity type, start times, and durations to be modeled. This is an advantage over earlier studies which assume a fixed hierarchical decision-making process restricting the way in which these facets influence each other. The time series representation also ensures the 24-h time constraints are satisfied by design itself. In addition, the complexities arising while configuring the interaction among sub-models (to ensure integrity) are also avoided (*29*).

While the time series representation of an activity schedule has been done in the past, the scope was limited to clustering and to understanding the behavioral patterns (*30*–*32*). To the authors' knowledge, none of the earlier studies use time series representation to model an

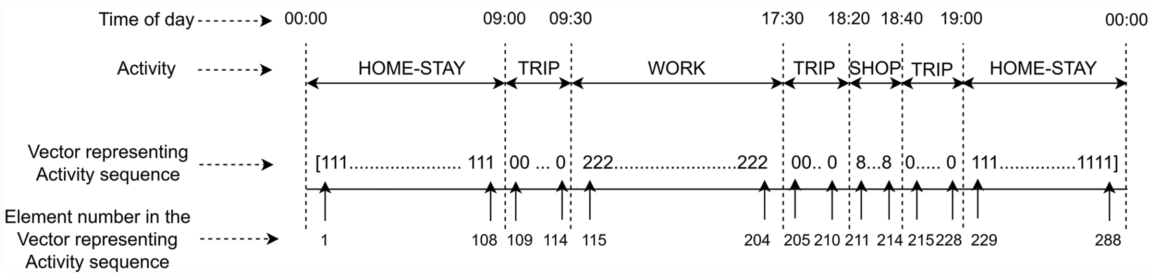

Figure 1. Activity pattern representation.

activity schedule and generate activity sequences. This could be attributed to the time series representation having a disadvantage as the 24-h schedule gets quite long and incorporating the influence of all the previous occurrences can be difficult as we get to the later part of the day. LSTMs, however, have the ability to model long sequences and remember long-term occurrences. This ability to model both long- and short-term dependencies suits the activity schedule modeling where activities executed immediately before and a long time ago influence current decisions. Besides, LSTMs, being data-driven, eliminate the need for modeler to make assumptions on the choice behavior (such as utility maximization) or on the distributions of parameters. Overall, this paper proposes a completely data-driven approach which derives all the relations from the data and does not require any assumptions from the modeler's side on the decision-making behavior or on the parameter distributions.

Though an earlier study (*33*) used LSTMs for schedule modeling, the model proposed in this paper is different in several ways. First, we use bidirectional LSTMs that can also include the influence of activities planned for the future on the present decisions. In addition, the time series representation brings several advantages to our proposed method (discussed before) and enables us to realize the LSTM networks' potential to a fuller extent. Moreover, the model proposed in this paper is more general compared with the student-specific model reported in Drchal et al. (*33*).

Methods

Time Series Representation of Activity Schedule

An individual's activity schedule is an outcome of multiple decisions interacting and influencing each other in a dynamic environment. Given this complexity, formulating a modeling system that can accurately represent the decision process becomes practically infeasible. Therefore, modelers are forced to make simplifying assumptions on the decision-making process to arrive at a practically feasible model. Often, modelers assume a sequential decision-making process where activities and their attributes are assumed to be decided one after the other based on priority and each decision is conditioned on all the preceding decisions. The credibility of the assumed sequence of priorities is often questioned in the literature (*17*) and an alternative could be to simultaneously predict different facets of the schedule by representing it as a time series.

In this study each activity schedule is represented as time series of 288 elements. Each element represents a 5-min interval of the individual's day (24-h period). The first element represents 00:00 to 00:05 a.m. The next element represents 00:05 to 00:10 a.m., and so on. The last element represents 23:55 to midnight. Each one of the 288 elements can take a value from the set, N = {0,1,2,3,4,5,6,7,8} representing an activity from the activity set, A = {0:*trip*, 1:*stay at home*; 2:*work*; 3:*pick-up/drop*; 4:*social visit*; 5:*leisure activities*; 6:*personal care*; 7:*education*; 8:*shopping*}. Figure 1 explains this representation with an example. This representation ensures that activities do not overlap over time, thus avoiding conflict resolution complexities. It also ensures that the time budget constraints are inherently satisfied. Incidentally, time series representation also adds a practical advantage that it imparts uniformity across individual schedules where each schedule has the same length and each occurrence can be identified by a specific time stamp. This is very useful in pattern recognition algorithms.

Time Series Modeling

A time series is a sequence of observations of a given variable at different points in time. Time series data are different from cross-sectional data in the sense that the observations at a time-step are not independent of previous observations. Given such a time series, it will be of interest to predict a particular outcome, y_t, at a time t, given the set of previous observations $(y_1, y_2, \ldots\ldots\; y_{t-1})$. Several methods, referred to as the time series models, have been developed for this purpose of which autoregressive (AR) models, Kalman filters (KF) and HMM are the most common (*28*, *34*).

A common drawback of these time series modeling methods is that the relationship between the variables in a sequence is assumed to be linear. Further, HMM, and KF assume that the present state is dependent only on its immediate predecessor. Attempts to extend this dependency further for HMMs, lead to an exponential growth in computational complexities (*35*). Though k-th order AR models include the dependency on k previous time-steps, they too fail to capture long-term dependencies. The same argument holds good for feed forward NNs that concatenate a few previous occurrences with the input to capture the sequential effect (*35*).

RNNs struggle to remember long-term occurrences and this makes them unsuitable for modeling long time series data. This is attributed to the so-called the vanishing/exploding gradient problem. For details, readers may refer to Graves (*28*). The vanishing/exploding gradient problem can be tackled using "gates" that regulate information flow across layers and restrict the gradients from vanishing or exploding. This concept of gates forms the basis for LSTM networks (*28*).

LSTM Networks

LSTMs are variants of the more basic feed forward neural networks (FFNNs). In fact, "unfolded" LSTMs can be viewed as a series of connected FFNNs which share the same parameters (Figure 2*a*). A thorough knowledge of FFNNs is necessary for a better understanding of LSTMs but for the sake of brevity they are not explained here. For a detailed discussion of FFNNs, readers are referred to Goodfellow et al. (*36*).

An LSTM cell can be represented as shown in Figure 2*a*. The "loop" in the figure indicates that the states of the hidden layer at the previous time instant are used as input at the current time instant (Figure 2*a*). In the figure, $X_0,\ldots,X_t$ is the external input sequence, A is a neural network and $h_0,\ldots,h_t$ is the output sequence. LSTMs aim to predict these sequences as close as possible to the observed sequences.

Figure 2*b* shows a typical LSTM cell. An LSTM cell takes as inputs $\boldsymbol{x_t}$, the external input at the present time-step (t), $\boldsymbol{h_{t-1}}$, the previous time-step output, and gives $\boldsymbol{h_t}$, the output of the present time-step. The LSTM cell is designed to assist the network in deciding what information to store, retain, and discard to optimize its prediction abilities. LSTMs use special units known as cell states ($\boldsymbol{C_t}$ and $\boldsymbol{C_{t-1}}$ in Figure 2*b*), which facilitate information flow across the cells (time-steps) efficiently (Graves 2008). These cell states can be viewed as memory units that store information on all the previous occurrences that the network deemed necessary (long-term memory).

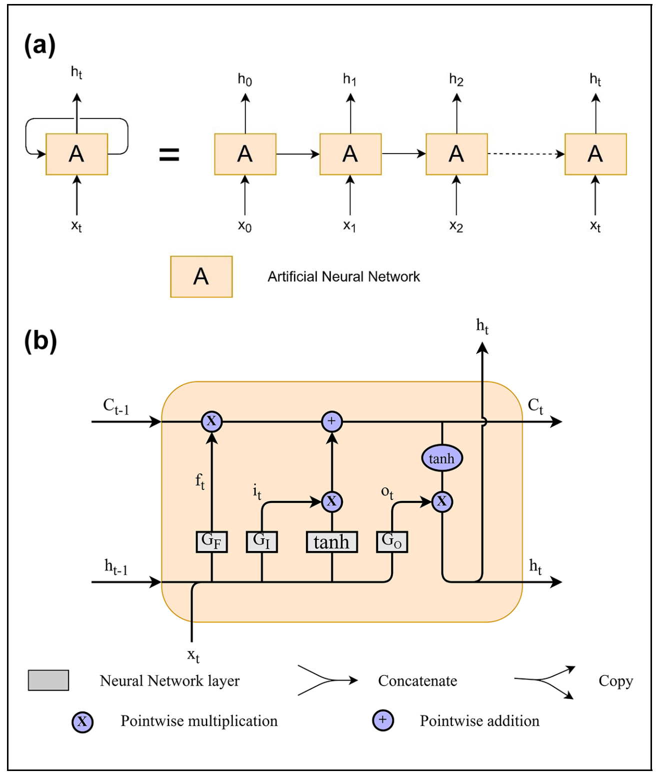

Figure 2. (*a*) Unrolled LSTM cell; and (*b*) structure of an LSTM cell.
Note: LSTM = long short-term memory.

Neural network structures termed "gates" are used to access and edit these cell states (*37*). Gates help the network to decide what information to retain and what to discard (forget gate), what information to update (input gate), and give the present time-step output (output gate).

A gate acts like a neural network layer which takes the weighted sum of its inputs and passes it through a quashing function known as activation and gives the output. All the LSTM gates use a sigmoid function as activation which gives outputs between zero and one (*28*). A typical LSTM cell contains three gates and they are organized as shown in the Figure 2*a*. The first gate with the output $\boldsymbol{f_t}$ is called the forget gate (G_F). As the name indicates, the forget gate compares the output in the previous time-step ($\boldsymbol{h_{t-1}}$) and the external input at the current time-step ($\boldsymbol{x_t}$) and decides what information from the previous cell state to discard (Equation 1).

$$\boldsymbol{f_t} = \sigma(\mathbf{W_f} \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x_t}] + \boldsymbol{b_f}) \tag{1}$$

$\mathbf{W_f}$ is known as the weight matrix in the above equation and $\boldsymbol{b_f}$ the bias vector. σ represents sigmoid activation. The next step is to decide the new information that must be added to the cell state. A *tanh* layer generates the values that must be considered for storage ($\tilde{\boldsymbol{C}}_t$). The input gate (G_I), with an output $\boldsymbol{i_t}$, decides what part of the

Observed	.........2,2,2,2,2,2,2,2,2,2,2,2,2,.......
Predicted	.........2,2,2,2,2,2,2,4,2,2,2,2,2,2,......

Figure 3. Comparison of observed and predicted sequences.

values generated by *tanh* layer must be added to the cell state. The input gate and the *tanh* layer outputs are combined to compute the updated cell state ($\boldsymbol{C_t}$) (Equations 2, 3, and 4).

$$\boldsymbol{i_t} = \sigma(\mathbf{W_i} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_i}) \tag{2}$$

$$\boldsymbol{\tilde{C}_t} = tanh(\mathbf{W_C} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_C}) \tag{3}$$

$$\boldsymbol{C_t} = \boldsymbol{f_t} * \boldsymbol{C_{t-1}} + \boldsymbol{i_t} * \boldsymbol{\tilde{C}_t} \tag{4}$$

$\mathbf{W_i}$ and $\mathbf{W_C}$ are weights for the input and tanh layers, and $\boldsymbol{b_i}$ and $\boldsymbol{b_C}$ are biases. "*" operator indicates element wise multiplication. The next step is to decide the cell's output for the present time-step ($\boldsymbol{h_t}$). The present cell state is passed through *tanh* activation, generating the output values. These values are passed through an output gate (G_O), whose output is $\boldsymbol{o_t}$, to decide which of the generated values to output (Equations 5 and 6). $\mathbf{W_O}$ and $\boldsymbol{b_O}$ are the weights and biases of the output gate.

$$\boldsymbol{o_t} = \sigma(\mathbf{W_O} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_O}) \tag{5}$$

$$\boldsymbol{h_t} = \boldsymbol{o_t} * tanh\ (\boldsymbol{C_t}) \tag{6}$$

The output h_t at each time-step is the predicted conditional probability mass function $\hat{y}_t = \{\hat{y}_{t,1}, \hat{y}_{t,2}, \ldots, \hat{y}_{t,M}\}$, where $\hat{y}_{t,k}$ is the predicted probability that the output is class (in this case activity type) k. M is the total number of output classes. This predicted output is compared with the observed output, $y_t = \{y_{t,1}, y_{t,2}, \ldots\ldots y_{t,M}\}$, where $y_{t,k} = 1$ if the expected output for that time-step is class k and 0 otherwise. A loss function is used as a measure of dissimilarity between the observed and the predicted sequences and the model parameters (referred to as "weights") are estimated by minimizing the total loss (across all time-steps for all individuals). This is achieved by using an efficient algorithm called the backpropagation algorithm which adjusts the weights by using the gradients of the loss function with respect to the weights so that the total loss is minimized (*28*). The loss function used in this study is the categorical cross-entropy given in Equation 7.

$$L(y_t, \hat{y}_t) = -\sum_{k=1}^{M} y_{t,k} log \hat{y}_{t,k} \tag{7}$$

As in the case of a deep NN, it is possible to stack the LSTM cells to create multi-layered LSTMs. Each layer output is fed as input to the next layer. The output of the final layer is the model output. It is noted here that in the context of this paper, x(t) does not exist as the input is socio-demographic information about the individual, which has no sense of sequentiality and can be fed all at once; whereas, the output is a sequence and predicted one after the other. For the sake of generality the complete version was presented here. The entire input is fed to the network before the first time-step itself.

Note that the forget gate acts on the cell state itself and allows the cell state to "reset" itself. The recurrent connection that connects the cell state to itself is assigned a weight of 1.0 (no other weight assigned in Equation 4) which ensures that the weight adjustments during back propagation do not blow up or diminish. This allows LSTMs to retain long-term information in cell states. The output at each time-step ($\boldsymbol{h_t}$) can be viewed as carrier of short-term memory. In the context of activity schedule modeling, the short-term memory of LSTM keeps track of the duration of the present activity along with the information on type and duration of previous activity. The long-term memory keeps track of the earlier participations and durations. The time of participation is built in to the time-step.

Suitability of the Proposed Method for Activity Schedule Modeling

It is clear from Equation 7 that the network tries to maximize the number of time-steps with correct predictions. This may not appear to be the most effective approach for modeling activity schedules. Take, for example, Figure 3. From Equation 7's perspective, the model's performance will be graded favorably because only one time-step is predicted incorrectly. However, from the perspective of a schedule, this has resulted in the division of a long-duration activity episode into two shorter-duration activities. As a result, the suitability of LSTM networks for such time series with long repeated occurrences, where the length of repetition is of interest, can be questioned. However, if we closely examine Equation

7, we will notice that the $\hat{y}_{t,j}$ actually represents the conditional probability

$$\hat{y}_{t,j} = P\big(y_{t,j} = 1|y_{t-1}, y_{t-2}, \ldots\ldots\ldots\ldots, y_0, x\big) \tag{8}$$

It can be said that the model aims to predict the probability mass function for the activity type at the present time-step given all the previous occurrences. Specifically, if we consider the case where the previous k occurrences are of a particular activity type a, the Equation 7 becomes

$$\begin{aligned}\hat{y}_{t,j} &= P(y_{t,j} = 1|y_{t-1,a} = 1, y_{t-2,a} = 1, \ldots\ldots, y_{t-k,a} \\ &= 1, y_{t-k-1}, y_{t-k-2}, \ldots\ldots, y_0, x).\end{aligned} \tag{9}$$

Now, for $\forall\ j \neq a,\ \hat{y}_{t,j}$ represents the probability that a new activity will be started given the activity participation history. The relative magnitudes of $\hat{y}_{t,j}$ will determine the activity type that is more likely to be participated in. For j = a, Equation 9 becomes,

$$\begin{aligned}\hat{y}_{t,j} &= P(y_{t,a} = 1|y_{t-1,a} = 1, y_{t-2,a} = 1, \ldots\ldots, y_{t-k,a} \\ &= 1, y_{t-k-1}, y_{t-k-2}, \ldots\ldots, y_0, x)\end{aligned} \tag{10}$$

where the probability term on the right-hand side can be viewed as a discretized version of the hazard function used in the hazard-based activity duration models. The only difference is that the present model also accounts for all the previous activity participations and their durations. As a result, we hypothesize that the model is attempting to jointly learn the distribution of activity start times and durations. This actually resembles schedule execution in reality where a person engaged in an activity at a particular point in time can either continue to be engaged in that activity or switch to another with some probability associated with each of the possibilities. When the switch happens, the completion of the previous activity duration and the start of a new activity happen simultaneously. Therefore, simultaneous prediction of the distributions of activity start times and durations (as explained by Equations 9 and 10) is more justified compared with predicting them sequentially.

Model Development

Data

This study uses OViN 2016 data, collected by Centraal Bureau voor de Statistiek (CBS) and Rijkswaterstaat (RWS) (*38*). The data set was downloaded from the DANS website after obtaining permission from the depositor (accessible through the following link: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:73441). The OViN data set contains information about the daily mobility of the Dutch population. The data set contains a 24-h activity-travel diary (from 00:00 to 00:00) of individuals from all over the Netherlands, with only out-of-home activities included. Along with the activity-travel information, the data also contain socio-demographic information of individuals and households. The total number of individual diaries in the data set is 37,229. In this study, only those individuals belonging to the capital municipalities of each of the 12 provinces are considered. This set contained 3,949 individual dairies. Of these, after the initial process of cleaning, 3,686 individuals were considered for the study. Individuals without any trips throughout the day are also included in the study to make the model more general.

Model Inputs

The inputs to the model are household demographics (household size, composition, number of people in different age groups, household income, vehicle ownership), individual characteristics (position of the individual in the house, sex, age, origin, working status, education status, driving license ownership status), day of the week and month, and the day for which travel diary is reported.

Feature Selection

Since the output of our model is a sequence, finding the correlations between input and output cannot be carried out directly as we need to incorporate the effect of features on sequntiality, and also incorporate long- and short-term dependencies. To address this issue, this study uses a feature embedding method called sequence graph transform (SGT) proposed in Ranjan et al. (*39*). SGT is suitable for this study as it can incorporate both long- and short-term dependencies and is computationally efficient (*39*). The SGT outputs an SGT matrix, whose elements represent a nonlinear transformation of the distance between each type of activity in the sequence. In other words, the SGT outputs an embedding corresponding to each pair of activity types *(a,b)* where $a, b \in N = \{0,1,2,3,4,5,6,7,8\}$ and N represents the activity set. More importantly, it gives a finite dimensional vector corresponding to each sequence that allows us to compute correlations with the input. Since the embeddings corresponding to each possible pair of activity types is output, the SGT provides a sparse 81-dimensional output in the present case. Therefore, the principal component analysis (PCA) is applied on the SGT matrix to reduce its dimensions. Finally the correlation between PCA transforms and the input is computed.

It is observed here that in the time series representation used in this study, embeddings with $a = b$ in the activity pair (a, b) will account for activity duration as the number of such occurrences for a particular activity will increase if the duration of activity is higher. Embeddings with $a \neq b$ will account of sequentiality between activity types. However, the embeddings are not position specific and SGT, thus, cannot account for activity start times. Another matrix of activity start times is, therefore, created, with each column representing an activity type. Using this matrix, correlations between activity start times and input features are also computed. Finally features that exhibit correlations with either SGT embeddings or with activity start times are used as model inputs. A descriptive summary of the selected input variables is given in Table 1.

Handling Class Imbalance

Activities which are less frequently performed are underrepresented in the schedules. Incidentally, small duration activities also get underrepresented in the time series representation of activity schedules. This underrepresentation leads to class imbalance, and the model tends to underpredict classes with lesser representation. A common method used to fix class imbalance is oversampling. The same method is adopted in this paper. In oversampling, data points containing patterns with the underrepresented activity types are sampled (with replacement) at random and are repeated in the data set until the relative frequency (total number of episodes of that activity type divided by the total number of episodes of all activity types) reaches at least 5%. It is, therefore, ensured that all the activities have at least 5% representation.

Another way of handling class imbalance is by using weighted loss functions. The categorical cross-entropy loss function (Equation 7) is modified by multiplying each term with a weight, w_i (Equation 11). The weights are inversely proportional to the relative proportion of each class in the data set ensuring that the model is penalized more severely for misclassification in the case of minority classes than for misclassification in the case of the majority class. In this study, we develop a model using weighted categorical cross-entropy and compare the performance with the oversampling method.

$$L(y, \hat{y}) = -\sum_{i=0}^{M} w_i y_i log \widehat{y_i} \tag{11}$$

Model Implementation

To train the model, the data was split into five sets of equal size, and in each step, the model performance was tested on one set while the model was trained on the remaining four sets. One-eighth of the samples from the training set were used for validation (10% of total data). The average performance over all the five test sets was finally reported. This process is called cross-validation (*40*) and is commonly adopted in ML-based transportation studies (*41*, *42*). Figure 4 presents the flow chart depicting the model implementation.

This paper also compares the impact of random oversampling and using a weighted loss function on model performance. Furthermore, it also compares the performance of unidirectional and bidirectional LSTMs. To test the presence of long-term dependencies in the schedules, a comparison with RNNs is also provided. Therefore, five models are developed, namely Models 1, 2, 3, 4, and 5. Their features are shown in Table 2.

The model architectures are reported here. Models 1, 3, and 4 have a three-layer bidirectional LSTM networks with 200 units in each layer. Each layer is regularized with 20% dropout. Model 2 has two unidirectional LSTM layers containing 300 units each and regularized with 30% dropout. The activation function used is "tanh." The optimal weights, for weighted categorical cross-entropy, corresponding to the activity types listed in the previous section were obtained to be 0.14, 0.01, 0.05, 0.22, 0.09, 0.06, 0.25, 0.07, and 0.1 respectively. The final layer is dense with nine neurons. The final layer uses the "softmax" activation function. The model is trained for 500 epochs, and the one with the least validation loss is finally chosen as the best. The computer codes are written in Python programming language and make use of KERAS library. The programs were run in Google Colab platform and made use of the free graphical processing units (GPUs) provided by Google Colab.

Results and Validation

Typically, activity-based models are validated by evaluating their performance on an observed data set and comparing predicted and observed patterns (*1*, *15*, *43*). A common practice in the literature is to validate the model at the aggregate level since the predictions at the aggregate level will be used for the final implementation. However, aggregate-level validation may not be reliable in the case of policy evaluation for several reasons. It is possible for the performance to be poor at the disaggregate level, but the errors compensate for each other, resulting in a decent performance at the aggregate level. Consequently, it is crucial to examine the performance of models for various subgroups. This paper provides a meso-level comparison of model performance across subgroups based on gender (male/female) and occupation (regular workers/students/others). Also presented is a performance comparison concerning the day of the week for which the pattern was developed (weekday/weekend).

Comparing Individual Activity Patterns

Comparison between observed and predicted patterns at individual levels is commonly made using SAM

Table 1. Model Inputs

Variable description	Possible values	#
Number of people in the household (HH)	1,2,….,10	2.83 (1.45)
HH composition	Single-person household	21.63%
	Couple/Couple + other(s)	24.19%
	Couple + child(ren)/Couple + child(ren) + other(s)	46.30%
	1 parent family + child (ren)/one-parent family + child(ren) + other(s)	7.89%
Position of the candidate in the HH	HH head (Single-person household)	21.63%
	HH head (Multi-person household)	3.26%
	Spouse/partner (of HH head)	46.14%
	Child	28.98%
Number of household members younger than 6 years	1,2,….,10	0.26 (0.60)
Number of household members from 6 to 11 years old	1,2,….,10	0.34 (0.66)
Number of household members aged 12 to 17 years	1,2,….,10	0.30 (0.64)
Number of household members aged 18 or older	1,2,….,10	1.92 (0.76)
Urbanity class of residential municipality	Very strong urban	57.90%
	Strong urban	36.68%
	Moderately urban	5.41%
Residential real estate class	20,000 to 50,000 inhabitants	2.61%
	100,000 to 150,000 inhabitants	8.83%
	150,000 to 250,000 inhabitants	16.89%
	250,000 inhabitants or more	33.67%
Age	1, 2, … 95	39.28 (23.4)
Origin	Native	74.17%
	Western immigrant	9.96%
	Non-Western immigrant	15.86%
Working status	Does not work	35.09%
	Works less than 30 h a week	13.56%
	Works more than 30 h a week	31.00%
	NA (less than 15 years of age)	20.26%
Sex	Male	48.32%
	Female	51.68%
Social participation	Employed	42.53%
	Homemaker	3.82%
	Student	23.86%
	Unemployed	2.83%
	Incapacitated	3.04%
	Retired	17.45%
	Younger than 6 years of age	6.46%
Highest completed education	Elementary education or less	7.62%
	Secondary education	14.09%
	Selective secondary education (HAVO, VWO)	25.80%
	Higher professional education	32.21%
	Younger than 15 years of age	20.28%
Household income	Less than € 10,000	3.96%
	€ 10,000–€ 20,000	25.05%
	€ 20,000–€ 30,000	32.40%
	€ 30,000–€ 40,000	22.25%
	€ 40,000–€ 50,000	8.91%
	Greater than € 50,000	7.43%
Driving license (DL) possession	Owns DL	61.49%
	Does not own DL	15.70%
	NA, younger than 17 years of age	22.81%

(continued)

Table 1. (continued)

Variable description	Possible values	#
HH Car ownership	0,1,2, 3 +	1.08 (0.76)
Primary user of the car	Candidate	39.18%
	Others	60.82%
Number of motorized two-wheelers	0,1,2,3 +	0.34 (0.68)
Number of bicycles	0,1,2,3,4,5,6 +	3.20 (2.1)

Note: NA = Not Applicable; HAVO = Hoger algemeen voortgezet onderwijs; VWO = Voorbereidend wetenschappelijk onderwijs.
For categorical variables, the percentage share of each category is reported. For continuous variables mean (standard deviation) are reported.

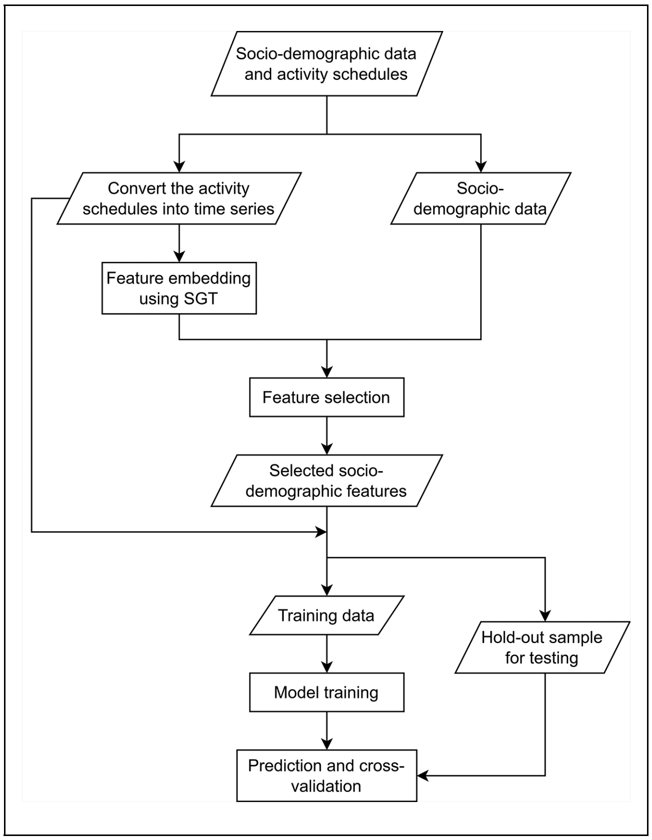

Figure 4. Flow chart depicting the model implementation.
Note: SGT = sequence graph transform.

Table 2. Model Types

Model name	Network type	Handling class imbalance
Model-1	Bidirectional LSTM	Weighted loss function
Model-2	Unidirectional LSTM	Weighted loss function
Model-3	Bidirectional LSTM	None
Model-4	Bidirectional LSTM	Oversampling
Model-5	Bidirectional RNN	Weighted loss function

Note: LSTM = long short-term memory; RNN = recurrent neural network.

(sequence alignment method) distance. SAM distance is used as a representative of the similarity between two sequences. It is measured based on the number of operations required to make one sequence the same as the other. A penalty of "1" is assigned for an addition/deletion of an activity episode, and "2" is assigned for a substitution. The final distance is the sum of all penalties (*44*). The greater the distance, the lesser is the similarity. The average SAM distances between observed and predicted sequences for Models 1, 2, 3, 4, and 5 were 2.55, 2.79, 2.89, 2.63, and 2.97 respectively. The results are comparable with, if not superior to, those reported in the literature (*21*, *45–47*).

Aggregate-Level Validation

The total number of observed activity episodes is 13,284. The number of episodes predicted by Models 1, 2, 3, 4, and 5 were 11,680, 12,392, 8931, 11,154, and 7,462 respectively. In general, all models underestimate the total number of activities participated and overestimate the duration of activities. However, when the performances are observed for individual activities, it can be seen that Model 1 performs the best. The macro-level results are tabulated in Table 3. Model 1 predictions for the total number of episodes for stay at home, work, education, and social visit differ from the original by −10%, + 3%, −2.8%, and −16.5%. Model 2 performs slightly worse with −12.6%, + 28.5%, + 12.1%, and 22.1% differences from the observed. For activities like pick-up/drop and personal care, Model 1 has prediction accuracies of −49.5% and −39.5% and Model 2 predicts −21.3% and + 9%. In predicting the number of episodes of each activity type, the results of Models 3, 4, and 5 are generally poor.

While understanding results related to activity start times, it is convenient to read Table 3 results along with Figure 5. Model 1 slightly underpredicts work, shopping, and education activity start times, while the performance is relatively good for other activities. The figure shows that the models generally follow the observed trends in

Table 3. Macro-Level Validation of Output

Activity type		Observed	Model 1 prediction	Model 2 prediction	Model 3 prediction	Model 4 prediction	Model 5 prediction
Stay at home	No. of activities	7,964	7,164	6,960	6,073	7,301	4,863
	Mean activity start time*	516.9 (507.2)	491.8 (529.0)	495.2 (544.4)	393.9 (504.0)	483.5 (510.4)	403.8 (618.9)
	Mean activity duration*	512.1 (369.4)	498.7 (328.7)	443.9 (263.7)	727.5 (446.6)	588.5 (397.6)	597.5 (683.9)
Work	No. of activities	907	934	1,166	857	1,058	606
	Mean activity start time*	614.1 (223.2)	575.8 (197.2)	560.7 (157.8)	603.0 (203.6)	626.2 (223.7)	317.9 (557.5)
	Mean activity duration*	370.8 (192.1)	381.0 (191.8)	393.8 (203.0)	385.8 (199.0)	325.7 (206.4)	935.8 (637.1)
Pick-up/drop	No. of activities	404	204	318	48	104	84
	Mean activity start time*	795.6 (241.3)	719.4 (237.1)	703.5 (244.8)	702.6 (238.5)	725.7 (231.1)	346.1 (594.0)
	Mean activity duration*	46.3 (89.2)	127.7 (123.1)	170.0 (176.2)	137.9 (123.2)	85.5 (100.6)	656.0 (683.0)
Social visit	No. of activities	607	507	741	359	489	332
	Mean activity start time*	901.2 (207.7)	898.2 (214.6)	919.4 (205.1)	929.4 (215.1)	927.7 (212.9)	341.2 (554.8)
	Mean activity duration*	164.9 (127.3)	203.4 (153.4)	249.0 (185.9)	191.6 (156.3)	153.1 (125.9)	815.9 (656.1)
Leisure activities	No. of activities	1,493	1,293	1,221	823	1,072	290
	Mean activity start time*	899.7 (234.0)	947.8 (248.9)	954.4 (235.5)	911.3 (235.6)	917.6 (233.9)	685.8 (622.7)
	Mean activity duration*	164.0 (139.5)	202.6 (168.6)	228.1 (189.4)	206.0 (167.8)	152.7 (137.3)	475.0 (627.1)
Personal care	No. of activities	167	101	182	25	82	126
	Mean activity start time*	754.2 (200.4)	728.6 (190.9)	714.6 (201.5)	717.0 (143.8)	762.2 (201.4)	557.9 (560.0)
	Mean activity duration*	82.4 (97.6)	137.0 (103.6)	162.9 (162.1)	167.2 (106.1)	82.4 (76.7)	319.6 (567.9)
Education	No. of activities	600	583	673	505	596	464
	Mean activity start time*	580.2 (156.4)	534.6 (89.8)	528.0 (91.7)	543.3 (107.9)	575.3 (183.7)	196.7 (438.8)
	Mean activity duration*	290.6 (148.3)	340.2 (138.7)	397.9 (143.8)	326.3 (126.9)	285.4 (165.9)	1,063.2 (590.5)
Shopping	No. of activities	1,142	893	1,131	241	452	698
	Mean activity start time*	818.4 (173.4)	760.8 (145.6)	721.9 (165.2)	808.0 (184.1)	791.1 (166.6)	410.7 (584.8)
	Mean activity duration*	69.5 (112.7)	172.9 (125.9)	239.4 (170.7)	167.7 (144.2)	100.2 (94.9)	682.0 (674.7)

Note: No. = number.
*All times in minutes. Numbers in brackets indicate standard deviation.

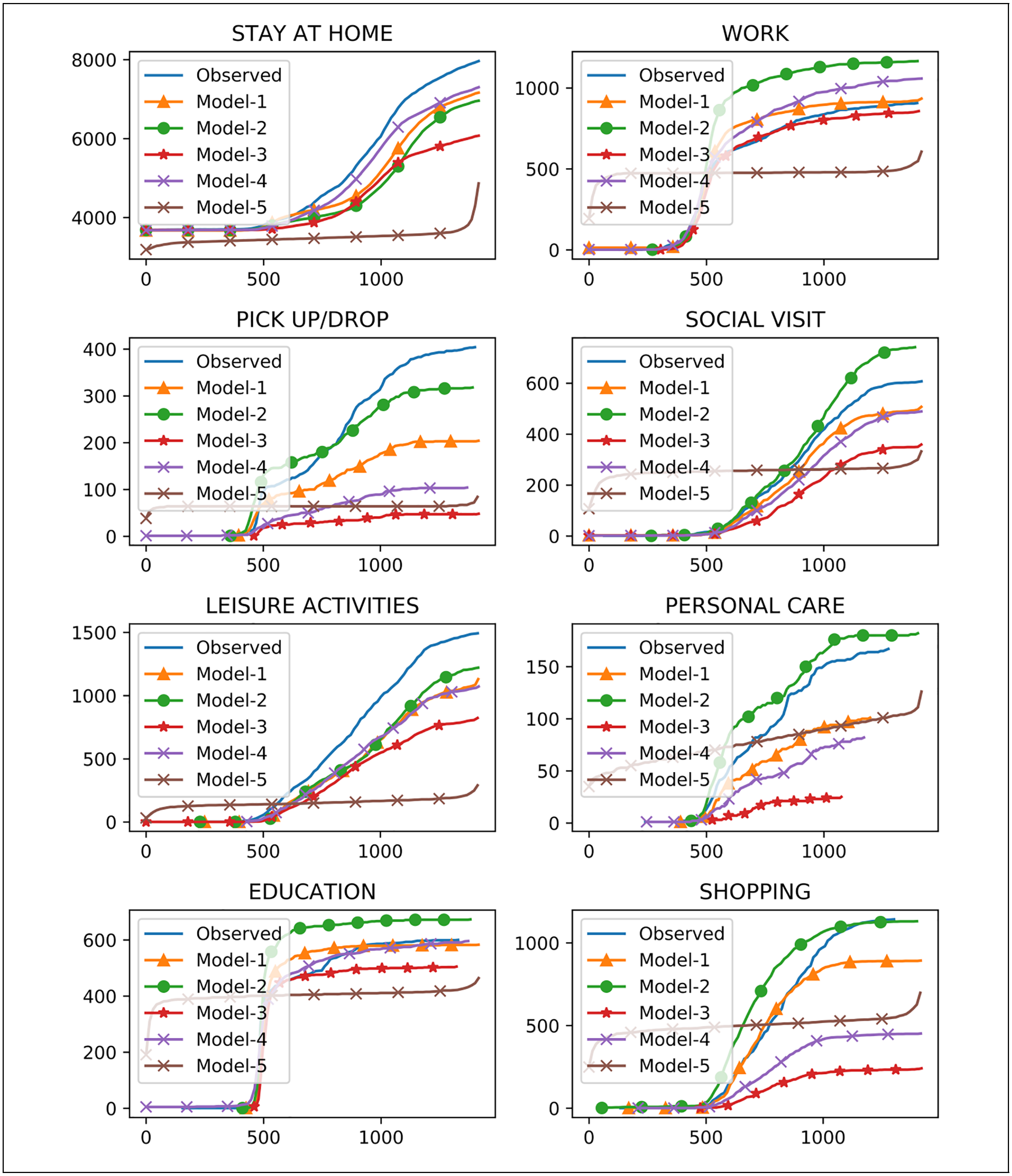

Figure 5. Activity start time patterns during a day. (The X-axis shows time of the day in minutes; the Y-axis shows the cumulative number of individuals starting an activity episode at that particular time.)

Table 4. Kruskal–Wallis Test Results

	Activity type	Model 1	Model 2	Model 3	Model 4	Model 5
Activity start time	Stay at home	0.27	0.92	0.00	0.00	0.00
	Work	0.61	0.01	0.64	0.08	0.00
	Pick-up/drop	0.00	0.00	0.03	0.02	0.00
	Social visit	0.54	0.07	0.06	0.04	0.00
	Leisure activities	0.00	0.00	0.52	0.09	0.00
	Personal care	0.31	0.05	0.46	0.62	0.00
	Education	0.60	0.02	0.00	0.76	0.00
	Shopping	0.00	0.00	0.08	0.00	0.00
Activity duration	Stay at home	0.04	0.00	0.00	0.00	0.00
	Work	0.81	0.06	0.21	0.00	0.00
	Pick-up/drop	0.00	0.00	0.00	0.00	0.00
	Social visit	0.00	0.00	0.05	0.02	0.00
	Leisure activities	0.00	0.00	0.00	0.00	0.00
	Personal care	0.00	0.00	0.00	0.20	0.00
	Education	0.02	0.00	0.00	0.50	0.00
	Shopping	0.00	0.00	0.00	0.00	0.00

activity start times; especially Models 1 and 2, generally run close to the observed trends. The performance of other models, except for Model 5, is also reasonably good, as can be seen from the table.

Table 4 shows the results of the Kruskal–Wallis test with a null hypothesis that the observed and the predicted start times/durations are from the same distribution. The table reports p-values indicating the maximum significance level at which the null hypothesis is accepted. Thus, if the p-value is greater than 0.01, it indicates that the null hypothesis is accepted at a 1% significance level. From the table, for activity start times, Models 1 and 2 have the null hypothesis accepted at a 1% significance level for all activities except pick-up/drop and shopping activities. In addition, the null hypothesis is rejected for leisure activities also for Model 2. For Model 3, the null hypothesis is accepted for most activities other than stay at home and education. For Model 5, the null hypothesis is rejected for all the activities.

For the activity durations, Models 1, 2, 3, and 4 predict mean work activity durations closely (+ 2.7%, + 6.2%, + 4.03%, and −12.7%). For the remaining activities, the durations are generally over predicted. However, the trends are picked for stay at home, leisure, and education activities (Figure 6). From Table 4, Model 1 has a p-value greater than 0.01 for stay at home, work, and education, while Model 4 has a p-value greater than 0.01 for social visit, education, and personal care activities. The results for the other three models are generally poor. The macro-level performance level of the models is similar to some of the results published in the literature (*10*, *15*, *48*).

Figure 7 compares the observed and predicted number of home-based tours for various models. Tours with zero stops correspond to walking tours that start and end at home without any stops in between. Models 1 and 2 accurately predict the number of tours with the exception of single-stop tours, where they underpredict the number of tours by 18%. Model 2 also overestimates the number of two-stop tours. Model 2 also slightly overpredicts the number of two-stop tours. The predictions of Models 3, 4, and 5 deviate significantly from the observed number of tours.

In comparison, Model 5 performs the poorest, not picking any trends of activity schedules and also predicting the number of episodes very poorly. The poor performance of the RNN-based model compared with the LSTM-based model emphasizes the presence of long-term dependencies in the data set. Model 3 also performs poorly, predicting a significantly lesser number of activity episodes with long durations, emphasizing the need for rectifying data imbalances. The prediction is significantly improved in the rest of the three models that adopt some method of rectifying data imbalance. Model 4 results are good for education activities but relatively poor for the other activities compared with the Models 1 and 2. Thus we can say that weighted categorical cross-entropy is preferred to the oversampling method for handling data imbalance.

The performance of Models 1 and 2 are similar in predicting the number of episodes, and activity start times and durations for most of the activities. However, Model 1 performs better in predicting the frequency of stay at home, work, and education activities. Further, considering the results of Table 4, we conclude that Model 1 performs better than Model 2, highlighting the presence of bidirectional influence.

In summary, the validation results indicate that the models could successfully predict work, education, and

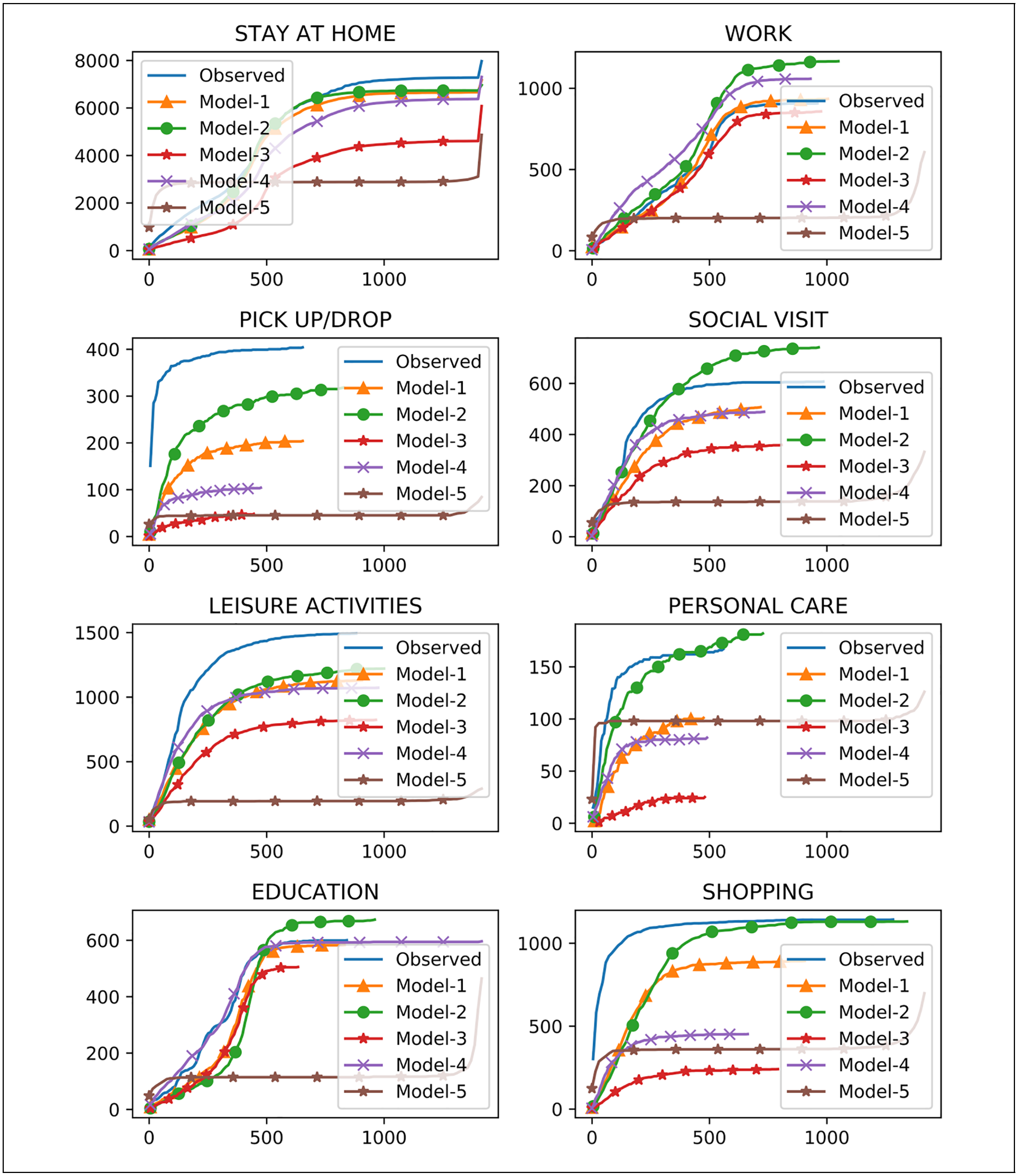

Figure 6. Activity duration patterns during a day (cumulative histogram plot of no. of episodes of a particular type of activity with activity duration in minutes on the X-axis).

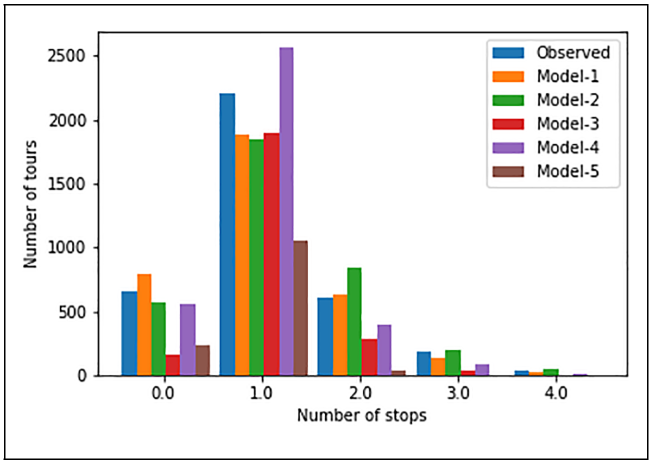

Figure 7. Comparison of number of tours.

stay at home activities. The models underpredict other types of activity but capture the patterns of shopping and other leisure activities very well. Furnishing the model with land-use information could enhance its performance. The models showed good performance in predicting mandatory activity start times and durations. For discretionary and maintenance activities, the models capture the patterns well but fall short in predicting the number of episodes. It was observed that, in general, all models perform poorly while predicting participation and durations of pick-up/drop activity.

This poor performance is because of these activities' short duration, with many episodes being less than 15 min (mean 46.4 min). Further, the household interactions involved in such activities cannot be represented in an individual-level activity schedule model with only individual socio-demographic data (*49*). The same argument can be extended to personal care and shopping activities (mean durations of 83.6 and 69.5 min). The performance can be improved by providing land-use data, opening and closing hours of shops, and so forth.

In general, all models perform poorly on short duration activities. This is evident from Figure 6. The gap between the observed and predicted curves for home, pick-up/drop and shopping activities is high for smaller values of durations. This is a result of these brief duration activities being represented by shorter length sequences in the time series representation, and LSTM, being a data-driven approach, is sensitive to the relative frequencies of each activity in the sequence. As illustrated in this study, the weighted loss function mitigates the problem to a great extent, but there is room for improvement in the case of very short duration actions. A possible approach would be to develop a mechanism to assign short-term dependencies more weight than long-term dependencies, so that short-term actions are also prioritized. This is a direction for future research.

Model Validation for Different Subgroups

Model validation at different sub-levels helps identify subgroups in which the performance is not good and devise ways to improve the model. Table 5 presents a comparison between model results for male and female respondents. Since Model 1 showed the best aggregate-

Table 5. Meso-Level Validation of Output: Comparison of Results Based on Gender

		No. of episodes		Mean activity start time*		Mean activity duration*	
	Activity type	Observed	Predicted	Observed	Predicted	Observed	Predicted
Male	Stay at home	3,755	3,454	102.2 (102.5)	98.5 (106.3)	515.0 (368.5)	491.0 (325.3)
	Work	527	511	123.3 (47.8)	111.9 (39.6)	370.9 (201.3)	410.2 (194.4)
	Pick-up/drop	134	48	162.6 (52.7)	143.0 (51.0)	53.8 (111.4)	134.0 (145.0)
	Social visit	268	231	182.1 (42.5)	180.5 (44.6)	163.5 (123.7)	197.1 (145.4)
	Leisure activities	750	586	180.1 (47.6)	185.3 (51.0)	179.1 (149.1)	224.5 (185.6)
	Personal care	77	51	151.8 (40.9)	142.0 (35.7)	96.9 (114.2)	143.4 (99.8)
	Education	287	287	115.0 (31.8)	107.3 (20.8)	297.0 (145.5)	329.8 (146.3)
	Shopping	456	333	165.2 (37.2)	154.8 (33.4)	72.8 (125.5)	165.2 (132.6)
Female	Stay at home	4,209	3,710	104.4 (100.5)	98.2 (105.3)	509.6 (370.2)	505.9 (331.6)
	Work	380	423	122.1 (39.8)	119.1 (38.9)	370.7 (178.6)	345.6 (182.5)
	Pick-up/drop	270	156	157.4 (45.8)	144.1 (46.3)	42.6 (75.6)	125.8 (115.4)
	Social visit	339	276	178.8 (40.8)	178.9 (41.5)	166.0 (130.1)	208.8 (159.6)
	Leisure activities	743	544	179.8 (46.0)	194.1 (48.0)	148.6 (127.3)	179.0 (144.5)
	Personal care	90	50	150.0 (39.3)	149.5 (40.2)	70.0 (78.7)	130.5 (106.9)
	Education	313	296	117.0 (30.8)	106.5 (14.7)	284.8 (150.5)	350.2 (130.0)
	Shopping	686	560	162.7 (32.8)	150.6 (26.1)	67.3 (103.3)	177.5 (121.5)

Note: No. = number.
*All times in minutes. Numbers in brackets indicate standard deviation.

Table 6. Meso-Level Validation of Output: Comparison of Results Based on Social Participation

		No. of episodes		Mean activity start time*		Mean activity duration*	
	Activity type	Observed	Predicted	Observed	Predicted	Observed	Predicted
Regular worker	Stay at home	3,532	3,126	110.0 (104.4)	103.4 (108.9)	460.6 (338.3)	436.4 (293.6)
	Work	801	853	120.0 (43.5)	112.0 (37.0)	382.9 (191.8)	393.4 (188.9)
	Pick-up/drop	213	102	161.2 (51.3)	150.3 (52.5)	33.4 (56.9)	95.2 (84.4)
	Social visit	228	169	186.3 (43.2)	187.5 (44.9)	169.4 (132.7)	209.5 (149.9)
	Leisure activities	626	478	187.4 (48.7)	201.4 (51.4)	165.1 (134.3)	194.5 (160.9)
	Personal care	74	37	152.7 (42.9)	148.4 (43.8)	62.2 (57.8)	119.3 (76.5)
	Education	28	15	145.5 (50.3)	144.3 (47.9)	102.7 (138.5)	169.7 (114.3)
	Shopping	528	362	168.1 (37.0)	159.7 (32.4)	60.0 (87.2)	164.7 (121.1)
Student	Stay at home	1993	1,771	107.1 (100.0)	101.2 (105.2)	479.5 (338.0)	476.1 (294.7)
	Work	64	56	149.6 (51.4)	158.1 (52.7)	298.5 (164.7)	244.2 (165.4)
	Pick-up/drop	34	12	177.1 (46.8)	143.8 (35.1)	60.1 (123.0)	98.3 (93.7)
	Social visit	143	119	183.5 (41.8)	182.7 (43.4)	163.9 (135.8)	189.7 (145.0)
	Leisure activities	387	319	182.2 (43.9)	190.6 (40.9)	149.3 (120.7)	186.1 (151.9)
	Personal care	20	11	156.4 (41.7)	150.9 (45.7)	129.5 (167.9)	111.8 (99.7)
	Education	453	464	112.9 (27.2)	105.6 (14.1)	303.6 (135.1)	342.5 (130.3)
	Shopping	158	110	176.5 (29.8)	169.9 (22.7)	61.9 (89.5)	130.0 (90.9)
Others	Stay at home	2,439	2,267	90.7 (97.0)	89.2 (101.4)	613.4 (414.1)	602.3 (371.8)
	Work	42	25	135.8 (39.6)	125.6 (24.0)	250.7 (169.9)	263.2 (186.7)
	Pick-up/drop	157	90	152.4 (42.7)	136.6 (41.3)	60.9 (111.0)	168.5 (148.7)
	Social visit	236	219	172.4 (38.4)	171.9 (39.7)	161.1 (116.0)	206.2 (159.9)
	Leisure activities	480	333	168.5 (44.3)	171.5 (49.9)	174.2 (158.2)	230.1 (190.0)
	Personal care	73	53	147.4 (36.2)	142.8 (31.4)	90.0 (98.0)	154.6 (116.8)
	Education	119	104	120.9 (35.4)	107.5 (19.4)	285.3 (166.9)	354.3 (160.4)
	Shopping	456	421	154.1 (30.7)	141.1 (22.6)	83.2 (141.2)	191.2 (134.0)

Note: No. = number.

*All times in minutes. Numbers in brackets indicate standard deviation.

level validation results, we present all the results in this section for that model only. It is seen that the model performance is very similar for the "Male" and "Female" categories in predicting activity participation, start times, and durations.

Table 6 compares prediction results for regular workers, students, and others. The "others" category includes part-time workers, homemakers, and incapacitated and retired individuals. The accuracies in predicting the number of episodes for work are + 6.5%, −12.5%, −40.5% for regular workers, students, and others, respectively. This variation is attributed to the total number of work activity episodes being significantly higher for "regular workers" than for "students" and "others." Similarly, the number of episodes of "education" activity are predicted with much better accuracy for "students" (−2.4%) than for "workers" (−46.4%) and "others" (−12.6%). It is seen that the activity types which are more primary for each subgroup are predicted with better accuracy than others. Similarly, shopping and leisure activities are predicted with much better accuracies for the "others" category, which consists predominantly of a non-working population, than for "regular workers" and "students." Barring these, the model's performance is consistent across the three subgroups.

Table 7 shows the comparison of results for weekdays and weekends, and it is observed that the model performance is mostly similar across both the categories but, in some cases, is slightly better for "weekdays" than for "weekends." Irregular scheduling behavior during weekends makes it difficult for clear patterns to emerge. In some cases, the difference in performance is rather large. For example, the number of episodes of education activity is predicted with an accuracy of −2.4% for weekdays, but for weekends, the accuracy is −25%. The reason for this difference could be the fewer events and irregular participation during the weekend. Overall, the model performance is consistent across different subgroups of the population, with a few exceptions, the reasons for which have been discussed.

The comparison between the observed and predicted number of tours for various subgroups is displayed in Table 8. The number of one-stop tours was underestimated across all subgroups, as was the total number of tours (Figure 7). A few exceptions notwithstanding, the performance is consistent across subgroups. For the below-18 age group, the frequency of two-stop tours is significantly overestimated. This group's number of 0-stop tours is also overestimated. The average SAM distance for this subgroup is likewise on the higher side. The

Table 7. Meso-Level Validation of Output: Comparison of Results Based on the Day of Week

	Activity type	No. of episodes		Mean activity start time*		Mean activity duration*	
		Observed	Predicted	Observed	Predicted	Observed	Predicted
Weekday	Stay at home	5,891	5,290	106.5 (101.2)	101.1 (105.6)	478.9 (346.4)	458.8 (300.0)
	Work	822	858	120.3 (42.7)	111.9 (35.9)	373.0 (192.7)	387.9 (187.9)
	Pick-up/drop	327	174	156.7 (49.2)	142.1 (48.7)	46.7 (92.2)	128.0 (126.1)
	Social visit	361	293	180.3 (41.6)	179.7 (44.0)	156.5 (118.5)	186.8 (147.5)
	Leisure activities	938	761	187.1 (46.5)	198.3 (47.9)	159.3 (136.0)	185.6 (160.6)
	Personal care	157	98	150.9 (40.1)	146.3 (38.5)	81.4 (97.4)	136.3 (104.3)
	Education	588	574	115.6 (31.2)	106.4 (17.4)	292.2 (147.2)	342.8 (137.0)
	shopping	804	621	165.6 (36.2)	154.0 (30.3)	71.6 (119.5)	175.4 (133.6)
Weekend	Stay at home	2,073	1,874	94.4 (101.5)	90.7 (105.9)	606.5 (413.9)	611.5 (376.4)
	Work	85	76	147.7 (54.0)	152.4 (55.4)	349.8 (185.1)	303.0 (217.0)
	Pick-up/drop	77	30	169.4 (42.4)	154.0 (37.3)	44.7 (75.4)	126.0 (104.0)
	Social visit	246	214	180.1 (41.4)	179.5 (41.4)	177.1 (138.4)	226.2 (158.3)
	Leisure activities	555	369	167.9 (44.8)	171.5 (48.6)	171.9 (145.0)	237.7 (179.0)
	Personal care	10	3	149.2 (40.2)	125.7 (12.8)	97.5 (99.3)	161.7 (74.1)
	Education	12	9	139.4 (27.8)	141.1 (19.7)	213.8 (175.2)	174.4 (144.0)
	Shopping	338	272	159.2 (30.4)	147.9 (25.8)	64.5 (94.6)	167.2 (105.8)

Note: No. = number.
*All times in minutes. Numbers in brackets indicate standard deviation.

Table 8. Meso-Level Validation of Output: Comparison of Number of Tours

Subgroup	No. of 0-stop tours		No. of 1-stop tours		No. of 2-stop tours		No. of 3 + stop tours		SAM distances
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	
Male	290	385	1,288	940	284	290	82	46	2.45
Female	373	410	1,468	944	323	341	102	90	2.67
Below 18	84	126	798	513	152	215	31	43	2.28
18–40	335	317	678	451	169	148	60	28	2.86
40–60	244	352	1,280	920	286	268	93	65	2.54
Above 60	70	97	453	367	110	101	33	25	2.36
Worker	344	401	1,190	797	289	281	105	62	2.78
Student	195	198	736	456	137	198	33	35	2.38
Non-worker	124	196	830	631	181	152	46	39	2.42
Weekday	528	591	2,073	1,405	463	525	150	113	2.64
Weekend	135	204	683	479	144	106	34	23	2.36

Note: No. = number; SAM = sequence alignment method.

number of two-stop tours is overpredicted for the students. These observations provide directions for future research to improve the model.

Model Validation Using Time-Use Data

To evaluate the robustness of the proposed method, it is applied to a time-use data set containing information on a wider range of activity types. The data set from the Time Use Survey 2019 (TUS2019) collected by National Statistical Office, Ministry of Statistics and Programme Implementation (MoSPI), Government of India, is used for this purpose. The data set was publicly available and was downloaded from the from the MoSPI website at the following link: https://www.mospi.gov.in/web/mospi/download-tables-data/-/reports/view/templateTwo/20702?q=TBDCAT. The data set contains information on the activity participation and time-use of the Indian population. It provides 24-h activity-diaries (from 04:00 a.m. to 04:00 a.m. of the next day) of a sample of individuals from all over India. In addition to information on activity participation, the data set also contains socio-demographic information about individuals and households. The activity set in the data set is far more detailed with several in-home activities also being recorded under separate categories. In total, the data set contains 141

Table 9. Activity Types in the Time-Use Data set

Activity code	Activity type	Description
0	Traveling	–
1	Employment in corporations, government, and non-profit institutions	–
2	Employment in household enterprises to produce goods, services, or both	Agriculture, animal husbandry, fishing, paid domestic service, and so on
3	Production of goods for own final use	Agriculture, animal husbandry, fishing, and so on
4	Unpaid domestic services for household members	Preparing food, storing, cleaning and maintaining HH, washing, and so on
5	Shopping for own household or members	–
6	Unpaid caregiving services for household members	Medical care, feeding, playing with children, attending to elderly, pick-up/drop, and so on
7	Pick-up/drop	–
8	Volunteering work	Community service
9	Education	At schools or training centers
10	Study related activities	Home-work, reports, studying
11	Socializing	Conversing, chatting, mailing
12	Participating in community cultural/social events	–
13	Religious practices	Private prayers, gatherings
14	Attending/visiting mass cultural, entertainment and sports events/venues	–
15	Cultural participation, hobbies, games, and sports	Arts, hobbies, and so on
16	Participating in sports	Sports, exercising
17	Reading for leisure, listening to audio devices	–
18	Watching television	–
19	Activities associated with reflecting, resting, relaxing	–
20	Sleep and related activities	Night sleep, nap, sleeplessness
21	Eating and drinking	–
22	Personal care	Haircut, makeup, or visit to saloon or parlor
23	Personal health care	Tending to injuries or other health issues, and so on

Note: HH = household.

activity types, which were categorized into 23 activity types for this study (Table 9). The 24-h reference period is divided into 48 30-min intervals, and the activity participation during each interval is recorded. The time series representation for activity sequences will, therefore, also consist of 48 elements, each of which represents a 30-min interval during the day.

The increase in the number of activity types demanded a deeper LSTM network. The architecture consisted of four bidirectional LSTM layers with 300, 200, 200, and 100 units respectively, followed by two dense layers with "ReLu" activation and 200 neurons in each layer. The output layer had 24 neurons and "Softmax" activation. Each of the hidden layers was regularized with 20% dropout. The optimizer used was "Adam" and weighted categorical cross-entropy loss function was used to account for class imbalance in output.

The models' results are shown in Table 10. An average of 13.37 episodes per sequence was observed in the data set, while the predicted average was 12.09. As before, the model slightly underestimates the total number of activities participated in and overestimates the duration of activities. From Table 10, the frequency of "employment at corporations and government offices" activity, which are generally regular is predicted very accurately with −0.62% closeness. In general, the frequencies of activities related to work, education, eating, sleeping, and leisure activities such as watching TV and reading are predicted within ± 10% closeness. However, frequencies of activities which are less regular in nature such as volunteering, social visits, personal health care, and attending mass cultural/social/sports events are predicted with poorer accuracy (difference higher than ± 20%). Figure 8 demonstrates that the majority of activity start time distributions are accurately predicted. Table 10's Kruskal–Wallis test results also indicate that the null hypothesis is accepted at a significance level of 1% for the majority of activity types. From Table 10, the distributions of activity durations are also predicted with good accuracy for the majority of activities, but not so well for activities with short durations, such as eating, personal care, and socializing. Overall, the model's performance on time-use data is found to be comparable to that on activity travel diary data.

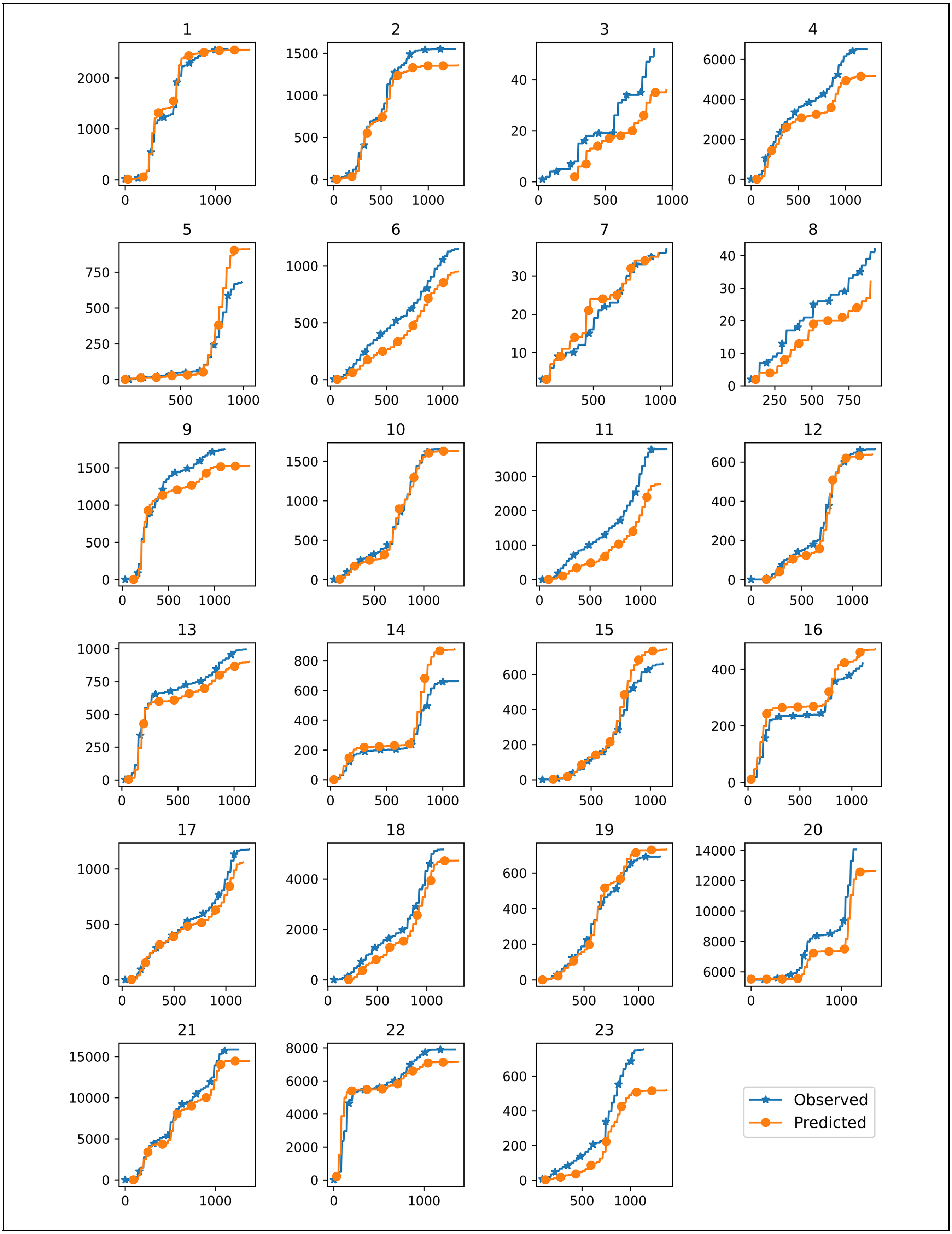

Figure 8. Activity start time patterns during a day. (The X-axis shows time of the day in minutes; the Y-axis shows the cumulative number of individuals starting an activity episode at that particular time.) Titles of the subplots indicate activity codes (refer Table 8).

Table 10. Validation Results for the Time-Use Data Set

Activity code	Frequency		Mean activity start time*		Mean activity duration*		Kruskal-Wallis test p-value	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Start time	Duration
1	2,569	2,553	469.6 (195.2)	441.2 (178.6)	40.8 (16.4)	47.0 (22.0)	0	0
2	1,551	1,354	477.3 (201.5)	459.7 (182.0)	41.0 (26.5)	43.2 (25.7)	0.1087	0.0001
3	52	36	543.5 (248.9)	580.0 (222.5)	29.8 (13.2)	32.4 (32.7)	0.464	0.1025
4	6,520	5,157	533.6 (326.5)	518.0 (325.9)	16.8 (11.2)	18.1 (13.6)	0.0398	0.0244
5	680	912	775.2 (150.3)	793.8 (120.0)	12.4 (5.5)	14.3 (7.2)	0.1428	0
6	1,147	950	623.0 (296.9)	675.4 (273.9)	13.8 (8.4)	19.8 (17.1)	0.0004	0
7	37	36	527.8 (262.0)	475.8 (246.0)	17.6 (15.8)	12.1 (6.4)	0.3765	0.2664
8	42	32	495.7 (263.8)	529.7 (262.2)	20.5 (17.6)	20.2 (15.7)	0.5513	0.8589
9	1,751	1,525	394.1 (237.6)	384.2 (257.3)	34.7 (18.4)	43.8 (22.1)	0.0015	0
10	1,654	1,630	713.5 (239.4)	724.1 (223.6)	20.0 (9.1)	22.3 (14.3)	0.8893	0.0176
11	3,790	2,770	729.5 (300.8)	812.2 (272.2)	12.1 (10.5)	14.9 (14.8)	0	0
12	665	638	682.2 (225.9)	696.3 (200.8)	17.7 (10.2)	19.2 (11.9)	0.4826	0.0539
13	996	901	381.7 (308.6)	382.8 (310.0)	10.2 (8.4)	11.1 (7.8)	0.8512	0.1151
14	663	877	624.2 (309.2)	650.1 (297.8)	13.6 (6.1)	14.8 (7.9)	0.2754	0.0161
15	661	743	715.3 (188.4)	709.6 (176.9)	16.7 (7.9)	20.1 (11.9)	0.4679	0
16	421	471	459.0 (363.7)	450.7 (377.0)	12.6 (6.5)	12.3 (7.0)	0.3489	0.3069
17	1,175	1,058	684.7 (332.9)	683.3 (348.2)	9.6 (6.5)	13.6 (10.2)	0.5501	0
18	5,171	4,729	748.4 (278.8)	807.6 (248.3)	13.0 (7.2)	14.0 (9.8)	0	0.5431
19	692	732	621.5 (210.1)	635.2 (195.4)	14.3 (8.3)	17.1 (12.0)	0.1303	0.0004
20	14,064	12,648	548.3 (479.7)	554.2 (513.4)	31.2 (16.5)	28.2 (16.3)	0	0
21	15,846	14,469	604.6 (315.6)	611.5 (318.8)	8.2 (5.1)	11.2 (6.1)	0.0218	0
22	7,896	7,156	332.6 (324.2)	263.1 (316.9)	8.3 (4.3)	13.6 (6.4)	0	0
23	753	519	727.1 (240.3)	766.5 (191.9)	8.8 (5.7)	11.2 (7.9)	0.2879	0

Conclusions

This paper presents a data-driven approach to model activity schedules that uses time series representation to simultaneously predict activity participation, start times, and duration. The results of this paper provide evidence that the proposed method has the potential to efficiently generate activity schedules to assist in transport planning. The LSTM networks used in this paper, being data-driven, reduce the role of domain expert knowledge in the modeling process and help obtain results based solely on data and observations. Further, the assumptions of sequential decision making are also relaxed, making the model more flexible. In aggregate-level validation, it was observed that the frequency of participation in activities had been predicted within 10% closeness. The Kruskal–Wallis test results showed that activity start times were predicted well by the bidirectional LSTM model for most activities. Among the four models compared, the bidirectional LSTM with weighted categorical cross-entropy loss function performs well in predicting the number of activity episodes and start times. It was also observed that the impact of class imbalance on model performance was significant. Also, bidirectional models performed better than unidirectional models. The model performance on different subgroups was consistent with that at aggregate levels. The validation of a model using time-use data revealed that it can efficiently handle data with a wide variety of activities.

The proposed model can be improved in several ways. Future work should incorporate other aspects of an activity pattern into this model such as location and mode choice. Further, ways to include land-use data, information on work style, working hours, opening and closing hours of shopping centers need to be explored so that the performance on short duration discretionary activities can be improved. More crucially, model transparency needs to be addressed. It may be said that the proposed model which encompasses several activity scheduling decisions and predicts everything in one go is hard to trust as a result of the lack of information on how the model predicts the outputs. The authors argue that this could indeed be an advantage as interpreting these models could help gain insights on the decision-making process not revealed before in the models which restrict themselves by assuming a predefined sequence in decision making. Indeed, the results of a recent study by Zhang et al. (*31*) which exploited time series representation to gain useful insights into the travel behavior, support this argument. Several methods to interpret NN using model weights are available in the literature. The applicability of these methods and modifications, if necessary, need to be explored in future studies.

Acknowledgment

The authors thank Centraal Bureau voor de Statistiek (CBS) and Rijkswaterstaat (RWS-WVL), Netherlands for granting permission to access the ODiN 2016 data set which has been used in this study.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: M Manoj, AP Prathosh, Anil Koushik, N Nezamuddin; analysis and interpretation of results: Anil Koushik; draft manuscript preparation: Anil Koushik, M Manoj, N Nezamuddin, AP Prathosh. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research presented in this paper is a part of the FIRP2019 project number MI02073G funded by the Industrial Research Development (IRD) unit, IIT Delhi. The authors thank IRD, IIT Delhi for this support.

ORCID iDs

Anil Koushik https://orcid.org/0000-0002-2927-8001
M Manoj https://orcid.org/0000-0003-0603-1839
N Nezamuddin https://orcid.org/0000-0002-7441-3989

References

1. Arentze, T., and H. J. P. Timmermans. A Learning-Based Transportation Oriented Simulation System. *Transportation Research Part B: Methodological*, Vol. 38, No. 7, 2004, pp. 613–633.
2. Pinjari, A. R., and C. R. Bhat. Activity-Based Travel Demand Analysis. In *A Handbook of Transport Economics* (A. de Palma, R. Lindsey, E. Quinet, and R. Vickerman, eds.), Edward Elgar Publishing, Cheltenham; Northampton, MA, 2011, pp. 213–248.
3. Miller, E. J.Agent-Based Activity/Travel Microsimulation: What's Next? In *The Practice of Spatial Analysis: Essays in Memory of Professor Pavlos Kanaroglou* (H. Briassoulis, D. Kavroudakis, and N. Soulakellis, eds.), Springer International Publishing, Cham, 2019, pp. 119–150.
4. Koushik, A. N. P., M. Manoj, and N. Nezamuddin. Machine Learning Applications in Activity-Travel Behaviour Research: A Review. *Transport Reviews*, Vol. 40, No. 3, 2020, pp. 288–311. https://doi.org/10.1080/01441647.2019.1704307.
5. Hagenauer, J., and M. Helbich. A Comparative Study of Machine Learning Classifiers for Modeling Travel Mode Choice. *Expert Systems With Applications*, Vol. 78, 2017, pp. 273–282. https://doi.org/10.1016/j.eswa.2017.01.057.
6. Drchal, J., M. Čertický, and M. Jakob. Data-Driven Activity Scheduler for Agent-Based Mobility Models. *Transportation Research Part C: Emerging Technologies*, Vol. 98, 2019, pp. 370–390. https://doi.org/10.1016/j.trc.2018.12.002.
7. Kitamura, R., and S. Fujii. Two Computational Process Models of Activity-Travel Choice. In *Theoretical Foundations of Travel Choice Modeling* (T. Garling, T. Laitila, and K. Westin, eds.), Elsevier, Amsterdam, Netherlands, pp. 251–279.
8. Bowman, J. L., 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, 2001, pp. 1–28. https://doi.org/10.1016/S0965-8564(99)00043-9.
9. Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1894: 57–66.
10. Habib, K. N. A Comprehensive Utility-Based System of Activity-Travel Scheduling Options Modelling (CUSTOM) for Worker's Daily Activity Scheduling Processes. *Transportmetrica A: Transport Science*, Vol. 14, No. 4, 2018, pp. 292–315. https://doi.org/10.1080/23249935.2017.1385656.
11. Lee, M., and M. G. McNally. An Empirical Investigation on the Dynamic Processes of Activity Scheduling and Trip Chaining. *Transportation*, Vol. 33, No. 6, 2006, pp. 553–565. https://doi.org/10.1007/s11116-006-7728-1.
12. Rasouli, S., and H. Timmermans. Activity-Based Models of Travel Demand: Promises, Progress and Prospects. *International Journal of Urban Sciences*, Vol. 18, No. 1, 2014, pp. 31–60. https://doi.org/10.1080/12265934.2013.835118.
13. Garling, T., M.-P. Kwan, and R. Golledge. Computational-Process Household Activity Modelling. *Transportation Research Part B: Methodological*, Vol. 28B, No. 5, 1994, pp. 355–364.
14. Pendyala, R. M., R. Kitamura, A. Chen, and E. I. Pas. An Activity-Based Microsimulation Analysis of Transport Control Measures. *Transport Policy*, Vol. 4, No. 3, 1997, pp. 183–192.
15. Miller, E. J., and M. J. Roorda. Prototype Model of Household Activity-Travel Scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1831: 114–121.
16. Auld, J., and A. Mohammadian. Activity Planning Processes in the Agent-Based Dynamic Activity Planning and Travel Scheduling (ADAPTS) Model. *Transportation Research Part A: Policy and Practice*, Vol. 46, No. 8, 2012, pp. 1386–1403. https://doi.org/10.1016/j.tra.2012.05.017.
17. Ruiz, T., and M. J. Roorda. Assessing Planning Decisions by Activity Type During the Scheduling Process. *Transportmetrica*, Vol. 7, No. 6, 2011, pp. 417–442. https://doi.org/10.1080/18128602.2010.520276.
18. Doherty, S. T., and A. Mohammadian. The Validity of Using Activity Type to Structure Tour-Based Scheduling

Models. *Transportation*, Vol. 38, No. 1, 2011, pp. 45–63. https://doi.org/10.1007/s11116-010-9285-x.
19. Adadi, A., and M. Berrada. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, Vol. 6, 2018, pp. 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.
20. Tsang, M., D. Cheng, and Y. Liu. Detecting Statistical Interactions From Neural Network Weights. *Proc., International Conference on Learning Representations (ICLR)*, Vancouver, Canada, 2018.
21. Vanhulsel, M., D. Janssens, G. Wets, and K. Vanhoof. Simulation of Sequential Data: An Enhanced Reinforcement Learning Approach. *Expert Systems With Applications*, Vol. 36, No. 4, 2009, pp. 8032–8039. https://doi.org/10.1016/j.eswa.2008.10.056.
22. Allahviranloo, M., and W. Recker. Daily Activity Pattern Recognition by Using Support Vector Machines With Multiple Classes. *Transportation Research Part B: Methodological*, Vol. 58, 2013, pp. 16–43. https://doi.org/10.1016/j.trb.2013.09.008.
23. Anda, C., and S. A. Ordóñez Medina. Extending the Hidden Markov Model for Activity Scheduling. *Proc., 6th Symposium of the European Association for Research in Transportation (hEART 2017)*, Haifa, Israel, September 12–14, 2017, pp. 1–6. https://www.research-collection.ethz.ch/handle/20.500.11850/183738
24. Bengio, Y., and P. Frasconi. An Input Output HMM Architecture. *Neural Information Processing Systems*, Vol. 7, 1995, pp. 427–434.
25. Ordóñez Medina, S. A. Inferring Weekly Primary Activity Patterns Using Public Transport Smart Card Data and a Household Travel Survey. *Travel Behaviour and Society*, Vol. 12, 2018, pp. 93–101. https://doi.org/10.1016/j.tbs.2016.11.005.
26. Anda, C., S. A. Ordonez Medina, and K. W. Axhausen. Synthesising Digital Twin Travellers: Individual Travel Demand From Aggregated Mobile Phone Data. *Transportation Research Part C: Emerging Technologies*, Vol. 128, 2021, p. 103118. https://doi.org/10.1016/j.trc.2021.103118.
27. Kato, K., S. Matsumoto, and K. Sano. Microsimulation for Commuters' Mode and Discretionary Activities by Using Neural Networks. *Proc., International Conference on Traffic and Transportation Studies (ICTTS) (2002)*, American Society of Civil Engineers, Reston, VA, 2002, pp. 1290–1297. https://doi.org/10.1061/40630(255)178.
28. Graves, A. *Supervised Sequence Labelling With Recurrent Neural Networks*. Technische Universität München, München, 2008.
29. Castiglione, J., M. Bradley, and J. Gliebe. *Activity-Based Travel Demand Models: A Primer*. Transportation Research Board, Washington, D.C., 2014.
30. Goulias, K. G. Longitudinal Analysis of Activity and Travel Pattern Dynamics Using Generalized Mixed Markov Latent Class Models. *Transportation Research Part B: Methodological*, Vol. 33, No. 8, 1999, pp. 535–558. https://doi.org/10.1016/S0191-2615(99)00005-3.
31. Zhang, J., K. C. Konduri, and A. Enam. Explore Daily Activity-Travel Behavior of the Elderly Using Multiyear Survey Data. In *Mapping the Travel Behavior Genome* (K. G. Goulias, and A. W. Davis, eds.), Elsevier, Amsterdam, Netherlands, 2020, pp. 323–337.
32. Joh, C.-H. H., 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*, 2001. 1752: 16–22.
33. Drchal, J., M. Čertický, and M. Jakob. VALFRAM: Validation Framework for Activity-Based Models. *Journal ofArtificial Societies and Social Simulation*, Vol. 19, No. 3, 2016. https://doi.org/10.18564/jasss.3127.
34. Hochreiter, S., and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
35. Lipton, Z. C., J. Berkowitz, and C. Elkan. A Critical Review of Recurrent Neural Networks for Sequence Learning. 2015, pp. 1–38. http://arxiv.org/abs/1506.00019. Accessed April 10, 2020.
36. Goodfellow, I., Y. Bengio, and A. Courville. *Deep Learning*. The MIT Press, Cambridge, 2016.
37. Staudemeyer, R. C., and E. R. Morris. Understanding LSTM - A Tutorial Into Long Short-Term Memory Recurrent Neural Networks. 2019, pp. 1–42. https://arxiv.org/abs/1909.09586v1.
38. Centraal Bureau voor de Statistiek (CBS) and Rijkswaterstaat (RWS). *Onderzoek Verplaatsingen in Nederland 2016 - OViN 2016 DANS*. 2017.
39. Ranjan, C., S. Ebrahimi, and K. Paynabar. Sequence Graph Transform (SGT): A Feature Embedding Function for Sequence Data Mining. *Data Mining and Knowledge Discovery*, Vol. 36, No. 2, 2022, pp. 668–708. https://doi.org/10.1007/s10618-021-00813-0.
40. Alpaydin, E. *Introduction to Machine Learning*. The MIT Press, Cambridge, 2010.
41. Ma, T., J. Y. J. Chow, and J. Xu. Causal Structure Learning for Travel Mode Choice Using Structural Restrictions and Model Averaging Algorithm. *Transportmetrica A: Transport Science*, Vol. 13, No. 4, 2017, pp. 299–325. https://doi.org/10.1080/23249935.2016.1265019.
42. Tang, L., C. Xiong, and L. Zhang. Spatial Transferability of Neural Network Models in Travel Demand Modeling. *Journal of Computing in Civil Engineering*, Vol. 32, No. 3, 2018. https://doi.org/10.1061/(asce)cp.1943-5487.0000752.
43. Yasmin, F., C. Morency, and M. J. Roorda. Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model. *Transportmetrica A: Transport Science*, Vol. 13, No. 3, 2017, pp. 222–249. https://doi.org/10.1080/23249935.2016.1249437.
44. Wilson, W. C. Activity Pattern Analysis by Means of Sequence-Alignment Methods. *Environment and Planning A: Economy and Space*, Vol. 30, No. 6, 1998, pp. 1017–1038. https://doi.org/10.1068/a301017.
45. Arentze, T., F. Hofman, H. Van Mourik, and H. Timmermans. Spatial Transferability of the Albatross Model System: Empirical Evidence From Two Case Studies. *Transportation Research Record: Journal of the Transportation Research Board*, 2002. 1805: 1–7.

46. Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. Arentze, and H. Timmermans. Integrating Bayesian Networks and Decision Trees in a Sequential Rule-Based Transportation Model. *European Journal of Operational Research*, Vol. 175, No. 1, 2006, pp. 16–34. https://doi.org/10.1016/j.ejor.2005.03.022.
47. Sammour, G., K. Vanhoof, T. Bellemans, D. Janssens, and G. Wets. Investigating the Predictive Performance of Computational Process Activity-Based Transportation Models. *Transportation Planning and Technology*, Vol. 39, No. 6, 2016, pp. 551–573. https://doi.org/10.1080/03081060.2016.1187807.
48. Deng, Y., E. J. Miller, and J. A. Vaughan. Application of Travel Activity Scheduler for Household Agents in a Chinese City. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2429: 121–128.
49. Cirillo, C., and K. W. Axhausen. Dynamic Model of Activity-Type Choice and Scheduling. *Transportation*, Vol. 37, No. 1, 2010, pp. 15–38. https://doi.org/10.1007/s11116-009-9218-8.