Transformer-based Generative Modelling on Activity Sequences with Time-of-Day Choices

Siyu Wang[12], Michel Bierlaire[2*], and Xin Ye[1]

[1]Key Laboratory of Road and Traffic Engineering of Ministry of Education, College of Transportation, Tongji University, China
[2]ENAC, Transport and Mobility Laboratory, EPFL, Switzerland

Short summary

In this study, a generative Multi-Dimensional Transformers (MDT) framework is introduced for activity-based travel demand modelling to enable the joint modelling of daily activity sequences and Time-of-Day (TOD) choices. The limitations of traditional econometric models in capturing high-dimensional correlations and sequential dependencies are addressed by this approach. Built on the Transformer architecture, multi-head attention and masking operations are leveraged to capture nonlinear relationships while customized embedding strategies and activity-temporal sequential matrices are designed to support variable-length sequences. Then the 2019 Household Travel Survey of Shanghai is utilized for the empirical application where real-world activity sequence patterns and TOD distributions are successfully estimated and simulated. Furthermore, the heterogeneous effects of exogenous variables such as gender and household structure on activity-temporal decisions are revealed. Overall, the MDT framework provide a scalable generative tool for simulating complex activity-travel behaviours and supporting transport policy evaluation.
Keywords: Activity-based Modelling, Activity Sequence, Activity Scheduling, Time-of-Day Choices, Transformers.

1 Introduction

In this study, a generative model for activity sequence is developed to jointly represent activity frequencies by type and their associated time-of-day (TOD) choices in sequence. Complete activity sequences within a single day are characterized by daily activity-travel analysis, and the dynamic interplay between travel decisions and temporal constraints is elucidated (Ye et al., 2017). Although temporal decisions are conventionally treated as either discrete or continuous variables (Kun et al., 2025), the discrete modelling of TOD is adopted in the present study.

The joint modelling of activity type and activity timing is increasingly prioritized within the econometric tradition. MDCEV models were widely extended to generate both activity episode counts and episode-level time allocations (Wang & Ye, 2024; Wang et al., 2025). Several frameworks were further proposed to address multidimensional activity-travel choice behaviours. Among these, the ALBATROSS framework was developed with reinforcement and social learning algorithms to dynamically generate activity schedules based on decision trees and adaptive constraints (Arentze & Timmermans, 2004). Furthermore, the OASIS framework was employed to simultaneously optimize utility across activity participation, location, schedule, duration, and travel mode (Janody Pougala et al., 2022; Pougala et al., 2023; Wang et al., 2023). Despite these developments, high-dimensional and nonlinear sequential dependencies among correlated daily activities and their TOD choices are not fully addressed by conventional models, and more flexible modelling approaches are therefore necessitated.

To address these limitations, machine learning models are introduced as a complementary paradigm. Unlike econometric models, loss functions such as mean squared error or maximum likelihood are optimized by machine learning approaches without the requirement for strong prior distributional assumptions (Hillel et al., 2021). Nonlinear and complex interactions among variables are effectively captured within such models, and deep generative models were accordingly proposed and extended to produce sequential outputs in the context of activity scheduling (Shone & Hillel, 2025). In parallel, the capabilities of Transformer architectures in capturing sequential dependencies were underscored by recent developments in large language models (Vaswani et al., 2017). These models are commonly built upon sequence-to-sequence (Seq2Seq) architectures and have demonstrated

strong performance in modelling temporal dynamics. Such architectures are increasingly applied in transportation research (Chen et al., 2025) and show considerable promise in temporal forecasting and joint choice modelling tasks.
However, a fundamental structural distinction is identified between natural language processing applications and activity-based modelling (ABM) contexts. In ABM systems, inputs are typically partially non-sequential as they consist of socioeconomic attributes and individual characteristics, whereas the outputs take the form of sequential results including activity types and temporal information. This input-output asymmetry presents distinct modelling challenges when Transformer-based architectures are applied in ABM settings, and these issues are not adequately addressed through conventional Seq2Seq formulations alone. A multidimensional and sequential framework is therefore required to jointly model activity types, TOD choices, and activity frequencies in an integrated and behaviourally consistent manner. This study aims at proposing a modelling framework for generating sequences of daily activities in association with activity start time choices and their explanatory variables. Two components are included in this framework, which are an estimation section and a simulation section. From a generative modelling perspective, significant implications for the understanding of time-use patterns and travel demand forecasting are offered by the proposed approach. The main contributions are summarized as follows:

(1) A multidimensional Transformer framework, termed Multi-Dimensional Transformers (MDT), is proposed to jointly model activity types, frequencies, and start time choices.

(2) Explanatory variables are incorporated into the encoder without positional encoding, while multidimensional sequence outputs are embedded in the decoder.

(3) Variable-length modelling is adopted with the decision steps, corresponding to activity frequencies as sequence lengths, with TOD decisions and activity categories as dependent variables.
The remainder of this paper is structured as follows. Section 2 presents the methodology of the MDT framework. Section 3 applies the model to the 2019 Shanghai Household Travel Survey, and Section 4 concludes with discussions of findings and implications.

2 Methodology

In this study, TOD is incorporated as a discrete decision dimension with particular emphasis on modelling activity start times. Activity-travel behaviour is represented as a generative process for sequential activity chains. Within this MDT framework, activity types and start times are modelled jointly as two interdependent choice dimensions while activity frequency is treated as the sequence length derived from the data.
A dataset of n individuals is defined where explanatory variables are represented by an input matrix $X \in \mathbb{R}^{n\times a}$. Each vector x_i contains a variables representing influencing factors. The activity frequency for individual i is denoted as l_i and corresponds to the number of activities in their daily sequence. The choice of activity type is encoded in matrix $A_i \in \mathbb{R}^{l_i\times m_A}$ and start time choices in $T_i \in \mathbb{R}^{l_i\times m_T}$ where m_A and m_T represent the number of alternatives for each dimension. Both matrices are one-hot encoded. The combined output is represented by $y_i = [A_i \quad T_i] \in \mathbb{R}^{l_i\times(m_A+m_T)}$ and the final output for the entire dataset is denoted as $Y \in \mathbb{R}^{n\times l_i\times(m_A+m_T)}$.

Overview of the MDT framework

The MDT architecture is fundamentally based on the original Transformer structure although specific adaptations are included to model multi-dimensional sequential decisions. The overall framework is illustrated in Figure 1.

- **Encoder and Decoder Modules:** Both the encoder and decoder are composed of Multi-Head Attention (MHA) and Feed-Forward Networks (FFN) where each component is followed by a residual connection and layer normalization. Complex variable correlations and sequential dependencies are captured by the MHA and FFN structure without reliance on strong distributional assumptions. Furthermore, causal masking and a shifted-right operation are employed by the decoder to facilitate autoregressive sequence generation.
- **Embedding of Output Sequences:** The sequential output Y_i is formed by the concatenation of the activity type and start time vectors. To address the critical task of variable-length sequences, three interpretable placeholder tokens are introduced:

1. The $\langle bocc \rangle$ (begin-of-choice-chain) placeholder consisting of all ones is prepended to each sequence to initiate autoregressive decoding as the initial input to the decoder and to enable causal masking.

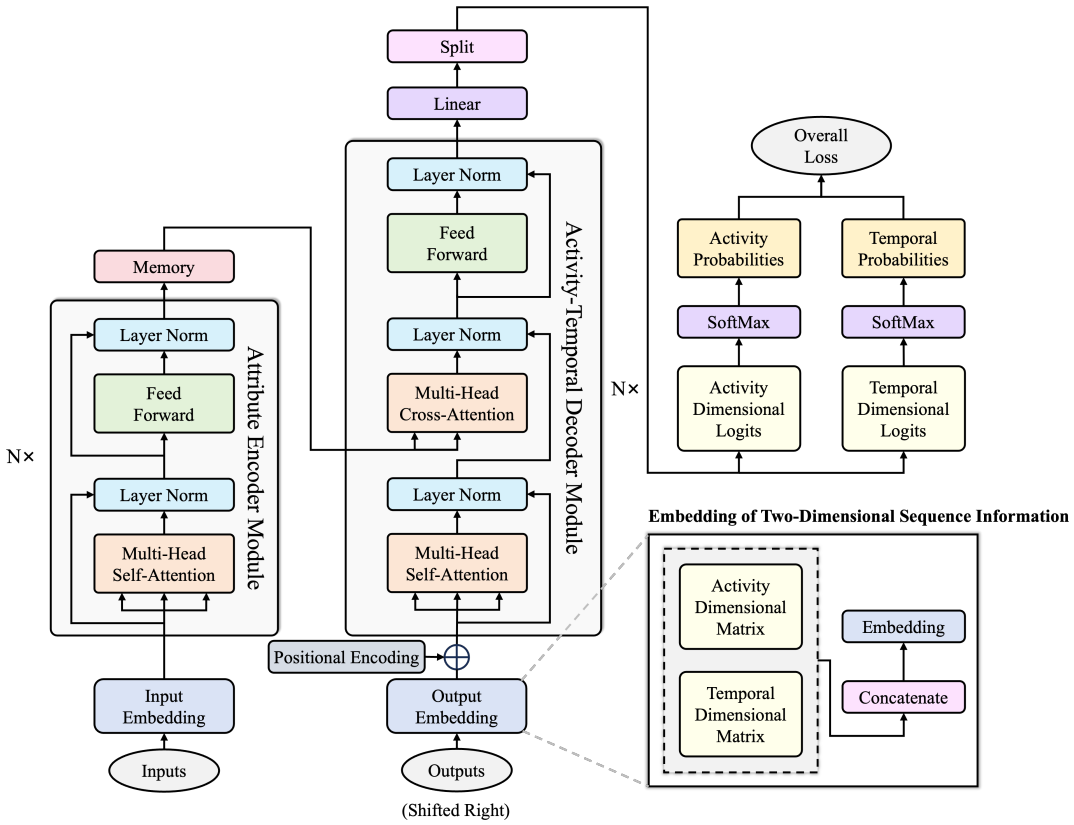


Figure 1: Overall framework of MDT

2. The $\langle eocc \rangle$ (end-of-choice-chain) placeholder representing the end of the choice chain is inserted after the last valid activity choice to signal the termination of the activity chain. This token occupies an extra dimension in the activity choice space.

3. The $\langle pocc \rangle$ (padding-of-choice-chain) placeholder consisting of all zeros is utilized to align all sequences in a batch to a uniform maximum length l_{max}.

It is ensured that matrix dimensions remain consistent during parallel computation without attention scores through this padding strategy.It is ensured by the model that $\langle eocc \rangle$ and $\langle pocc \rangle$ placeholders are masked out during both attention computation and loss aggregation so that training is focused exclusively on valid decision steps.

Definition of loss function

For the variable-length sequences addressed in this study, where each individual i corresponds to a distinct sequence length l_i, step-wise probability distributions over activity types and start time choices are produced by the decoder, formulated as:

$$P_{A_i}^{(s)} = \text{Softmax}\left(z_{A_i}^{(s)}\right), P_{T_i}^{(s)} = \text{Softmax}\left(z_{T_i}^{(s)}\right) \tag{1}$$

As expressed in Equation (1), $P_{A_i}^{(s)}$ and $P_{T_i}^{(s)}$ are obtained via Softmax transformations of the output logits $z_{A_i}^{(s)}$ and $z_{T_i}^{(s)}$ from the decoder for each decision step s. The per-sample loss $\mathcal{L}_i$ is computed as the normalized sum of cross-entropy losses across valid decision steps:

$$\mathcal{L}_i = -\frac{1}{m_A(l_i+1)+m_T l_i} \times \left(\sum_{s=1}^{l_i+1} \log P_{A_i}^{(s)}(a_i^{(s)}) + \sum_{s=1}^{l_i} \log P_{T_i}^{(s)}(t_i^{(s)}) \right) \tag{2}$$

where l_i+1 includes the additional $\langle eocc \rangle$ placeholder for the activity category, while time predictions stop at l_i since $\langle eocc \rangle$ is not associated with time. The total loss over n individuals is:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_i \tag{3}$$

All padding placeholders $\langle pocc \rangle$ are masked out during both attention computation and loss aggregation. This ensures that model training focuses only on valid decision steps and is unaffected by differences in sequence lengths.

Generation procedure of synthetic activity schedules

To evaluate the applicability of the trained MDT model within ABM contexts, frequency distributions of generated sequences are obtained through numerical random simulation. The data generation of the synthetic group is implemented via a Gumbel-Softmax inverse sampling procedure which consists of the following three core computational steps:

• Step 1: Random Disturbance Generation: Uniform random numbers $U_k \sim \text{Uniform}(0, 1)$ are generated to simulate standard Gumbel noise through the transformation $G_k = -\log(-\log(U_k))$.

• Step 2: Utility Computation: The latent utility for each alternative is calculated by adding the Gumbel noise to the raw log-odds produced by the Transformer model where this combined value is expressed as $\text{logit}_j + G_k$.

• Step 3: Probabilistic Mapping and Selection: A SoftMax transformation is applied to these combined values to derive differentiable choice probabilities. The final choice outcome is determined by applying the argmax function to select the alternative with the maximum utility.

This procedure effectively approximates the inverse transform sampling of a discrete distribution while remaining compatible with the gradient-based Transformer framework. Complex activity-travel behaviours are simulated by preserving the probabilistic mapping principles inherent in discrete choice modelling. Because the test set is not utilized during the training procedure, this group is considered the synthesis input for the empirical study.

3 RESULTS AND DISCUSSION

The empirical analysis utilizes data from the 2019 Shanghai Household Travel Survey (SHTS). Since the end time of a previous trip corresponds to the start time of the current activity, trip-based survey data are converted into individual-level daily activity chain information. Given the greater diversity and flexibility in daily schedules among non-commuters, this study focuses on this group to better capture heterogeneity in activity patterns. After the filtering process is completed, data from 35,711 individuals comprising 125,422 daily activity records are selected for the study.

Model specification and performance

The model architecture includes 4 encoder layers and 4 decoder layers where each is equipped with 4 multi-head attention mechanisms. The embedding dimension is set to 128 and the Feed-Forward Network (FFN) within each Transformer layer has a dimensionality of 512. The dataset is randomly divided into a 70% training set, a 20% validation set, and 10% testing subsets. After 63 iterations, the model converges and reaches a final loss value of 0.698 on the entire dataset. Table 1 compares the frequencies of the top 10 most common activity chains and associated TOD decisions between the true SHTS dataset and the MDT-generated outputs.

From Table 1, the trained MDT model accurately reproduces the structural distribution of activity sequences and chain complexity as shown in Table 1. The frequencies of the top 10 activity chains in the trained MDT model align closely with the observed data where the framework correctly identifies both simple and complex sequences with high precision. In terms of overall generation intensity, the average activity frequency per person is 3.53 in the synthetic results compared to 3.51 in the true dataset and this proximity indicates that the Transformer architecture successfully internalizes the underlying scheduling logic.

Generation results

For the generation part, the test set is withheld from the model training process to evaluate the generative capacity of the proposed framework. Following the allocation of 10% of the total sample, explanatory variables representing individual characteristics and socioeconomic attributes for 3,572 individuals are fed into the encoder. The trained MDT framework is then applied to simulate activity sequences for this group where the simulated distributions are obtained via inverse sampling. Table 2 presents the comparison between the observed test data and the generated results for the top 10 activity chains.

Table 1: Comparison of observed and trained activity sequence distributions.

Activity and TOD Decisions	Frequency	
Activity Chain with Start Time (Top 10 Chains)	Observed Data	Trained MDT
(At Home, 00:00–02:00) → (Shopping, 08:00–10:00) → (At Home, 08:00–10:00)	3216	2607
(At Home, 00:00–02:00) → (Shopping, 06:00–08:00) → (At Home, 08:00–10:00)	2704	2473
(At Home, 00:00–02:00) → (Shopping, 06:00–08:00) → (At Home, 06:00–08:00)	1583	1849
(At Home, 00:00–02:00) → (Shopping, 08:00–10:00) → (At Home, 10:00–12:00)	1438	1295
(At Home, 00:00–02:00) → (Others, 08:00–10:00) → (At Home, 10:00–12:00)	620	654
(At Home, 00:00–02:00) → (Recreation, 06:00–08:00) → (At Home, 08:00–10:00)	606	778
(At Home, 00:00–02:00) → (Recreation, 08:00–10:00) → (At Home, 10:00–12:00)	575	678
(At Home, 00:00–02:00) → (Recreation, 10:00–12:00) → (At Home, 10:00–12:00)	415	497
(At Home, 00:00–02:00) → (Others, 06:00–08:00) → (At Home, 08:00–10:00)	399	459
(At Home, 00:00–02:00) → (Others, 08:00–10:00) → (At Home, 08:00–10:00)	379	385
Different Type of Activity Sequences		
Simple Sequences (Activity Number $\leqslant 3$)	25628	25388
Complex Sequences (Activity Number > 3)	10083	10323
Overall Activity Frequency	125422	126031
Average Activity Frequency per Person	3.51	3.53

Based on the comparison in Table 2, the joint distribution of activity types and start times demonstrates a high degree of statistical consistency between the observed test data and the generated sequences. To provide a rigorous and objective assessment beyond the Top-10 patterns, the full sequence space is also evaluated, which is composed of over 1,500 unique chains, using standard distributional similarity metrics. The Jensen–Shannon Divergence (JSD) is 0.1610, the Total Variation Distance (TVD) is 0.3278, and the Overlap Coefficient reaches 0.6722. These quantitative measures suggest that the MDT framework effectively captures the underlying probability surface and the complex scheduling logic inherent in the empirical patterns. Meanwhile, the model is capable of synthesizing not only dominant behaviors, such as "At Home-Shopping-At Home" sequences, but also reproducing the long-tail characteristics. As indicated in Table 2, the discrepancy in the remaining categories, which represent approximately 65% of the total probability mass, is limited to 2.46 percentage points.

Given the inherent stochasticity of human activity-travel behavior and the high cardinality of the sequence space, these results confirm that the MDT model reproduces the empirical data with high statistical fidelity. This alignment is further validated by the marginal distributions shown in Figure 2. The results confirm that the MDT model accurately reproduces the empirical data in terms of both overall activity categories and frequency estimations, effectively representing the

Table 2: Comparison of observed test set and generated sequences

Activity and TOD Decisions	Frequency	
Activity Chain with Start Time (Top 10 Chains)	Observed Test Data	Generated Result
(At Home, 00:00–02:00) → (Shopping, 08:00–10:00) → (At Home, 08:00–10:00)	333	263
(At Home, 00:00–02:00) → (Shopping, 06:00–08:00) → (At Home, 08:00–10:00)	272	261
(At Home, 00:00–02:00) → (Shopping, 06:00–08:00) → (At Home, 06:00–08:00)	167	178
(At Home, 00:00–02:00) → (Shopping, 08:00–10:00) → (At Home, 10:00–12:00)	139	134
(At Home, 00:00–02:00) → (Others, 08:00–10:00) → (At Home, 10:00–12:00)	65	55
(At Home, 00:00–02:00) → (Recreation, 06:00–08:00) → (At Home, 08:00–10:00)	62	69
(At Home, 00:00–02:00) → (Recreation, 08:00–10:00) → (At Home, 10:00–12:00)	56	68
(At Home, 00:00–02:00) → (Others, 08:00–10:00) → (At Home, 08:00–10:00)	46	55
(At Home, 00:00–02:00) → (Shopping, 10:00–12:00) → (At Home, 10:00–12:00)	44	53
(At Home, 00:00–02:00) → (Recreation, 8:00–10:00) → (At Home, 08:00–10:00)	33	41

full distribution spectrum of urban mobility.

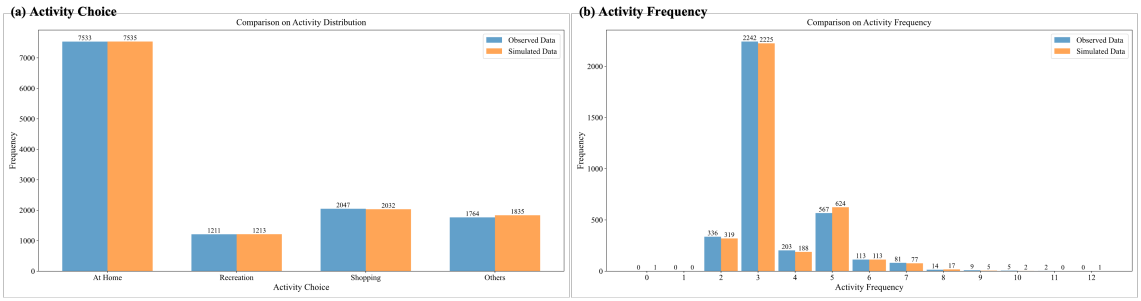


Figure 2: Comparison on activity categories and frequencies between observed SHTS and simulated results via MDT

On the other hand, Figure 3 provides a detailed comparison of the start time distributions for typical activity types between the observed and simulated data. The overall discrepancy remains relatively small, and this consistency demonstrates that the MDT framework offers a valid simulation approach for complex activity scheduling. Both shopping and recreation activities exhibit a primary peak between 06:00 and 10:00 while a secondary peak for recreation is observed in the 18:00 to 20:00 interval. Although the model somewhat underestimates the proportion of recreation activities during the evening peak and slightly overestimates other activity types, the framework effectively generates reasonable schedule distributions for the target population. These results confirm that the Transformer architecture successfully captures the temporal dependencies required for high-fidelity travel demand synthesis.

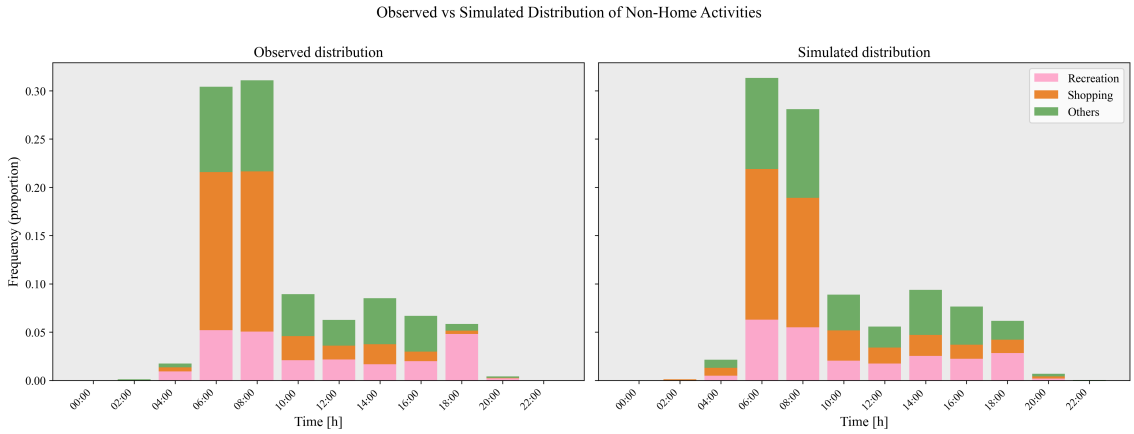


Figure 3: Comparison of empirical and simulated distribution of activities in SHTS

Marginal effect analysis

A key advantage of the MDT architecture is that the influence of covariates such as socioeconomic attributes is explicitly incorporated into the analysis while simultaneously leveraging the sequential modelling and simulation capabilities of the Transformer. Based on the trained model and descriptive statistics, marginal effect analysis is conducted for three representative exogenous variables including age, gender, and employment or education status. The baseline scenario is defined as a 60-year-old employed male from an all-elderly household with school-age children who resides in Shanghai and exhibits an average daily travel distance of 4.57 km.

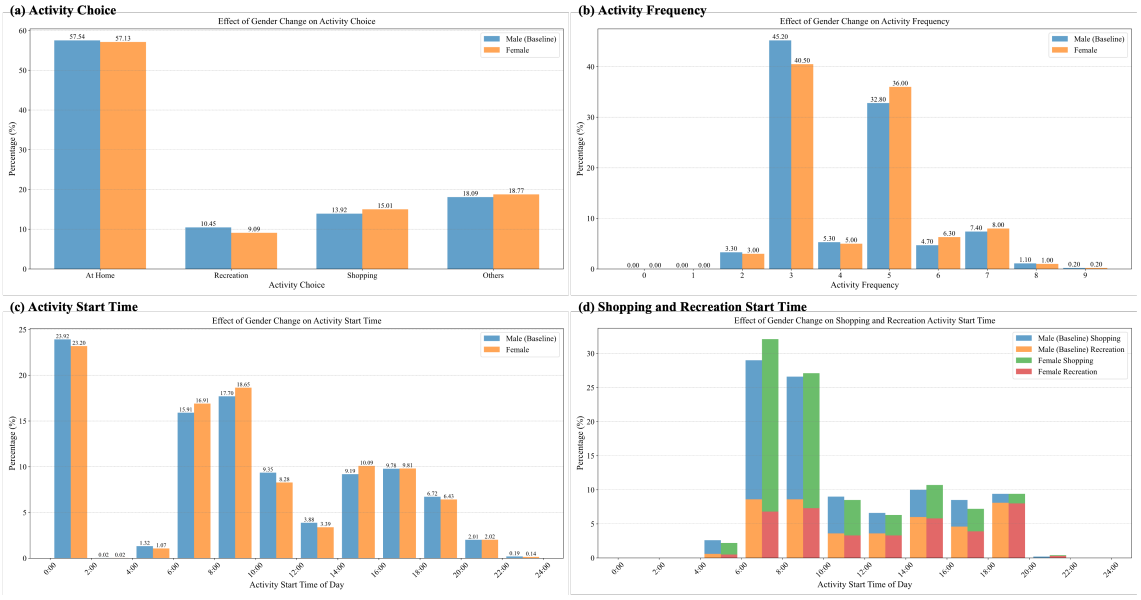


Figure 4: Marginal effects of gender on activity-temporal attributes

With gender as the primary variable, Figure 4 highlights the impact of individual characteristics on activity-temporal behaviour. Figure 4(a) indicates that women participate more in shopping activities whereas men engage more in recreational activities. In terms of activity frequency shown in Figure 4(b), males exhibit simpler chains with fewer daily activities while females tend to engage in more complex sequences with higher activity participation. Figure 4(c) reveals that activity start times for both genders follow a bimodal distribution where women have a higher proportion of activities between 06:00 and 10:00 compared to the male preference for the 10:00 to 14:00 interval. Furthermore, Figure 4(d) shows that while both genders prefer morning shopping and evening recreation, women maintain a higher proportion of shopping during morning hours. Overall, gender exhibits significant heterogeneity in the categorical and temporal distribution of daily activities within the synthetic population.

4 Conclusions

The Multi-Dimensional Transformer (MDT) achieves the joint modelling of activity types, frequencies, and Time-of-Day (TOD) choices. This framework represents a successful trade-off between econometric insights and machine learning in capturing high-dimensional and sequential dependencies. By leveraging the attention mechanism, the MDT excels in complex choice environments where traditional methods struggle. The integration of custom placeholders such as ⟨eocc⟩ acts as an endogenous decision step which ensures both accuracy and flexibility when generating variable-length activity chains.

The superior performance of the MDT model is validated using Shanghai Household Travel Survey (SHTS) data through the accurate reproduction of joint activity-temporal distributions. The model successfully quantifies behavioural heterogeneity and reveals significant patterns such as the heightened schedule rigidity observed in employed individuals compared to retirees. These findings demonstrate the practical value of the MDT framework as a powerful simulation tool for transportation planners. The model enables more precise demand forecasting than static models and effectively supports the design of efficient urban transportation interventions.

Despite these strengths, the MDT framework requires further development in several areas. The current discrete representation of time should be expanded to continuous time modelling for greater precision. Furthermore, the framework remains to be extended to additional choice dimensions including destination and mode choices, and given more systematic comparisons and integrations with econometric approaches such as OASIS framework. Such advancements facilitate progress toward a truly integrated and comprehensive next-generation Activity-Based Modelling system that captures the full complexity of urban mobility.

References

Arentze, T. A., & Timmermans, H. J. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, *38*(7), 613–633. (ISBN: 0191-2615)

Chen, Y., Chi, B., Li, C., Zhang, Y., Liao, C., Chen, X., & Xie, N. (2025). Toward interactive next location prediction driven by large language models. *IEEE Transactions on Computational Social Systems*. Retrieved 2025-07-25, from `https://ieeexplore.ieee.org/abstract/document/10835157/`

Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y. (2021, March). A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling*, *38*, 100221. Retrieved 2025-02-03, from `https://www.sciencedirect.com/science/article/pii/S1755534520300208` doi: 10.1016/j.jocm.2020.100221

Janody Pougala, Tim Hillel, & Michel Bierlaire. (2022, June). Capturing trade-offs between daily scheduling choices. *Journal of Choice Modelling*, *43*, 100354. Retrieved 2025-02-14, from `https://www.sciencedirect.com/science/article/pii/S1755534522000124` doi: 10.1016/j.jocm.2022.100354

Kun, H., Ye, X., & Geng, S. (2025, March). *A Univariate Probability Density Model Based on Partially Monotone Neural Network —— an Application to Time-of-Day Choices of Shopping Activities* [SSRN Scholarly Paper]. Rochester, NY: Social Science Research Network. Retrieved 2025-07-25, from `https://papers.ssrn.com/abstract=5192839` doi: 10.2139/ssrn.5192839

Pougala, J., Hillel, T., & Bierlaire, M. (2023, October). OASIS: Optimisation-based Activity Scheduling with Integrated Simultaneous choice dimensions. *Transportation Research Part C: Emerging Technologies*, *155*, 104291. Retrieved 2026-01-20, from `https://www.sciencedirect.com/science/article/pii/S0968090X23002802` doi: 10.1016/j.trc.2023.104291

Shone, F., & Hillel, T. (2025, October). Synthesising activity participations and scheduling with deep generative machine learning. *Transportation Research Part C: Emerging Technologies*, *179*, 105273. Retrieved 2026-03-10, from `https://www.sciencedirect.com/science/article/pii/S0968090X25002773` doi: 10.1016/j.trc.2025.105273

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, M., Pougala, J., & Bierlaire, M. (2023). Initial Comparisons Between MDCEV model and OASIS Framework. In *23rd Swiss Transport Research Conference (STRC).*

Wang, M., & Ye, X. (2024). Development of Multiple Discrete-Continuous Extreme Value Model with Non-Monotonic Utilities: Formulation and Application to Outdoor Non-Mandatory Time-Use Decisions. *Transportation Research Record*, 03611981241230533. (ISBN: 0361-1981 Publisher: SAGE Publications Sage CA: Los Angeles, CA) doi: https://doi.org/10.1177/03611981241230533

Wang, M., Ye, X., & Wang, K. (2025). A Modification of Multiple Discrete-Continuous (MDC) Choice Model to Consider Nonmonotonic Preference in Episode-Level Time-Use Behaviors. *Journal of Advanced Transportation*, *2025*(1), 7114605. (ISBN: 2042-3195)

Ye, X., Garikapati, V. M., You, D., & Pendyala, R. M. (2017). A practical method to test the validity of the standard Gumbel distribution in logit-based multinomial choice models of travel behavior. *Transportation Research Part B: Methodological*, *106*, 173–192. doi: https://doi.org/10.1016/j.trb.2017.10.009