



# How do you go where? Improving next location prediction by learning travel mode information using transformers

Ye Hong  
hongy@ethz.ch  
Institute of Cartography and  
Geoinformation, ETH Zurich  
Zurich, Switzerland

Henry Martin  
martinhe@ethz.ch  
Institute of Cartography and  
Geoinformation, ETH Zurich  
Zurich, Switzerland  
Institute of Advanced Research in  
Artificial Intelligence (IARAI)  
Vienna, Austria

Martin Raubal  
mraubal@ethz.ch  
Institute of Cartography and  
Geoinformation, ETH Zurich  
Zurich, Switzerland

## ABSTRACT

Predicting the next visited location of an individual is a key problem in human mobility analysis, as it is required for the personalization and optimization of sustainable transport options. Here, we propose a transformer decoder-based neural network to **predict the next location an individual will visit based on historical locations, time, and travel modes**, which are **behaviour dimensions often overlooked in previous work**. In particular, the prediction of the next travel mode is designed as an auxiliary task to help guide the network's learning. For evaluation, we apply this approach to two large-scale and long-term GPS tracking datasets involving more than 600 individuals. Our experiments show that the proposed method significantly outperforms other state-of-the-art next location prediction methods by a large margin (8.05% and 5.60% relative increase in F1-score for the two datasets, respectively). We conduct an extensive ablation study that quantifies the influence of considering temporal features, travel mode information, and the auxiliary task on the prediction results. Moreover, we experimentally determine the performance upper bound when including the next mode prediction in our model. Finally, our analysis indicates that the performance of location prediction varies significantly with the chosen next travel mode by the individual. These results show potential for a more systematic consideration of additional dimensions of travel behaviour in human mobility prediction tasks. The source code of our model and experiments is available at <https://github.com/mie-lab/location-mode-prediction>.

## CCS CONCEPTS

• **Information systems** → **Geographic information systems**; *Location based services*; • **Computing methodologies** → **Neural networks**; • **Applied computing** → *Transportation*.

## KEYWORDS

Mobility, Deep learning, Location prediction, Travel behaviour

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGSPATIAL '22, November 1–4, 2022, Seattle, WA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9529-8/22/11...\$15.00

<https://doi.org/10.1145/3557915.3560996>

## ACM Reference Format:

Ye Hong, Henry Martin, and Martin Raubal. 2022. How do you go where? Improving next location prediction by learning travel mode information using transformers. In *The 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, November 1–4, 2022, Seattle, WA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3557915.3560996>

## 1 INTRODUCTION

The rapid urbanization process in the last decades has caused a constant increase in individual travel, imposing significant challenges in achieving sustainable cities. To meet the sustainable development goals of the United Nations [12], mobility behaviour change and new mobility concepts to promote these changes will play indispensable roles [26]. These mobility concepts, such as mobility as a service (MaaS) [30], smart charging [43] and ride-sharing [16], all rely on the capability to proactively provide personalized services that are tailored to the travel context and individuals' characteristics [23]. Individual mobility prediction to know when and where travel will occur is a crucial technique driving the development and application of these new concepts and, therefore, a key technology for sustainable transportation.

The prediction of where an individual will go given her historical mobility information is central to individual mobility prediction. The problem is also known as the next location prediction and has attracted much attention over the last decade. Researchers are increasingly interested in tackling this problem using learning-based methods thanks to the booming of deep learning (DL) models [22]. As it can be formulated as a **sequence prediction problem**, similar to the tasks encountered in natural language processing and audio processing, models that have shown success in these two fields are often directly applied. In particular, the transformer model [39] that utilizes a multi-head self-attention mechanism has revolutionized various sequence modelling tasks due to its powerful and efficient network structure. Transformer models are also starting to gain attention in predicting individual mobility, as they tackle some challenges in mobility prediction by design: (1) Multiple periodicities co-exist in the location visitation patterns of individuals [9] (e.g., daily, weekly). These periodicities vary considerably across individuals. The multi-head self-attention module allows the network to focus on multiple steps in the input sequence, effectively capturing these periodicities. (2) The long-term dependency of mobility behaviour. Studies have shown that the current mobility depends on behaviours conducted days or weeks before [4, 37], which requires

the prediction model to capture long-term dependencies. The design of the transformer model enables efficient learning of these dependencies. However, human mobility also exhibits unique characteristics, such as complex spatio-temporal dependencies [9, 20] and the inherent stochasticity of location visits [36], which hinder the performance when applying sequence learning models on raw location visit sequences. Therefore, learning to predict the next location directly from historical location visits is challenging. An accurate prediction model should consider context information that influences individuals' choice of locations.

Results from travel behaviour studies that aim to understand individuals' activity location choices could guide the consideration of context information. Empirical evidence suggests that the selection of activity locations is highly correlated to other aspects of individual travel behaviour, such as the availability of travel modes [27] and the day of the week [7]. However, the comprehensive information regarding individuals' travel behaviour is not fully utilized in location prediction problems. **To date, it is still unclear (1) how strong the influence of these long-term factors is on choosing the immediate next location and (2) whether the DL network can benefit from this knowledge and learn the complex dependency patterns directly from data.**

To close this research gap and answer the above questions, we propose a transformer-based model that utilizes historical travel behaviour to predict individuals' next location. More precisely, the model **aims to learn mobility transition patterns from historical location, temporal and travel mode sequence information.** Inspired by travel behaviour studies, we encourage the model to also predict the next travel mode the individual will choose. We anticipate that this ancillary task will help the prediction of the next location. Through experiments on two real-world GPS datasets, we demonstrate the effectiveness of our model design and quantify the dependency of location prediction performance on travel mode. Our results show that careful consideration of individual travel behaviour significantly benefits human mobility prediction. In short, our contributions are summarized as follows:

- We propose a transformer decoder-based neural network that utilizes location, travel mode and time-related information for the next location prediction task. The proposed model achieves state-of-the-art performance.
- We show that jointly learning the next location and next mode improves the prediction performance for both tasks.
- We conduct extensive experiments on two real-world GPS tracking datasets and conclude that considering additional aspects of travel behaviour significantly increases the performance of next location prediction.

The rest of this paper is organized as follows. We first systematically review related work in Section 2. In Section 3, we formulate the next location prediction problem. Next, we introduce details of the network architecture in Section 4. We apply our model to two real-world GPS datasets and analyze its performance in Section 5. Finally, we summarize the main findings and conclude the paper in Section 6.

## 2 RELATED WORK

### 2.1 Next location prediction

The next location prediction problem has found application in many different fields, such as recommendation systems [44], sensor networks [28], and mobility behaviour analysis [41, 42]. The exact definition of the problem varies across studies due to different objectives and employed datasets. For example, location-based social network (LBSN) applications focus on predicting the next check-in point-of-interest (POI) [40, 44]. In contrast, mobility behaviour studies aim to understand the next location for a user to conduct an activity [34]. Here we focus on the methods proposed for mobility applications.

The last decade has witnessed the expansion of studies focusing on next location prediction. Markov Chain and its variants are probably the most often employed methods for the task [22]. These models regard locations as states and construct a transition matrix that encodes the transition probability between states for each individual. Ashbrook and Starner [1] and Gambs et al. [11] both proposed identifying significant locations from GPS data and building a Markov model to predict location transitions. Later Markov model variants that consider collective movements [3] and incorporate location importance [17] further increased the prediction performance. However, Markov-based models struggle to represent the complex sequential patterns in human mobility because of their inherent assumption that the current state only depends on the states of previously limited time steps [20].

Recent advances in DL have also promoted their application in location prediction. As a widely adopted sequence modelling method, recurrent neural network (RNN)-based models, such as Long Short-Term Memory (LSTM) [34] and spatial-temporal (ST)-RNN [21], were reported to outperform Markov models by a large margin in the task. Still, vanilla RNN models tend to underweight long-term dependencies when the input sequence length increases. Therefore, studies employed the attention module to capture both short-term and long-term dependencies dynamically [9, 20]. Moreover, the transformer model that builds on top of the multi-head self-attention mechanism [39] have started to gain interest in the field. In particular, Xue et al. [44] proposed MobTcast for considering various contexts with a transformer-based structure and achieved state-of-the-art POI prediction results for LBSN data. Although having great potential in learning the complex spatio-temporal dependencies, limited studies have applied transformer for the location prediction problem.

### 2.2 Factors affecting activity location choice

Understanding the factors affecting activity location choice is beneficial for predicting individuals' mobility, as they can be regarded as prior knowledge and potentially guide the learning of DL models. In the travel behaviour field, the choice of locations is regarded as an integral part of individuals' activity-travel behaviour and has been studied within the activity-based framework [32]. Studies that focus on analysing travel behaviour over time suggest that both stability and variability are found in individuals' activity location choices. For example, Dharmowijoyo et al. [6] showed that the variability of location visits is much larger between weekend-weekday pairs

than between weekday-weekday and weekend-weekend pairs. Empirical studies also demonstrate the correlation of different aspects of individual travel behaviour. For example, Susilo and Axhausen [38] reported high repetition in location-mode combinations, suggesting that individuals use the same travel mode to reach their locations. Similar conclusions were reported by Hong et al. [14], where they found that only a subset of all location-mode combinations is essential for describing the mobility behaviour. From this perspective, aspects of travel behaviour can be considered constraints for individuals' choice of activity locations.

A similar problem as the next location prediction is the formulation of an individual's location choice set, which is a crucial component in microscopic traffic simulation models [19]. Instead of predicting the exact next location, the problem aims at generating a set containing all possible locations. Based on time geography theory, potential path areas analysis has been applied to tackle the problem, suggesting that the choice set is constrained by the travel time [33], time of day [45] and the available travel mode [27]. However, this empirical knowledge is not fully utilized in models for location predictions.

Building on the previous studies, we utilize a transformer-based model for learning spatio-temporal dependencies in individual location visits. In addition, we consider other aspects of travel behaviour that constrain an individual's activity location choices in the learning process.

### 3 PROBLEM FORMULATION

We formulate the next location prediction problem with the following definitions:

**DEFINITION 1 (TRAJECTORY).** Let  $u^{(i)}$  be a user in a set of users  $\mathcal{U} = \{u^{(1)}, u^{(2)}, \dots, u^{(|\mathcal{U}|)}\}$ , a trajectory  $T^{(i)} = [q_1, q_2, \dots, q_{n_{u^{(i)}}}]$  is a time-ordered sequence composed of  $n_{u^{(i)}}$  track points visited by  $u^{(i)}$ . A track point can be represented as a tuple of  $q = \langle p, t \rangle$ , where  $t$  records the time when the user visits, and  $p = \langle x, y \rangle$  represents spatial coordinates in a given reference system, e.g., latitude and longitude.

**DEFINITION 2 (LOCATION SEQUENCE).** Location is defined when a user remains within a certain geographical radius for a defined time, and is a sub-sequence of the user's trajectory. A location sequence  $S^{(i)} = [L_1, L_2, \dots, L_{w_{u^{(i)}}}]$  is a time-ordered sequence composed of  $w_{u^{(i)}}$  locations visited by  $u^{(i)}$ . Each location  $L = \langle l, t, e \rangle$  is described by the arrival time  $t$ , the (main) travel mode  $e$  used to reach that location, and the identifier  $l \in O$ , where  $O$  is the set containing all known locations.

**PROBLEM 1 (NEXT LOCATION PREDICTION).** Consider the historical location sequence  $S_{hist}^{(i)} = [L_{n-m+1}, L_{n-m+2}, \dots, L_n]$  that  $u^{(i)}$  visited, where  $n$  is the current time step and  $m$  is the number of considered previous locations, the goal is to predict the location identifier the same user  $u^{(i)}$  will visit in the next time step, i.e.,  $l_{n+1} \in O$ .

In this study, we construct the historical location sequence using locations visited in the previous 7 days. Thus, the number of considered previous locations  $m$  depends on the user  $u^{(i)}$  and the current time step  $n$ , making the next location prediction a sequence prediction problem with variable sequence length.

## 4 METHODS

Figure 1 presents the overview of the proposed model. Specifically, the model considers **historical location, time and travel mode patterns** as input and learns from **the auxiliary task of predicting the next travel mode** to improve the prediction of the next location. The model consists of three major components: (1) feature embedding (Section 4.1); (2) transformer decoder and fully connected (FC) layers (Section 4.2); and (3) loss design and prediction (Section 4.3). The detailed description of each component is presented in the following subsections.

### 4.1 Embedding learning

The representation and modelling of historical information is a vital step for accurately predicting the next location. Most existing prediction models utilize the location visitation sequence to understand the complex spatial-temporal dependencies. **We additionally consider temporal and travel mode information that can provide context for visitation patterns.** Moreover, user information is beneficial for the network to identify location sequences travelled by different users and learn **user-specific movement patterns.** Therefore, we regard the **location  $l_k$ , the start time  $t_k$ , and the travel mode  $e_k$**  at any time step  $k$  as well as the user  $u^{(i)}$  (represented as a **unique identifier**) of the current sequence as the input features (see Figure 2). Specifically, we extract the time of the day  $h_k$  (grouped into 15-minute bins) and the day of the week  $d_k$  from the start time  $t_k$ , aiming to separate different levels of periodicity in location visits.

We now introduce the embedding method to transform each feature **from the categorical type to a finite-dimensional real-valued vector.** This is a more efficient and effective way to represent the correlation between different categorical values compared to the commonly used one-hot encoding method [10, 20]. **The embedding layers are parameter matrices that provide mappings between the original variable and the real-valued vector, whose parameters are jointly optimized with the entire network.** The feature construction process can be formulated as follows:

$$emb_k^l = l_k W_l \quad (1)$$

$$emb_k^h = h_k W_h \quad (2)$$

$$emb_k^d = d_k W_d \quad (3)$$

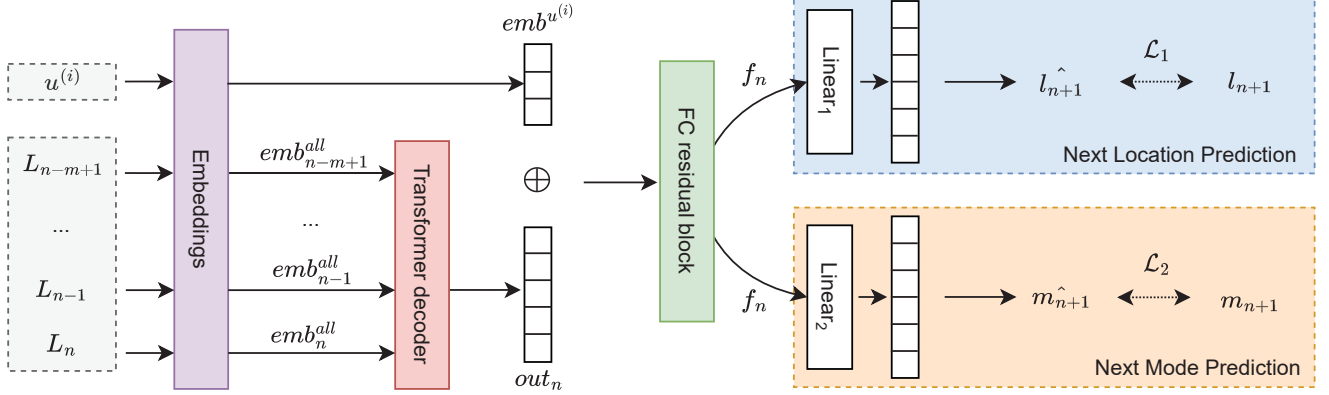
$$emb_k^e = e_k W_e \quad (4)$$

$$emb^{u^{(i)}} = u^{(i)} W_u \quad (5)$$

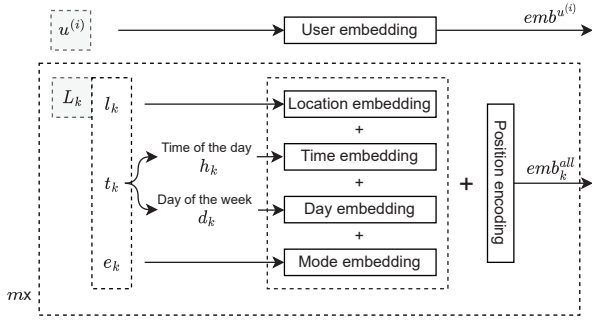
where  $emb_k^l, emb_k^h, emb_k^d, emb_k^e \in \mathbb{R}^{d_{base}}$  and  $emb^{u^{(i)}} \in \mathbb{R}^{d_{user}}$  are the respective embedding vectors, and  $l_k, h_k, d_k, e_k$  and  $u^{(i)}$  are the respective one-hot encoded original categorical features.  $W$  terms stand for weight matrices that are optimized during training. See Figure 2 for an illustration of the embedding process. Finally, the total embedding vector  $emb_k^{all}$  for each time step  $k$  is obtained by adding all sequence features together with a position encoding  $PE$ :

$$emb_k^{all} = emb_k^l + emb_k^h + emb_k^d + emb_k^e + PE \quad (6)$$

We use the original positional encoding proposed by Vaswani et al. [39] that utilizes sine and cosine functions to encode sequence



**Figure 1: Structure of the proposed next location prediction model.** The model learns to represent input features into embedding vectors, which are fed into a transformer decoder neural network. The network’s output is concatenated with user embedding and processed through a fully connected residual block for predicting the next location. We introduce the next mode prediction as an auxiliary task to guide the learning process.



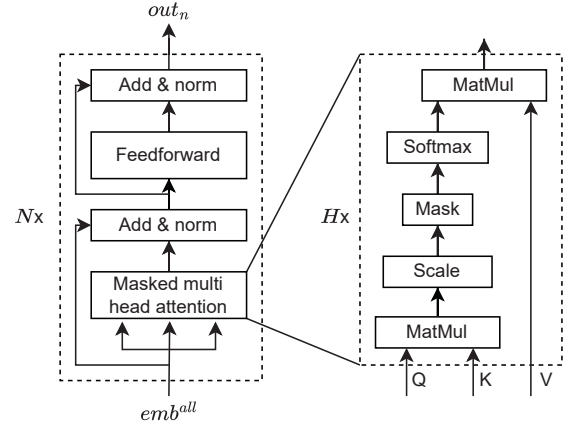
**Figure 2: Pipeline of generating the embedding vectors.**

information in the embedding. Positional encoding is essential for the training as the self-attention module does not implicitly assume sequential order [39]. Note that we use **addition** instead of a concatenation operation to combine embedding vectors [46]. This process **allows the model to flexibly balance the importance of each feature**, as compared to manually assigning and tuning the sizes for different embedding vectors. After the embedding process, we obtain a single user embedding  $emb^{u(i)}$  and a sequence of embedding vectors  $emb_k^{all}$  representing features at each time step  $k$ , which can be further processed with the transformer-based network.

#### 4.2 Transformer and fully connected layers

An accurate next location prediction model should be able to extract regularities and capture the multilevel periodicity from the complex spatio-temporal historical sequences [9]. We utilize a transformer-based network to learn location transition patterns from historical location, temporal and travel mode information encoded in the final embedding vector  $emb_k^{all}$ . We adopt an architecture similar to the Generative Pre-trained Transformer (GPT) model, originally designed for language modelling [29], where only the transformer

decoder part is used. The decoder consists of a stack of  $N$  identical blocks (see Figure 3), each with two layers. The first is the **masked multi-head attention**, and the second is the **fully-connected feedforward network with two linear layers and a ReLU activation function**. Residual connections and layer normalization components are added to each layer to facilitate learning [13]. The input and output dimension of each block is designed to be the same as the embedding vector, i.e.,  $d_{model} = d_{base}$ .



**Figure 3: Structure of the transformer decoder neural network and the masked multi-head attention.**

The core of the transformer structure is the multi-head self-attention mechanism. The attention function can be understood as **obtaining an output value based on a query and a set of key-value pairs, all of which are vectors of size  $d_k$** . As described by Vaswani et al. [39], transformer uses the scaled dot-product attention and implements the calculation efficiently by packing the set of query, key and value values into matrices Q, K and V. This process can be



formulated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

Then, multi-head attention is constructed by concatenating the results of  $H$  attention functions (see Figure 3):

$$\text{MultiHead}(Q, K, V) = (\text{head}_1 \oplus \dots \oplus \text{head}_H)\mathbf{W}^O \quad (8)$$

$$\text{where } \text{head}_i = \text{Attention}(Q\mathbf{W}_i^Q, K\mathbf{W}_i^K, V\mathbf{W}_i^V) \quad (9)$$

where  $\oplus$  represents the concatenation operation and  $\mathbf{W}$  terms are parameter matrices learned by the network. In each masked multi-head attention, key, value and query matrices are identical, all corresponding to the output of the previous block. In the first block, they are set as the embedding matrix  $\text{emb}^{\text{all}}$ , obtained through stacking all embedding vectors according to their sequence. Note that we additionally include the forward-masking operation to prevent the attention function from accessing information from “future” time steps (see Figure 3); that is, the entry at time step  $i$  can only focus on the information preceding (and including)  $i$ . The self-attention mechanism enables the model to access information from every step in the historical sequence and evaluate its importance by learning parameter matrices. This ensures that long-term dependencies in the historical pattern can be extracted. Additionally, thanks to the multi-head design, the model retains multiple sets of parameter matrices that focus on various places in the historical sequence, efficiently capturing the multi-level periodic nature of human mobility.

Then, the output vector of the transformer-decoder model  $\text{out}_n$  is concatenated with the user embedding  $\text{emb}^{u(i)}$ , and fed into a fully-connected residual block:

$$f_n = FC(\text{out}_n \oplus \text{emb}^{u(i)}) \quad (10)$$

where  $FC(\cdot)$  represents the operation by the fully-connected residual block, whose structure is shown in Figure 4. It consists of a single linear feedforward layer and a ReLU activation function, followed by a dropout layer, residual connection and a batch normalization layer. This block learns the dependencies of the transformer decoder output and the user embedding, helping the model extract user-specific movement patterns. Finally, the aggregated vector representation  $f_n$  encapsulates location, temporal, travel mode and user information of the entire historical sequence, and can be further used to predict the next location.

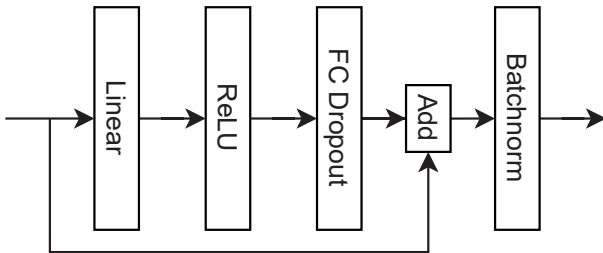


Figure 4: Structure of the fully-connected residual block.

### 4.3 Loss and training

The aggregated vector representation  $f_n$  is then fed into two branches: the next location prediction branch and the next mode prediction branch. For the first task, we calculate the probability of each location through a linear projection and a softmax function:

$$P(\hat{l}_{n+1}) = \text{Softmax}(\text{Linear}_1(f_n)) \quad (11)$$

where  $P(\hat{l}_{n+1})$  of size  $|O|$  contains the probability of all locations to be visited at the next time step. Therefore, the most likely visited location  $\hat{l}_{n+1}$  at time step  $n + 1$  is the location with the highest probability in  $P(\hat{l}_{n+1})$ .

Given the ground truth next location  $l_{n+1}$  from the training dataset, the task can be regarded as a multi-class classification problem. We obtain our location prediction loss  $\mathcal{L}_1$  using the multi-class cross entropy:

$$\mathcal{L}_1 = - \sum_{i=1}^{|O|} P(l_{n+1})^{(i)} \log(P(\hat{l}_{n+1})^{(i)}) \quad (12)$$

where  $P(\hat{l}_{n+1})^{(i)}$  represents the predicted probability of visiting the  $i$ -th location and  $P(l_{n+1})^{(i)}$  is the one-hot represented ground truth, i.e.,  $P(l_{n+1})^{(i)} = 1$  if the actual visited next location is the  $i$ -th location, and  $P(l_{n+1})^{(i)} = 0$  otherwise.

In addition to predicting the next location, we enforce the network to predict the next travel mode and optimize the network's parameters based on the joint prediction loss. This design is motivated by the assumption that learning to correctly predict the next travel mode can improve the prediction of the next location. Analogous to the next location probability, the probability of choosing each travel mode  $P(e_{n+1})$  can be obtained as follows:

$$P(e_{n+1}) = \text{Softmax}(\text{Linear}_2(f_n)) \quad (13)$$

Therefore, we introduce the mode prediction loss  $\mathcal{L}_2$  using the multi-class cross entropy:

$$\mathcal{L}_2 = - \sum_{i=1}^{|\mathcal{M}|} P(e_{n+1})^{(i)} \log(P(\hat{e}_{n+1})^{(i)}) \quad (14)$$

where  $|\mathcal{M}|$  is the total number of travel mode categories,  $P(\hat{e}_{n+1})^{(i)}$  represents the probability of using the  $i$ -th mode at time step  $n + 1$  and  $P(e_{n+1})^{(i)}$  is the one-hot represented true next mode.

The final loss is the combination of the two losses. Given training samples in mini-batches of size  $B$ , the network's parameters are optimized by minimizing the following loss:

$$\mathcal{L} = \sum_{b=1}^B (\mathcal{L}_1(b) + \theta \mathcal{L}_2(b)) \quad (15)$$

where  $\mathcal{L}_1(b)$  and  $\mathcal{L}_2(b)$  correspond to the location prediction loss and the mode prediction loss of the  $b$ -th training sample in the mini-batch, respectively, and  $\theta$  balances the relative weights of the two losses.

## 5 EXPERIMENTS

### 5.1 Datasets

We test the proposed method using longitudinal tracking datasets from two tracking studies performed in Switzerland.

**The Green Class (GC) study** [24]. The study contains 139 participants based in Switzerland who got access to a comprehensive mobility package consisting of a general public transport pass valid in Switzerland, as well as access to several car- and bike-sharing programs and a battery electric vehicle for their personal use. All participants were asked to record their movements using a tracking app on their phones between November 2016 and January 2018. The app pre-processed the movement of participants into staypoints (stationary behaviour) and triplets (continuous movement without changing travel mode). The study participants provided high-level activity labels for staypoints (*home, work, errand, leisure, wait, and unknown*) and mode labels for triplets (*car, e-car, train, bus, tram, bicycle, e-bike, walk, airplane, boat, coach*).

**The yumuv study** [30]. The study involves 498 participants based in and around the city of Zürich, Switzerland. Participants were tracked over three months between July and September 2020, when COVID-19 case numbers, non-medical interventions and impact on people’s daily lives were low in Switzerland. Participants got access to a mobility-as-a-service app called yumuv that provided simplified access to several micro-mobility modes. All participants were tracked via an app on their phones. Similar to the GC study, the app pre-processed movements into staypoints and triplets that participants labelled with activity labels and travel modes.

We pre-process the raw movement data from the GPS tracking studies into corresponding analysis units for the next location prediction task using Python and the open-source *Trackintel* framework [25]. We only consider users tracked for more than 300 days in GC and more than 30 days in Yumuv to ensure high temporal tracking coverage. 93 users in GC and 422 users in Yumuv remain. Based on the activity label, we regard a stay point as an activity if its duration is **longer than 25 minutes** or if it was **labelled with a non-trivial purpose** (any available purpose except for *wait* or *unknown*). Then, the activity stay points are spatially aggregated into locations to account for visits to the same place at different times. We utilized the function provided in *Trackintel* with parameters  $\epsilon = 20$  and  $num\_samples = 2$  to generate *dataset* locations [15]. Locations are also attached with the arrival travel mode. Since travel to a location may involve multiple stages with various travel modes, we determine the main travel mode as the mode with the longest distance [2]. We further group the main travel mode into seven groups: *walk, bicycle, train, tram, bus, car* and *other* (including *ski, airplane* and *coach*). Table 1 shows the number of locations in both datasets, categorized based on the main travel mode.

**Table 1: Main travel mode frequency of locations.**

Mode category	GC	Yumuv
Walk	55,000	80,617
Bicycle	4,926	24,745
Train	25,495	20,260
Tram	2,182	10,943
Bus	3,465	10,962
Car	88,664	55,289
Other	1,751	476
<b>Total</b>	<b>181,483</b>	<b>203,292</b>

We split the location visitation sequence of each user into **train, validation and test sets with the ratio of 6:2:2 based on time**, such that location records that occurred in the first 60% of days are considered as train and the last 20% of days as test. The hyper-parameters are tuned on the validation set, and the test set is only used for reporting the prediction performance. We **train a single population-level model using records from the training set of all users**. We do not pre-filter the locations visited less often, which is a pre-processing step usually found in the location prediction models using GPS traces (e.g., [34]). Although preserving the original location transition patterns, this setting makes the problem a more challenging task, as a large proportion of mobility consists of exploring new locations that are difficult to predict [5].

## 5.2 Setup

The hyper-parameters for the network are reported in Table 2. We chose the best performing set of hyper-parameters using grid search: the number of layers  $L$  from  $\{2, 4, 6\}$ , the number of heads  $H$  from  $\{4, 8\}$ , the size of the embedding  $d_{base}$  from  $\{32, 64, 128, 256\}$  and the dropout of the FC layer from  $\{0.1, 0.2, 0.5\}$ . The  $\theta$  in the final loss (Eq. (15)) is set to 1. Adam optimizer with an initial learning rate of  $1e^{-3}$  and L2 penalty of  $1e^{-6}$  is used to optimize the model’s parameters. We implement learning rate warm-up and decay as reported by Vaswani et al. [39]. To alleviate the model over-fit on the training dataset, we use an early stopping strategy to pause the learning if the validation loss stops decreasing for 3 epochs. We then drop the learning rate by 0.1 and continue the training from the model with the lowest validation loss. This process is repeated 3 times.

**Table 2: Experiment settings and network parameters.**

Setting	Value	Network parameter	Value
Learning rate (lr)	$1e^{-3}$	Emb. $d_{base}$	64
L2 penalty	$1e^{-6}$	User emb. $d_{user}$	16
lr warm up	2 epochs	# layers $N$	4
lr decay	0.02	# heads $H$	8
Early stop patience	3 epochs	Feedforward	256
Early stop lr drop	0.1	FC dropout	0.1

We implement several popular methods as baselines to compare with our proposed model. The approaches we choose are as follows: (1) **Markov models**. The *de facto* standard method for this task [11, 17, 18]. We implement the 1<sup>st</sup> order Mobility Markov Chain (1-MMC) [11], as our pre-experiment shows that higher-order methods can not improve the prediction performance. (2) **LSTM network**, a widely adopted recurrent neural network (RNN) model for sequence data, has been successfully applied to predict the next location [34, 42]. We implement the model proposed by Solomon et al. [34]. (3) **Deepmove** [9]. It incorporates a historical attention module to extract historical movement patterns with RNN-based networks. We consider the most recent two days as the current sequence and the remaining five days as the historical sequence for the input to the model. (4) **MobTcast** [44]. It considers temporal, semantic, social, and geographical contexts with a transformer encoder-based

**Table 3: Performance evaluation results for next location prediction. The mean and the standard deviation (in parentheses) are reported. Numbers marked in bold and underline represent the best and the second-best performing method respectively.**

Dataset	Method	F1	Acc@1	Acc@5	Acc@10	MRR
GC	1-MMC [11]	25.56	35.29	59.10	62.95	46.10
	LSTM [34]	30.26 (0.42)	36.25 (0.18)	60.86 (0.15)	65.90 (0.27)	47.54 (0.13)
	Deepmove [9]	30.82 (0.19)	36.27 (0.13)	60.96 (0.13)	66.06 (0.17)	47.57 (0.10)
	MobTcast [44]	<u>32.29</u> (0.28)	37.20 (0.29)	60.34 (0.37)	65.51 (0.35)	47.76 (0.30)
	LSTM attn	32.16 (0.48)	<u>37.73</u> (0.08)	<u>61.65</u> (0.21)	<u>66.57</u> (0.37)	<u>48.70</u> (0.12)
	Ours	<b>34.89</b> (0.16)	<b>39.93</b> (0.14)	<b>61.89</b> (0.17)	<b>66.92</b> (0.10)	<b>49.94</b> (0.11)
Yumuv	1-MMC [11]	34.81	45.00	68.61	71.58	55.72
	LSTM [34]	41.46 (0.34)	45.32 (0.26)	68.72 (0.23)	72.25 (0.29)	56.10 (0.22)
	Deepmove [9]	42.29 (0.21)	46.11 (0.26)	69.47 (0.08)	73.01 (0.06)	56.91 (0.19)
	MobTcast [44]	42.65 (0.28)	46.13 (0.25)	69.18 (0.17)	72.81 (0.07)	56.69 (0.16)
	LSTM attn	<u>43.02</u> (0.28)	<u>46.66</u> (0.23)	<u>69.87</u> (0.11)	<u>73.48</u> (0.16)	<u>57.38</u> (0.15)
	Ours	<b>45.43</b> (0.32)	<b>48.67</b> (0.20)	<b>70.17</b> (0.09)	<b>73.50</b> (0.08)	<b>58.49</b> (0.13)

structure to forecast the next POI. Also, the model includes two losses to encourage predicted and ground truth locations to be close in space. We incorporate all components except for the social context since the overlap between location visits for our users is low due to the high-resolution GPS tracking. (5) **LSTM with self-attention (LSTM attn)**. Inspired by Li et al. [20], we implement an LSTM-based network with self-attentions between the current and previous hidden states.

We use the following metrics to evaluate the next location prediction performance of various methods: (1) **Accuracy**. It indicates the number of correctly predicted locations by the network. We rank the predicted probability from  $P(l_{n+1})$  in descending order and count the number of times that the ground truth location appears in the top-k predicted locations (Acc@k). Acc@1, Acc@5, and Acc@10 are reported. (2) **F1 score**. Since certain locations are visited more often than others [35], we employ the F1 score weighted by location visitation frequency. We argue that the F1 score is a better metric than Acc@1 as it considers the unbalanced location visits. (3) **Mean Reciprocal Rank (MRR)**. The metric is commonly applied to measure performance in information retrieval and re-identification tasks. MRR is the harmonic mean of the ground truth label’s rank in the prediction.

## 5.3 Results

**5.3.1 Overall performance.** Table 3 shows the performance of next place prediction methods on the two considered datasets. We train each DL model 5 times with different random seeds and report the mean and the standard deviation of the respective performance indicators.

The 1-MMC method provides a strong baseline for the task. However, its performance is worse than any DL model, indicating that the Markov property cannot fully capture mobility patterns. On the contrary, the naïve LSTM model outperforms the 1-MMC method by a relatively large margin, demonstrating the effectiveness of DL-based models in predicting mobility. The worse performance compared to other DL methods can be attributed to the insufficient consideration of long-term dependencies in location visit sequences,

an inherent shortcoming of the LSTM model [39]. This deficiency is partly tackled by introducing the historical attention module in the Deepmove model that explicitly focuses on extracting historical information, which consistently outperforms the naïve LSTM on all indicators for both datasets. MobTcast obtains comparable performance in Yumuv and slightly better F1 and Acc@1 scores in GC compared to Deepmove, suggesting the additional consideration of context and the introduction of two losses have limited effect on predicting an individual’s next location.

However, we observe a significant performance increase by adding a self-attention mechanism to the naïve LSTM model. The self-attention’s ability to focus on specific steps during prediction indicates that successfully mining historical patterns can greatly benefit location prediction. By incorporating the multi-head self-attention and introducing the next mode prediction loss, the proposed model outperforms all other models by a relatively large margin. The relative increase of the F1 score is 8.05% and 5.60% for two datasets, respectively, which is calculated using the formula  $(F_{11} - F_{12})/F_{12}$ , where  $F_{11}$  is the F1 score of our model and  $F_{12}$  represents the F1 score of the second-best performing model. The consistent pattern for both considered datasets demonstrates the effectiveness of our network architecture and loss design. Additionally, we report that the performance difference between LSTM with self-attention and our transformer decoder-based model is relatively small in Acc@5 and Acc@10 compared to the other indicators. We assume that the single head self-attention can already extract multiple likely visited next locations but lacks the ability to distinguish the importance within this set successfully.

**5.3.2 Ablation study.** We perform an extensive ablation study to understand the importance of each component in our proposed model. In Table 4, we show the performance of different variants of our model, considering whether or not to include temporal features  $h_k$  and  $d_k$ , the travel mode feature  $e_k$ , and the next travel mode prediction loss  $\mathcal{L}_2$ . The performance results are consistent across both datasets. We first report that the model performs significantly better with temporal features. This increase is intuitive as the time encodes different levels of periodicity in location visits compared

**Table 4: Performance of the ablation study for next location prediction. We consider the model with (✓) and without (-) adding temporal features (T), the travel mode feature (F), and the mode prediction loss ( $\mathcal{L}_2$ ).**

Dataset	Module			F1	Acc@1	Acc@5	Acc@10	MRR
	T	F	$\mathcal{L}_2$					
GC	-	-	-	31.24 (0.18)	36.22 (0.22)	60.57 (0.25)	65.75 (0.23)	47.42 (0.24)
	-	-	✓	31.20 (0.18)	36.64 (0.19)	60.99 (0.10)	66.09 (0.07)	47.83 (0.12)
	-	✓	-	31.61 (0.23)	36.80 (0.18)	60.88 (0.07)	65.98 (0.10)	47.86 (0.09)
	✓	-	-	34.55 (0.19)	39.10 (0.20)	61.10 (0.24)	66.18 (0.19)	49.07 (0.17)
	✓	-	✓	<u>34.80</u> (0.29)	<u>39.81</u> (0.12)	<u>61.75</u> (0.11)	<u>66.67</u> (0.10)	<u>49.76</u> (0.07)
	✓	✓	-	34.60 (0.09)	39.39 (0.10)	61.23 (0.07)	66.28 (0.16)	49.34 (0.08)
	✓	✓	✓	<b>34.89</b> (0.16)	<b>39.93</b> (0.14)	<b>61.89</b> (0.17)	<b>66.92</b> (0.10)	<b>49.94</b> (0.11)
Yumuv	-	-	-	42.11 (0.17)	45.90 (0.12)	69.47 (0.12)	73.12 (0.12)	56.73 (0.09)
	-	-	✓	42.83 (0.24)	46.45 (0.19)	69.47 (0.17)	72.90 (0.11)	57.01 (0.13)
	-	✓	-	42.68 (0.18)	46.30 (0.18)	69.68 (0.14)	73.15 (0.18)	57.01 (0.16)
	✓	-	-	44.53 (0.10)	47.78 (0.05)	69.68 (0.27)	73.26 (0.19)	57.82 (0.10)
	✓	-	✓	<u>45.04</u> (0.22)	<u>48.37</u> (0.19)	<u>70.08</u> (0.12)	<u>73.41</u> (0.14)	<u>58.28</u> (0.12)
	✓	✓	-	44.65 (0.11)	47.97 (0.11)	69.86 (0.12)	73.38 (0.12)	58.01 (0.10)
	✓	✓	✓	<b>45.43</b> (0.32)	<b>48.67</b> (0.20)	<b>70.17</b> (0.09)	<b>73.50</b> (0.08)	<b>58.49</b> (0.13)

to merely having the sequence information. While considering the travel mode feature and the mode prediction loss are both beneficial for the model’s performance, the latter is more important for the task. Moreover, we observe a further performance gain when simultaneously including the travel mode feature and the mode prediction loss, as indicated by the highest performance achieved by the complete model for both datasets. These experiments show that temporal features are indispensable for accurate next location prediction. They also confirm that both historical travel mode patterns and the ability to predict the next travel mode are essential for the task, justifying the effectiveness of our model design.

Since our proposed model outputs both the probability of the next location and the next travel mode, we can utilize the same model structure to evaluate the performance of predicting the next travel mode. From this viewpoint, time and location can be regarded as additional features, and predicting the next location can be seen as the auxiliary task. We evaluate the usefulness of these components in the next mode prediction task using an ablation study. The performance result are shown in Table 5. We again observe a large performance increase after including temporal features. Furthermore, historical location visits and the additional location prediction loss benefit the prediction. The top performance is achieved by the complete model that utilizes all three components in the network. These results indicate that the next travel mode prediction performance can be improved by considering location visit information. Combined with the location prediction ablation study, we conclude that the next location and the next travel mode prediction tasks are inherently dependent and should be tackled together to achieve optimal performance.

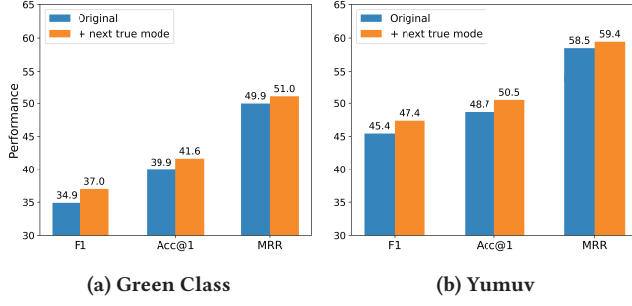
**5.3.3 Importance of travel mode information.** We further investigate the influence of travel mode on the next location prediction task. To understand to what extent the next location prediction result depends on the prediction of travel mode, we input the ground truth next travel mode into the model. This is

**Table 5: Performance for next mode prediction. We consider the model with and without adding temporal features (T), the location feature (F), and the location prediction loss ( $\mathcal{L}_1$ ).**

Dataset	Module			F1	Acc@1	MRR
	T	F	$\mathcal{L}_1$			
GC	-	-	-	54.29 (0.16)	58.17 (0.09)	75.95 (0.05)
	✓	-	-	57.30 (0.18)	60.14 (0.09)	77.28 (0.04)
	✓	✓	-	<u>58.64</u> (0.33)	<u>61.03</u> (0.29)	<u>77.67</u> (0.17)
	✓	✓	✓	<b>58.75</b> (0.08)	<b>61.16</b> (0.25)	<b>77.72</b> (0.16)
Yumuv	-	-	-	50.84 (0.12)	52.53 (0.13)	71.90 (0.07)
	✓	-	-	52.59 (0.11)	53.96 (0.13)	72.74 (0.09)
	✓	✓	-	<u>53.70</u> (0.21)	<u>54.64</u> (0.10)	<b>73.05</b> (0.06)
	✓	✓	✓	<b>54.03</b> (0.19)	<b>54.84</b> (0.11)	<b>73.05</b> (0.08)

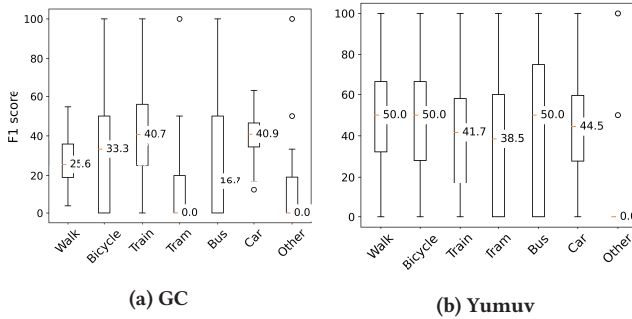
achieved by feeding the travel mode at the next time step into the mode embedding layer to generate the embedding vector  $emb_{n+1}^e$ . Then, the vector is added to the user embedding and further processed with the fully connected layer, i.e., Eq. (10) becomes  $f_n = FC((out_n + \text{Linear}(emb_{n+1}^e)) \oplus emb^{u(i)})$ . We can regard this new problem as predicting the next location conditioned on the next travel mode, where the scenario is to predict the location when knowing which travel mode the user will take. Figure 5 compares the performance of the original model with the model that includes the next travel mode. When knowing the next travel mode, we observe a notable increase in all indicators for the two datasets. In line with the conclusion from the ablation study, this result illustrates that the choice of the next location depends strongly on the travel mode. The performances shown in Figure 5 can also be interpreted as the upper bound of performance gain that can be achieved by including the prediction of travel mode in our model. This result suggests that working on a better mode prediction still holds great potential to increase the prediction accuracy of the next location.





**Figure 5: Performance of the original model (blue) and the model that considers the ground truth next travel mode (orange).**

This mode-location dependency is evident by checking differences in predicting various categories of locations. In particular, we group the location prediction result based on the ground truth next travel mode and plot the F1 score in Figure 6. Each “box” in the figure is generated by the F1 score of every user and represents the model’s performance in a single travel mode. We also labelled each box with the median F1 score across users. The differences in the F1 scores suggest that location prediction depends on the travel mode: locations with specific travel modes are more challenging to predict than others. For example, the locations reached by *tram* and *bus* obtained relatively lower performance than other modes in GC (Figure 6a). For Yumuv, the mode *tram* achieves a relatively low median F1-Score, and the mode *bus* has a high variance across users (Figure 6b). This observation could be attributed to the more variate choice of locations when users travel with these modes. The failure to predict the mode *other* (including *airplane*, *ski* and *coach*) is related to the exploration nature of human mobility: users often travel with these modes to new locations that are difficult for the model to predict. Moreover, the distinct pattern for the GC and Yumuv datasets can be explained by the mobility patterns of the two groups of participants. GC participants are more active in their mobility and often travel a longer distance by car and train [14], whereas Yumuv users mostly commute within the city [30] and have a more balanced use of travel modes (Table 1).



**Figure 6: F1 score boxplot of individual categorized by the next travel mode.**

## 6 DISCUSSION AND CONCLUSION

Travel behaviour researchers have long realized that individuals’ choice of locations is highly influenced by other travel behaviour dimensions, such as the day of the week and the availability of travel modes. While this dependency is evident in an increasing number of empirical studies, limited attention has been placed on integrating this information into the next location prediction task. In this study, we present a transformer-based model to predict the next location visit of an individual. In particular, the model is designed to accurately predict the next location and the next travel mode at the same time. We test our model on two large-scale real-world GPS tracking datasets and compare its performance with most recent methods for the task. Our experiments show that the proposed model learns the dynamics of human mobility from historical time, travel mode and location sequences and obtained state-of-the-art next location prediction performance. Our extensive ablation study demonstrates that temporal and travel mode features and the ability to predict the next travel mode are essential components in an accurate location prediction model. Moreover, our analysis indicates strong mode-location dependency, which suggests the next location and the next travel mode prediction tasks should be tackled together to achieve optimal performance.

We conclude that considering historical travel mode patterns increases the accuracy of the next location prediction. This result connects travel mode detection models with next location prediction models. Mode detection models aim to classify a movement’s travel mode given its spatio-temporal characteristics [31], which can enrich general tracking datasets that lack accurate user-provided travel mode labels. Therefore, mobility prediction can be tackled using a two-step approach: (1) identify the travel mode with detection models based on spatio-temporal movement characteristics; and (2) predict the next location with prediction models based on temporal, travel mode and location information. This approach indicates that an optimal system for mobility prediction should develop accurate models for both tasks.

Next location prediction is an essential backbone of many sustainable transport solutions; however, it is a challenging problem that is not yet fully tackled. Although we improve the state-of-the-art considerably compared to the other methods, the field still needs more research to achieve a more accurate prediction performance. We propose several future directions based on the results of this work. This study demonstrates the importance of travel mode for predicting an individual’s immediate next location visit. Since the mode-location dependency is more evident on a longer time scale [14, 38], future research should consider integrating travel mode information into mobility simulation models for generating more realistic mobility sequences. Furthermore, besides the inter-dependent travel behaviour dimensions, studies have shown that external contexts, such as the built environment, play an indispensable role in shaping individuals’ mobility [8]. This knowledge can be effectively extracted and combined with individual mobility in DL models based on land-use maps and POI data. Whether or not the additional information will improve the performance of next location prediction or mobility generation tasks is a direction worth exploring.

## ACKNOWLEDGMENTS

This work is supported by the Hasler Foundation (grant number 1-008062). The yumv dataset was collected as part of a joint study with Swiss Federal Railways that was financed through the ETH Mobility Initiative (grant number MI-01-19).

## REFERENCES

- [1] Daniel Ashbrook and Thad Starner. 2002. Learning Significant Locations and Predicting User Movement with GPS. In *International Symposium on Wearable Computers (ISWC) 2002*. 101–108.
- [2] Kay W. Axhausen. 2007. Definition of movement and activity for transport modelling: Contribution to the handbooks in transport: transport modelling. In *Handbook of transport modelling: transport modelling*. Vol. 1. Elsevier, 329–343.
- [3] Meng Chen, Yang Liu, and Xiaohui Yu. 2014. NLPMM: A Next Location Predictor with Markov Modeling. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2014*. 186–197.
- [4] Elisabetta Cherchi, Cinzia Cirillo, and Juan de Dios Ortúzar. 2017. Modelling correlation patterns in mode choice models estimated on multiday travel data. *Transportation Research Part A: Policy and Practice* 96 (2017), 146–153.
- [5] Andrea Cuttone, Sune Lehmann, and Marta C. González. 2018. Understanding predictability and exploration in human mobility. *EPJ Data Science* 7 (2018), 2.
- [6] Dimas B.E. Dharmowijoyo, Yusak O. Susilo, and Anders Karlström. 2017. Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: A case study in the Bandung Metropolitan Area, Indonesia. *Journal of Transport Geography* 64 (2017), 1–12.
- [7] Dimas B.E. Dharmowijoyo, Yusak O. Susilo, and Anders Karlström. 2016. Day-to-day variability in travellers' activity-travel patterns in the Jakarta metropolitan area. *Transportation* 43, 4 (2016), 601–621.
- [8] Temitope Farinloye, Emmanuel Mogaji, Stella Aririguzoh, and Tai Anh Kieu. 2019. Qualitatively exploring the effect of change in the residential environment on travel behaviour. *Travel Behaviour and Society* 17 (2019), 26–35.
- [9] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. DeepMove: Predicting Human Mobility with Attentional Recurrent Networks. In *International World Wide Web Conference (WWW) 2018*. 1459–1468.
- [10] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to Simulate Human Mobility. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2020*. 3426–3433.
- [11] Sébastien Gams, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2012. Next Place Prediction Using Mobility Markov Chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility (MPM '12)*. 3:1–3:6.
- [12] David Griggs, Mark Stafford-Smith, Owen Gaffney, Johan Rockström, Marcus C. Ohman, Priya Shyamsundar, Will Steffen, Gisbert Glaser, Norichika Kanie, and Ian Noble. 2013. Sustainable development goals for people and planet. *Nature* 495, 7441 (2013), 305–307.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR) 2016*. 770–778.
- [14] Ye Hong, Henry Martin, Yanan Xin, Dominik Bucher, Daniel J. Reck, Kay W. Axhausen, and Martin Raubal. 2022. Conserved quantities in human mobility. From locations to trips. (2022). <https://doi.org/10.3929/ethz-b-000551831>
- [15] Ye Hong, Yanan Xin, Henry Martin, Dominik Bucher, and Martin Raubal. 2021. A Clustering-Based Framework for Individual Travel Behaviour Change Detection. In *International Conference on Geographic Information Science (GIScience) 2021 - Part II*. 4:1–4:15.
- [16] Haosheng Huang, Dominik Bucher, Julian Kissling, Robert Weibel, and Martin Raubal. 2019. Multimodal Route Planning With Public Transport and Carpooling. *IEEE Transactions on Intelligent Transportation Systems* 20 (2019), 3513–3525.
- [17] Quanying Huang. 2017. Mining online footprints to predict user's next location. *International Journal of Geographical Information Science* 31, 3 (2017), 523–541.
- [18] Vaibhav Kulkarni and Benoît Garbinato. 2019. 20 Years of Mobility Modeling & Prediction: Trends, Shortcomings & Perspectives. In *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL) 2021*. 492–495.
- [19] Gabriel Leite Mariante, Tai-Yu Ma, and Véronique Van Acker. 2018. Modeling discretionary activity location choice using detour factors and sampling of alternatives for mixed logit models. *Journal of Transport Geography* 72 (2018), 151–165.
- [20] Fa Li, Zhipeng Gui, Zhaoxu Zhang, Dehua Peng, Siyu Tian, Kunxiaojia Yuan, Yunzeng Sun, Huayi Wu, Jianya Gong, and Yichen Lei. 2020. A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction. *Neurocomputing* 403 (2020), 153–166.
- [21] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. In *AAAI Conference on Artificial Intelligence (AAAI) 2016*. 194–200.
- [22] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. 2021. A Survey on Deep Learning for Human Mobility. *Comput. Surveys* 55, 1 (2021), 7:1–7:44.
- [23] Zhenliang Ma and Pengfei Zhang. 2022. Individual mobility prediction review: Data, problem, method and application. *Multimodal Transportation* 1, 1 (2022), 100002.
- [24] Henry Martin, Henrik Becker, Dominik Bucher, David Jonietz, Martin Raubal, and Kay W. Axhausen. 2019. Begleitstudie SBB Green Class - Abschlussbericht. *Arbeitsberichte Verkehrs- und Raumplanung* 1439 (2019). <https://doi.org/10.3929/ethz-b-000353337>
- [25] Henry Martin, Ye Hong, Nina Wiedemann, Dominik Bucher, and Martin Raubal. 2022. Trackintel: An open-source Python library for human mobility analysis. arXiv:2206.03593
- [26] Henry Martin, Daniel J. Reck, and Martin Raubal. 2021. Using Information and Communication Technologies to facilitate mobility behaviour change and enable Mobility as a Service. *GI Forum Journal for Geographic Information Science* 9, 1 (2021), 187–193.
- [27] T. Neutens, F. Witlox, N. Van De Weghe, and P. H. De Maeyer. 2007. Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science* 21 (2007), 1061–1076.
- [28] Poria Pirozmand, Guowei Wu, Behrouz Jedari, and Feng Xia. 2014. Human mobility in opportunistic networks: Characteristics, models and prediction methods. *Journal of Network and Computer Applications* 42 (2014), 45–58.
- [29] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.
- [30] Daniel J. Reck, Henry Martin, and Kay W. Axhausen. 2022. Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D: Transport and Environment* 102 (2022), 103134.
- [31] Avipsa Roy, Daniel Fuller, Trisalyn Nelson, and Peter Kedron. 2022. Assessing the role of geographic context in transportation mode detection from GPS data. *Journal of Transport Geography* 100 (2022), 103330.
- [32] Stefan Schönfelder and K. W. Axhausen. 2016. *Urban rhythms and travel behaviour: spatial and temporal phenomena of daily travel*. Routledge, Farnham, England.
- [33] Darren M. Scott and Sylvia Y. He. 2012. Modeling constrained destination choice for shopping: a GIS-based, time-geographic approach. *Journal of Transport Geography* 23 (2012), 60–71.
- [34] Adir Solomon, Amit Livne, Gilad Katz, Bracha Shapira, and Lior Rokach. 2021. Analyzing movement predictability using human attributes and behavioral patterns. *Computers, Environment and Urban Systems* 87 (2021), 101596.
- [35] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. 2010. Modelling the scaling properties of human mobility. *Nature Physics* 6, 10 (2010), 818–823.
- [36] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.
- [37] Lijun Sun, Kay W. Axhausen, Der-Horng Lee, and Xianfeng Huang. 2013. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences* 110, 34 (2013), 13774–13779.
- [38] Yusak O. Susilo and Kay W. Axhausen. 2014. Repetitions in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman Index. *Transportation* 41, 5 (2014), 995–1011.
- [39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Conference on Neural Information Processing Systems (NeurIPS) 2017*. 5998–6008.
- [40] Dongjie Wang, Kunpeng Liu, Hui Xiong, and Yanjie Fu. 2022. Online POI Recommendation: Learning Dynamic Geo-Human Interactions in Streams. arXiv:2201.10983
- [41] Dongjie Wang, Pengyang Wang, Kunpeng Liu, Yuanchun Zhou, Charles E. Hughes, and Yanjie Fu. 2021. Reinforced Imitative Graph Representation Learning for Mobile User Profiling: An Adversarial Training Perspective. In *AAAI Conference on Artificial Intelligence (AAAI) 2021*. 4410–4417.
- [42] Yang Xu, Dan Zou, Sangwon Park, Qiuping Li, Suhong Zhou, and Xinyu Li. 2022. Understanding the movement predictability of international travelers using a nationwide mobile phone dataset collected in South Korea. *Computers, Environment and Urban Systems* 92 (2022), 101753.
- [43] Yanyan Xu, Serdar Çolak, Emre C. Kara, Scott J. Moura, and Marta C. González. 2018. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nature Energy* 3, 6 (2018), 484–493.
- [44] Hao Xue, Flora D. Salim, Yongli Ren, and Nuria Oliver. 2021. MobTCast: Leveraging Auxiliary Trajectory Forecasting for Human Mobility Prediction. In *Conference on Neural Information Processing Systems (NeurIPS) 2021*. 30380–30391.
- [45] Seo Youn Yoon, Kathleen Deutsch, Yali Chen, and Konstantinos G. Goulias. 2012. Feasibility of using time-space prism to represent available opportunities and choice sets for destination choice models in the context of dynamic urban environments. *Transportation* 39, 4 (2012), 807–823.
- [46] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *AAAI Conference on Artificial Intelligence (AAAI) 2021*. 11106–11115.