Journal of Urban Mobility 2 (2022) 100022

Contents lists available at ScienceDirect

Journal of Urban Mobility

journal homepage: www.elsevier.com/locate/urbmob

Self-Attention based encoder-Decoder for multistep human density prediction☆

John Violos[a,1,*], Theodoros Theodoropoulos[b,2], Angelos-Christos Maroudis[b,3], Aris Leivadeas[a,4], Konstantinos Tserpes[b,5]

[a] Department of Software and IT Engineering, École de technologie supérieure, 1100 Notre-Dame St W, Montreal, H3C 1K3, Quebec, Canada
[b] Department of Informatics and Telematics, Harokopio University of Athens, 9 Omirou, 177 78, Tavros, 16671, Greece

ARTICLE INFO

Keywords:
Mobility
Encoder-decoder
Self attention
Time series
Deep learning
Points of interest
Regions of interest

ABSTRACT

Multistep Human Density Prediction (MHDP) is an emerging challenge in urban mobility with lots of applications in several domains such as Smart Cities, Edge Computing and Epidemiology Modeling. The basic goal is to estimate the density of people gathered in a set of urban Regions of Interests (ROIs) or Points of Interests (POIs) in a forecast horizon of different granularities. Accordingly, this paper aims to contribute and go beyond the existing literature on human density prediction by proposing an innovative time series Deep Learning (DL) model and a geospatial feature preprocessing technique. Specifically, our research aim is to develop a highly-accurate MHDP model leveraging jointly the temporal and spatial components of mobility data. In the beginning, we compare 29 baseline and state-of-the-art methods grouped into six categories and we find that the statistical time series and Deep Learning Encoders-Decoders (ED) that we propose are highly accurate outperforming the other models based on a real and a synthetic mobility dataset. Our model achieves an average of 28.88 Mean Absolute Error (MAE) and 87.58 Root Mean Squared Error (RMSE) with 200,000 pedestrians per day distributed in multiple regions of interest in a 30 minutes time-window at different granularities. In addition, the geospatial feature transformation increases 4% further the RMSE of the proposed model compared to the state of the art solutions. Hence, this work provides an efficient and at the same time general applicable MHDP model that can benefit the planning and decision-making of many major urban mobility applications.

1. Introduction

The modeling and prediction of human mobility is a topic of increasing interest due to its applications in multiple domains of urban mobility, such as personalised recommender systems (Zheng et al., 2018), urban planning (Du et al., 2020) and the design of smart cities (Chen et al., 2019), just to mention a few. In particular, human mobility refers to the movement of human beings (individuals as well as groups) in urban areas in time periods that span from a few minutes to a few hours (Barbosa et al., 2018). It is evident that modeling the human mobility, urban and transport planners can identify movement behavior patterns and suggest corrective actions to improve the livable urban spaces.

This generates an important opportunity for urban mobility and planning stakeholders by leveraging smart mobility data and analytics to not only analyze how existing infrastructures facilitate the life of community members but also how to create a sustainable environment based on the forecasted human mobility. Obviously, the core of the smart mobility data analytic tools is the data related to the movement of the users in an urban environment. Fortunately, nowadays humans, transport infrastructures, and even entire cities are equipped with sensors included in mobile devices, GPS tracking tools and social media geotagging systems that generate continuously mobility data that reflect the every-day activities of citizens.

Following, these data should be carefully analyzed to extract the appropriate knowledge that will make relevant applications more efficient and more intelligent at the same time. To this end, the key theme of this

☆ This document is the results of the research project funded by CHIST-ERA-2018-DRUID-NET project "Edge Computing Resource Allocation for Dynamic Networks.
* Corresponding author.
E-mail addresses: ioannis.violos.1@ens.etsmtl.ca (J. Violos), ttheod@hua.gr (T. Theodoropoulos), it21863@hua.gr (A.-C. Maroudis), aris.leivadeas@etsmtl.ca (A. Leivadeas), tserpes@hua.gr (K. Tserpes).
[1] Conceptualisation, Methodology, Writing - Original Draft.
[2] Encoder Decoder Modeling, Self-attention Mechanism, Geospatial Transformation.
[3] Time Series Analysis, Machine Learning Prediction, Mobility Simulation.
[4] Edge Computing Contextualize.
[5] Smart City Contextualize.

https://doi.org/10.1016/j.urbmob.2022.100022
Received 17 October 2021; Received in revised form 7 April 2022; Accepted 20 May 2022

paper is to design a data driven and machine learning model that processes the mobility data in order to provide timely and accurate insight for an optimal decision making and what-if (Arman et al., 2019) analysis in the context of urban mobility. Specifically, the problem we address is the Multistep Human Density Prediction (MHDP). This problem is defined as the real time prediction of the distribution of moving entities into multiple Regions of Interest (ROIs) or Points of Interest (POI) through different temporal granularities. In this context, moving entities are individuals or human groups moving in an urban area, whereas ROIs are locations which the moving entities frequently visit. The involved prediction can be in a next-step or a multi-step granularity, according to how far in the future the prediction should be made.

Most of the current models are designed to provide single next-step predictions. However, there are many contemporary applications that require predictions with different time granularities. For instance, three major application categories that require multi-step ahead forecasting of the amount of people gathered in multiple ROIs are the following: (a.) The epidemic spreading modeling (Balcan et al., 2010) which defines the crowded ROIs and the time duration they will remain crowded. (b.) The wireless networks (Kapoor et al., 2017), especially in the context of smart cities, where multiple users and sensor devices try to connect in an access point creating network planning bottlenecks. (c.) Edge Computing task offloading mechanisms (Saeik et al., 2021) in which user devices, e.g. Augmented Reality (AR) glasses in touristic attractions, offload their computational intensive workloads in nearby processing nodes at the edge of the network. By leveraging the sequential density of users in different ROIs this can lead to a better planning of these applications both in a short and in a long-term time window.

The above applications are only few of a vast range of applications that an MHDP can be used. Nonetheless, what is import to understand is that to reap the benefits of the MHDP we need first to understand the people's behavior. The mobility and the density of people/application-users into ROIs is characterised by the properties of periodicity and self-similarity making a model that analyzes and forecasts time series a rational approach, since data are usually collected within equal time intervals. These data must include spatial and temporal information of the mobility. Additionally, based on different use cases the models can be enriched with exogenous information like the weather, the terrain characteristics, and various events that affect the mobility decisions of people. Nonetheless, in the particular research we focus only in the spatio-temporal data represented by timestamps and geo-locations feature vectors, which are structured in a time-series dataset (e.g. a person's position every 30 minutes).

For many years, the main approach for time series problems, such as the problem at hand, was the statistical modeling and forecasting. Later, Machine Learning (ML) methods have been proposed as alternatives to the statistical ones. Although, ML models are based on more complicated and advanced mathematical models, their accuracy is criticized to be often below than their statistical counterparts (Makridakis et al., 2018). For this reason, lately, the hype in forecasting prediction is around other types of Artificial Intelligence (AI) techniques such as the Deep Learning (DL) and specifically the gated variations of Recurrent Neural Networks (RNN).

The above techniques are not always performing the same for different types of applications (De Saa and Ranathunga, 2020; Yamak et al., 2019), signifying that there is a strong connection between the use-case and the most appropriate prediction model to be used. Accordingly, in this paper, one of our research aims is to find the best prediction model for any urban mobility use-case that can benefit from the MHDP mechanism. More precisely, our research is trying to answer which model should be incorporated into a MHDP mechanism in order to provide a high prediction accuracy for multiple ROIs in a sequence of time-steps. Hence, this motivated us to make extended experiments among statistical, machine learning and deep learning models in a real dataset and a synthetic human mobility simulator in order to find which has the best performance for the multi-step forecasting of the human density. Our research reveals, that a technique of deep learning, the attention-based Encoder-Decoder (ED) architecture, provides the best accuracy for the mobility prediction.

Towards our path to design the most accurate predictive model for the urban density prediction we have identified the four following major research contributions:

- We discuss how a data-driven methodology applies for the multi-step forecasting of the number of people gathered in ROIs.
- We make an extended experimental comparison in statistical, machine learning and deep learning time series approaches in order to find the best performance model.
- We propose the self-attention based Encoder-Decoders architecture which surpasses the accuracy of the other methods in the literature.
- We propose a geospatial feature transformation that scales the density of people in a ROI based on the density of its neighborhood ROIs weighted by the lengths of their borders. This data transformation improves further the performance of the multi-step predictions.

The rest of the paper is structured as follows: Section 2 provides a short overview of the MHDP in three popular use cases. Section 3 highlights the related work in multi-step forecasting of human density. Section 4 explains how the mobility prediction can be modeled with self-attention based ED architecture and the geospatial feature transformation. Section 5 describes the experimental setup and the evaluation results of our proposed methods. Finally, Section 6 concludes the paper and suggests future directions.

2. Applications of multistep human density prediction

Multistep human density prediction has applications in multiple domains such as smart cities, edge computing, wireless networks and epidemiology modeling. The density prediction in a future time-window give us a comprehensive understanding of the moving entities in a spatio-temporal framework. Doing so, the dynamicity of the moving entities can be reflected in the component of time and the component of space taking into consideration their correlation. The aggregation of the predictions in next location and also the crowd flow prediction miss the depth in the component of time while, the trajectory prediction of single entities misses their relations on the component of space.

2.1. Smart cities

The domain of smart cities includes multiple fields of the information and communication technology in order to improve the quality of human life, wellbeing, economic development and optimise the utility of different resources and services. Mobility data in smart cities are acquired by mobile devices, vehicles equipped with global positioning system devices, smart cards (bank cards and transport cards), and embedded sensors. These data monitored by intelligent systems for traffic management, traffic lighting control, tourism recommendation systems, civil protection experts, intelligent transportation and many more services. Decisions can be real-time based on the dynamic changes of the human density. Every service has its own requirements in the time-steps granularity. For instance, traffic lighting systems need time steps of one minute while intelligent transportation systems of ten minutes or more. Even in these examples the time requirements change based on the scale of geospatial regions and the context of the use cases. Specific use cases in which the multistep density prediction can be leveraged by urban and transport planning are the transportation demand analysis over time (Verma et al., 2021), the sustainable mobility planning (Singh et al., 2022) and the bike-sharing services (Arias-Molinares et al., 2021)

2.2. Edge computing

Edge Computing is another emerging technology where mobility can play a vital role. Edge computing refers to adding the necessary computational and communication resources closer to the end-user at the edge

of the network. This approach can allow fast data processing and real-time decision for mission critical applications. Accordingly, new and future internet technologies, such as Internet of Things (IoT) and 5G and beyond are and will be largely based on the Edge Computing concept. An inherent part of these technologies is mobility. Hence, the adequate prediction of user mobility can facilitate the resource allocation at the edge, and the proactive planning of a relatively limited infrastructure. Additionally, Edge computing is quite dynamic and distributed in its nature due to the type of the applications it supports (Dechouniotis et al., 2020). Specifically, when the application supports mobility, the computational resources allocated to a user should follow its direction by traversing the edge infrastructure close to the ROIs. At the same time, the requirements of ultra reliable and low latency communications that these applications impose, make a necessity the high accurate prediction of mobility models with different granularity of time-windows.

2.3. Epidemiology modeling

Epidemic spreads depend on the human density as there is a significant positive correlation of the likelihood of infection with the close humans interactions. The temporal and spatial dynamics of disease transmission within a population can be modeled by a sequence of multiple time-steps of human density in ROIs. This provides a useful tool to decision-makers who use non-pharmaceutical interventions policies (Ilin et al., 2021), such as quarantines, perimeter closures and social distancing. Specifically, infectious disease epidemiologists use models based on population size, population density, and travel distance. The gravity and radiation models are the most commonly used (Sallah et al., 2017). The gravity model is based on the assumption that the mobility between two locations has a positive correlation with the population size and a negative with the distance, whereas the radiation model assumes that the mobility depends on the population density.

3. Related work

In this paper we address the human density prediction problem in multiple RoIs and multi-steps with a time-series dataset. Accordingly, in this Section we first review and categorize pertinent prediction mechanisms proposed in the literature that could be used and adapted for the human density prediction. Secondly, we present relevant works that have used mobility prediction mechanisms. Finally, we investigate the research gaps of each of the prediction mechanism categories and we reason the need of the proposed novel Self-attention based Encoder-Decoder equipped with the geospatial density transformation mechanism, that tackles the problem of human density prediction.

3.1. Time-series prediction mechanism

1. Statistical methods (Faghih et al., 2020) are based on the assumption that the data are stationary, They may use one polynomial for the Autoregression (AR), which regresses the variable on its own past values, and the Moving Average (MA) polynomial which is a linear combination of error terms occurring contemporaneously and at various times in the past. The sum of these two polynomials gives the ARMA model. In case the data are not stationary then ARIMA model can be used. ARIMA is based on ARMA but also includes the integration part which is a number of differences in the sequences of data observations.
2. Linear Regression (LR) (Fernández-Delgado et al., 2019) is based on the assumption that the output of the model is a linear function of the input variables. Lasso, Ridge and Elastic Net (EN) are variations that also assign a regularization penalty. Huber regression is a variation robust in outliers, whereas Passive Aggressive Regression (PAR) is an online regression approach. We also experiment with the Stochastic Gradient Descent Regression (SGDR) which is an extension of the stochastic gradient descent classification to the regression case. Finally, Least-Angle Regression (LARS), the Random Sample Consensus (RANSAC) and the Lasso model fit with the Least Angle Regression (LLARS).
3. Machine Learning (ML) (Xie et al., 2020) is a broad field, spanning an entire family of different techniques. However, all these techniques have the common principle that they automate the model building from data without being explicitly programmed. From this category we first examine the Support Vector Regression (SVMR), which is based on drawing the maximum margin hyperplane in an n-dimensional feature space. Following, we study the Regression Trees (Extra, CART), which is based on tree structures combined with decision rules. Finally, we evaluate the K-Nearest Neighbors (KNN) algorithm, which is based on the average of the values of K Nearest Neighbors of the testing instance.
4. Ensemble ML (Raj S. and M., 2021) are models that include multiple weak predictors and aggregate their individual outputs in order to improve their performance. They are mostly divided in the bagging methods that decrease the variance error, such as Bagged Decision Trees (Bag) and Random Forest (RF), and the Boosting methods that mostly decrease the bias error, such as Adaboost and Gradient Boosting Machines (GBM). The Bagging methods include homogeneous weak predictors that learn independently and in parallel while the predictors of the Boosting methods learn sequentially and adaptively.
5. Deep Learning (DL) (Luca et al., 2021) is a prominent subfield in ML that includes Artificial Neural Networks (ANN) with different types of hierarchical layers. In mobility modeling and prediction the Feed-forward, Long Short-Term Memory (LSTM), Convolutional layers (CNN) and many variations and combinations of them have been used successfully. Each of them captures the spatio-temporal dependencies of the mobile entities through different representation formulations.
6. Encoder-Decoders (ED) (Luca et al., 2021) are DL topologies with one ANN that compress the input feature vector in a latent space and one ANN to decode the latent vector to the output feature vector. Various types of DL models have been used in the literature for the encoder and decoder such as simple LSTM for both of them (LSTM-ED), bidirectional and simple LSTM (BD-LSTM-ED), unidirectional and bidirectional LSTM (Uni-BD LSTM ED, also named HB ED) and CNN encoder and LSTM decoder (CNN-LSTM ED). Special emphasis has been given to the ED that also include an attention mechanism which mimics cognitive attention and solves the bottleneck problem focusing on the most significant parts of the mobility features. A specific architecture of ED topology is the transformer (TRNF) with significant results in sequence to sequence problems which also lately adapted in mobility challenges.

3.2. Prediction mechanisms for mobility

Regarding the use of some of the above techniques specifically for the mobility prediction problem, only a few studies exist. For example, a work that compares statistical time series with DL methods for multi-step crowd distribution (Cecaj et al., 2020) has shown that generally statistical time series methods have better performance than simple DL models. But, an ED with a CNN for encoder and an LSTM for decoder outperforms the other DL models and it has similar performance to that of ARIMA. Comparing ED CNN-LSTM with ARIMA, we see that both approaches have advantages and limitations. The ARIMA has slightly better performance in the mean error metrics for most time-steps. In contrast, DL approaches reduce the maximum errors and they are more robust to spikes and sudden changes in the sequential values. Other important conclusions from this work are that there is a steady increase in the forecasting errors when going from a few steps forecast to more steps in the future. In addition, when the data sample size grows, DL methods significantly improve their predictive performance. This work presents

experimentally that the ED has the potential to be the state-of-the-art approach for multi-step density prediction using only an off-the-shelf ED. Our work goes a step beyond by trying to understand theoretically how the ED topologies can be adapted to the mobility characteristics. Furthermore, we conduct research with different ED topologies that also include the attention mechanism.

Our work is also related to the WiFiMod (Trivedi et al., 2021) model which uses an off-the-shelf TRNF taking into consideration long term mobility dependencies and multiple spatial scales. Even though this study also makes density prediction, it differentiates from our work because it focuses only on indoor modeling. The authors explain that the indoor mobility modeling has major differences from outdoor mobility because in a finer spatial scale, the mobility becomes more frequent, the prediction space expands and there is a more complex sequential periodicity.

Two important advances in time series forecasting that worth to be investigated by the mobility researchers but do not match with our work are the AR-NET (Triebe et al., 2019) and the DeepAR (Salinas et al., 2020). The AR-NET combines the best of traditional statistical models and neural networks using the stochastic gradient descent for the estimation of the AR weights. AR-NET can deal efficiently with long-range dependencies with fine granularity data. Nonetheless, as the authors mention, AR-NET currently has been designed for one-step forecasting while they leave multi-cast forcasting as future work. The DeepAR is based on an autoregressive recurrent neural network model and has the ability to learn a global model from multiple related historical time series. Its main advantage is that it provides probabilistic forecasts which is a characteristic we do not require in density prediction.

Recently, an Encoder-Decoder with Attention mechanism has been used in mobility prediction but for different challenges, such as the trajectory prediction (Zhou et al., 2019) and the prediction of users next PoI (Gao et al., 2019). In addition, it is important to mention that the Attention mechanism has different variations such as the (a.) Multi-dimensional Attention, (b.) Hierarchical Attention, (c.) Self Attention, and (d.) Memory-based Attention (Hu, 2020) just to mention the most popular. These variations can be used in different ways and have a different role in the DL ED topologies.

3.3. Comparison of related works

In all of the aforementioned works, related to mobility prediction, only a subset of the possible solutions from the 6 categories were used. This paves the way for a research investigation and for an analysis opportunity to examine which solution is the most appropriate for the MHDP. To the best of our knowledge, this is the first research endeavor of such a holistic analysis that tries to expose the limitations and advantages of the six categories in the context of urban mobility.

Some of these limitations can be extracted by analyzing the main functional blocks of each method. For instance, the methods of the categories from 1 to 5 have the limitation that they are not designed to provide predictions jointly in the spatial and temporal component. In addition to that, the statistical methods (category 1) are based on the stationarity assumption, which does not always characterise mobility data. LR models (category 2) include simplistic methods based on the assumption that the data have linear dependencies and can be represented using a linear model. Unfortunately, this assumption does not always characterise mobility data. The ML, ensemble methods and DL (categories 3, 4 and 5) can work with no-stationary and non-linear dependent data. However, their multi-output outcomes are limited to only one single step for the multiple ROIs or multi-step predictions for one single ROI. ED models (category 6) can provide sufficiently multiple output predictions in a forecasting time window but they are intrinsically designed to process natural language data. From a first look, it seems that the ED models may be a promising solution for the MHDP, yet they have not been designed for the particular problem. In other words, this means that there is a research gap since the current ED models have the structure to provide multi-step predictions but they have not been tailored for urban mobility data.

Thus, the key theme of this paper is to address the Multistep Human Density Prediction problem, by comparing and evaluating the related machine learning methodologies, while in the end proposing an innovative Self-attention based Encoder-Decoder and a geospatial density transformation mechanism. The latter, is a secondary problem that we address, namely, how to design an ED model to process efficiently urban mobility data in order to provide accurate density multi-step and multi RoIs predictions. It should be noted, however, that the novelty of this approach is not just the reuse of a state-of-the-art model in the urban mobility context but the optimization of the particular model through the design of an innovative Attention based ED that leverages the spatial and temporal properties of urban mobility data. This proposed model, as we will see in the experimental evaluation, surpasses the accuracy of all other methods.

4. Attention based encoder-decoder

A geospatial area, in which people move, can be separated in a set of POIs or ROIs (Kuo et al., 2018). A POI is defined as a specific physical location that people find useful or interesting and visit with high frequency. A ROI in the geospatial domain is defined as a polygonal selection in a 2D map that is important to be examined. The two terms, for the purpose of our research, can be used interchangeably since in both of them a high density of people concentration can be noticed.

Accordingly, in every POI/ROI people are concentrated and their density is changing over time. The POIs/ROIs are characterized by geospatial properties which affect the mobility and the flow of people among them. These factors make us to design a prediction model that should leverage: (a.) the geospatial properties among the POIs/ROIs, (b.) the number of people grouped in every POI/ROI and (c.) the time that is a key factor that shows how people density changes. Regarding the factor (a.), we propose a geospatial density value transformation that scales the number of people grouped in a POI/ROI, the number of people grouped in its neighborhood POIs/ROIs and the length of the borders between the ROIs. For the factor (b.), we empirically know that the density of people has a self similarity property with temporal dependencies (Trivedi et al., 2021) and can be predicted through a time series approach. Lastly, the time component (c.) can be leveraged through a sequence to sequence approach that takes as input a look back window of n previous values and provides m sequential predictions in a future time window.

As an example, Fig. 1 depicts an urban-area of interest, that we will also examine later in the experimental evaluation. We see that different regions are characterized by polygonal borders and a representative centroid. The density of people change over time and it is illustrated with different colors, according to the depicted heat bar. For each ROI we predict the density of people given the number of people it has in the n previous steps and the number of people in its neighboring regions. The predictions are multistep meaning that the models provide forecasts for the density in a sequence of m steps.

In the next subsections, we will firstly present the overview of the model and then we will focus on the technical details of our proposed model.

4.1. Model overview

The strong part of our proposed model is the mapping ability between the input and output sequences. The relationships and the prediction between the sequences is enhanced by the self-attention mechanism, as explained later, that focuses on the most significant parts of the temporal density dependencies. The attention mechanism learns which patterns in the input sequences should be considered as relevant and which as background noise in order to predict the sequential density

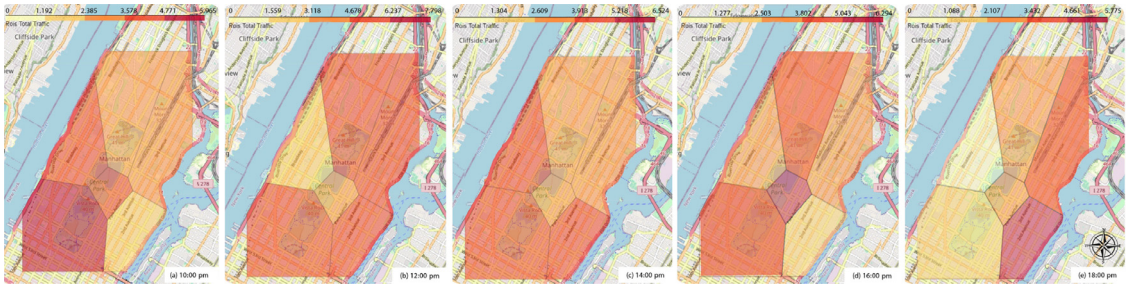

Fig. 1. Human density evolution over time.

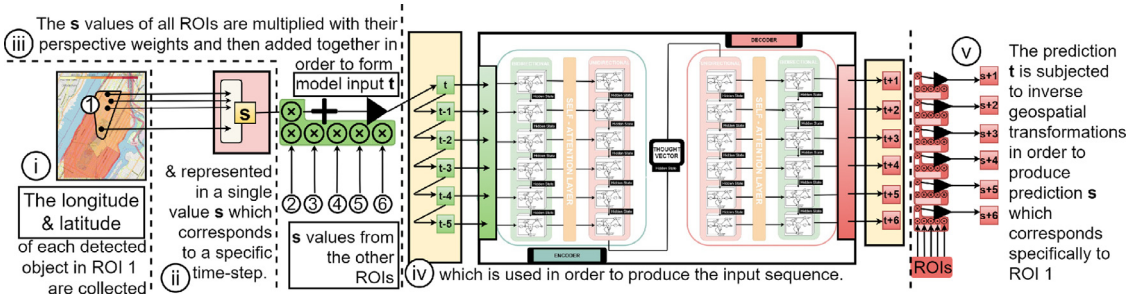

Fig. 2. The pipeline of human density prediction.

values. However, first and foremost, we need to understand which are the inputs and the outputs of our prediction model.

In more detail, we use an end-to-end mobility prediction model that begins from the sensoring of the geolocation data in constant intervals and ends up in the prediction of the human density distribution in a set of POIs/ROIs as it is depicted in the Fig. 2. The sensored mobility data, in its simplest form, are sequences of timestamps, latitude and longitude. The mobile entities that they represent can be physical entities or mobile devices. The output of the human density can be expressed either as a scalar number declaring the percentage of people in each POI or the real amount of people.

Furthermore, an additional characteristic of our model is that we assign the users to the nearest ROI and aggregate them in a single value s that represents them. The processing of aggregated values of people's density has three benefits compared to the processing and prediction of batches with individual geolocations. Firstly, it outputs smaller errors because it does not aggregate the errors of all the individual predictions. Secondly, it is more computationally lightweight because it makes less calculations. Finally, it preserves the privacy of individuals according to the General Data Protection Regulation (EU 2016/679 (GDPR))- the EU regulation law on data protection and privacy.

After the sensoring step (i) and representation step (ii) of individuals geolocations, which are also depicted in the Fig. 2, the time series are constructed by the ordered transformed density values of POIs/ROIs step (iii). Each POI/ROI has its own version of time series based on its own spatial properties (i.e. centroid distances and borders sizes) and the temporal users' mobility behaviors step (iv). As an example, users may have different mobility behaviors in regions with museums than regions with open-air concerts. Each POI/ROI has its own univariate prediction model with a self-attention based ED. As we will discuss in the experimental evaluation, the univariate models with the geospatial preprocessing have better performance than the multivariate models. At the last step (v) we apply the inverse geospatial transformation in the ED outputs to take the real predicted density values.

4.2. Geospatial feature engineering

As stated above, our input data are generated by sensoring devices, such as GPS tracking tools, which provide raw features of timestamps, longitudes and latitudes. These raw mobility data contain latent knowledge regarding the density of people. A feature engineering technique using geospatial domain knowledge can transform the raw data into a more suitable representation for the input of the ED model. The geospatial feature representation (Geo) leverages the length of the region borders and the contextual human density information.

In particular, we combine the representations of the various regions in a single variable which can be digested by the univariate proposed model. Given that the current density state of a region could affect the future states of other regions, two main points should be properly considered. The first point is to decide which of the other regions are affected by considering (a.) the data based on the activity of people, (b.) the actual size of the regions formulated and (c.) the established time-steps. Additionally, it is safe to assume that during each time-step an individual can traverse two regions at most. Thus, the state of a region can only be affected by the states of its neighboring regions within the time-frame which is currently being examined. The second point is the extent at which the states of the neighboring regions are affected. The next state of a region depends on the current states of its neighboring regions in accordance with the length of the borders they share. A longer common border between two regions means that it is statistically more likely for people to cross it.

In order to extract this inter-regional correlation it is essential to incorporate two additional mechanisms. The first mechanism is in charge of the formulation of the variables in an $Input\_Matrix$, which contains information regarding the states of all the regions. The backbone of this mechanism is the implementation of the $r \times r$ weight matrix named $Weight\_Matrix$ which contains representations of the r regions, normalized in a $0-0.5$ zone. For instance, the first row of the $Weight\_Matrix$ will contain information regarding the borders that the region i forms with the other regions. In the case of not neighboring regions between two ROIs, it will take the value 0. The declared border that a region forms with its own is equal to the sum of the borders which are formulated with the rest of the regions. In this way, the sum of the weights of each row will be equal to 1, the elements of the main diagonal will be equal to 0.5 and the rest of the weights of each row will have a sum which is equal to 0.5. Following, the states of all regions for each one of the m specified time-steps are collected, thus creating a $r \times m$ $Weight\_Matrix$. The two matrices are then multiplied with each other.

Out of the r rows of the resulting matrix only the one which corresponds to our area of interest is going to be selected. The following equation describes the formulation of our model input.

$$Model\_Input[i][j] = \sum_{l=1}^{r}(Input\_Sequence\_Matrix[i][l] \cdot Weight\_Matrix[j] \cdot k) \quad (1)$$

where i is the index of the input sequence, j is the index of the area of interest, and k is an additional weight vector which was introduced to enhance the efficiency of the model. In more detail, to guarantee the generality of the Eq. (1) and to estimate the optimal significance between different ROIs and their neighbors, we multiply the resulting vector with the k vector, which is the updated weighted matrix. This updated matrix entails weights that correspond to the significance of each region in accordance to our area of interest. Much like before, the elements of the k should have a sum which is equal to 1. The values of this vector can be formulated via processes such as Grid Searching. The final product is named $Model\_Input$ (Eq. 1) and it is consumed by the proposed Self-attention based ED.

The second mechanism is responsible for transforming the produced prediction sequence back into a form which corresponds to only one specified region. In order to achieve this functionality every prediction output is subjected to the transformation which is described in the Eq. (2).

$$Prediction\_Specific[i][j] = \frac{1}{k} \cdot (Prediction[i][j] + Input\_Matrix[i][j] \cdot k - \sum_{l=1}^{r}(Input\_Sequence\_Matrix[i][l] \cdot Weight\_Matrix[j])) \quad (2)$$

The Eqs. 1 and 2 are derived by the geometrical properties of the ROIs based on the assumption that more people can move between two ROIs through a longer border than a shorter one.

4.3. Sequence to sequence prediction

As mentioned before, we treat the MHDP problem as a time series problem. For our study, the look-back window of a time series forms a sequence of previous human density values. The multi-step-ahead prediction also constitutes a sequence of density values. This intrinsic structure of the input and output data, makes us to use a sequence to sequence (seq2seq) approach.

Hence, we resort to Encoders-Decoders (ED) models that their structures allow us to follow this seq2seq approach. Specifically, (ED) with RNN is a prominent DL model that maps the input sequence to the output sequence. RNNs neurons send feedback signals to each other through hidden states and keep prior inputs in memory while they process the current inputs and outputs. The training of long sequences in simple RNNs has the vanishing gradient problem, which can be addressed by the gates of LSTM units. The gates regulate what information of the time series the model should learn, forget or remember.

Fig. 3 illustrates the form of the encoder, which processes the sequential input values and encapsulates the temporal and spatial information into the context vector, also named thought vector. The encoder utilizes a bidirectional and a unidirectional LSTM layer. The bidirectional layer provides one hidden state output for each time-step in an n-dimensional space form which is then utilized as input by the unidirectional layer. The synergy between the heterogeneous layers is capable of exploiting the temporal correlations in the look back-window leveraging past and both past and future information. This structural symmetry enables the decoder later on, to capture the sequential patterns in both directions.

The context vector is the last hidden state of the Encoder and the input of the Decoder. It is calculated by the hidden and cell states of the LSTM units. Its main role is to summarize the information of the input sequence in a fixed-length representation. This representation captures the similarity relations among the sequential density values and it can work as the compressed information that the decoder will unfold in order to predict the density values in the next steps of a POI/ROI.

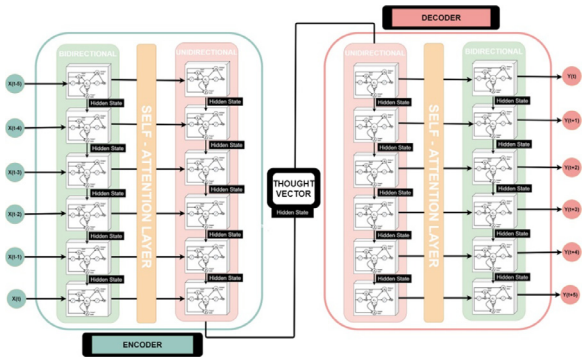

Fig. 3. Self-attention encoder decoder.

In addition, as it will be described in the next subsection we enhance the information of the context vector using a self-attention mechanism. This mechanism, focuses on the most informative patterns between the uni-directional and the bidirectional layers.

Regarding the decoder, it interprets the context vector and generates sequentially the output values. The decoder is also implemented by utilizing a unidirectional and a bidirectional LSTM layer with the self-attention mechanism. In addition, there is an interpretation layer and an output layer. The purpose of the fully connected interpretation layer is to interpret each time-step in the decoder output sequence and send the product to the output layer. We also wrap both the interpretation and the output layers inside a time-distributed wrapper. By doing so, the output provided by the decoder will be processed by the same fully-connected and output layer. This results at enabling the wrapped layers to be used for each time-step by the decoder.

4.4. Innovative encoder-decoder with attention mechanism

In this part of the section, we present the final model of our proposed mechanism, which is the Hybrid Encoder-Decoder (HB ED) model. The Encoder and the Decoder parts of the HB ED consist of Unidirectional and Bidirectional layers. In this particular architecture, we also utilized two Self-Attention (SATT) based layers alongside the recurrence-based ones. The first one, which is present at the encoder, receives as input the output of the bidirectional layer and its output will be utilized as input by the unidirectional layer. The second one, which is present at the decoder, receives as input the output of the unidirectional layer and its output will be utilized as input by the Bidirectional layer.

This final architecture, which is named SATT-HB-ED, is differentiated from the existing ones because the Attention mechanism is incorporated inside the encoder and the decoder parts respectively. The de facto use of the Attention mechanism is that it enables the decoder to examine the various states of the encoder and to provide an output by selectively focusing on specific elements from the sequence. At the same time, in this particular architecture, the Attention layer is utilized in a manner which aims to enhance the ability of the Hybrid bidirectional-unidirectional structures to encapsulate temporal dependencies. This enhancement manifests in the form of more robust encoding / decoding capabilities when compared to various alternative options.

In more details, the purpose of the incorporation of the Attention mechanism in the Encoding / Decoding process is to encourage the formation of homogeneous representations. The level of similarity to its counterparts each of the intermediate products of the encoding / decoding process holds, shall determine its impact on the overall encoding / decoding process. Since the Weight Vectors of the Attention mechanism are trained alongside the rest of the network and not in an independent manner, the same logical process is applicable in the opposite direction as well. Therefore, the elements of the input sequence which are more significant (in regards to affecting the output of the model) are more likely to be conceptually represented via the Encoding / Decoding pro-

cess in a similar way. This process enables the important elements of each input sequence to be represented in a more stable manner.

The Self-Attention mechanism, which is used in this particular architecture, is a variation of the Additive Attention mechanism (Cheng et al., 2016). The Additive Attention mechanism was introduced in order to provide more efficient sequence-to-sequence modeling by aligning the decoder with the relevant input elements. The variation which is used in this architecture implements the following process. The first recurrence-based layer produces the hidden states for each element of the input sequence. Following, the Alignment Scores between each element's hidden state and the rest of the hidden states are calculated. Each Alignment score corresponds to a specific element and is indicative of the similarity between the Hidden State of this specific element i and the Hidden States of the other elements j of the input. The Alignment Score for each pair is calculated using the Eq. (3).

$$scorealignment(i, j) = W_t \cdot tanh(W_i \cdot H_i + W_i \cdot H_j) \quad (3)$$

where H represents the Hidden State and W the Trainable Weight Vectors.

In the above equation, the $W_i \cdot H_j$ vector consists of N times as many columns as the $W_i \cdot H_i$ vector, since it is derived by the Hidden States of N elements of the input. Thus, the latter is added to each column of the former. The resulted vector is then passed through a tanh activation function and multiplied with another trainable vector. This process enables the Alignment Scores to be combined and represented in a single vector. The Attention Scores are then produced by passing the Alignment Scores through a softmax layer. The softmax layer will force the vector values to sum up to 1. By doing so, the importance of each time-step is properly encapsulated. Finally, the hidden states of the input are combined with their perspective Attention Scores in order to produce the Context Vector. The Context Vector is ultimately consumed by the second recurrence-based layer.

The proposed SATT-HB-ED is trained with the preprocessed data using the Adam optimizer, the backpropagation technique for the ED and also the Teacher Forcing technique for the Decoder. The Adam optimizer (Kingma and Ba, 2017) involves a combination of Momentum and Root Mean Square Propagation (RMSP), where both the exponentially weighted average and the exponential moving average of the past gradients are taken into consideration. The backpropagation technique (Hecht-nielsen, 1992) is used to minimize the cost function which evaluates the performance of the model by adjusting the network's weights and biases. The Teacher Forcing (Williams and Zipser, 1989) technique is a method that uses the ground truth from a prior time step as input, instead of model output from a prior time step. This is used to achieve faster and more efficient training for recurrent neural network models.

5. Experimental evaluation

The proposed methodology has been implemented and experimentally evaluated in the Python 3 programming language using the libraries NumPy, pandas, Scikit-learn, SciPy, GeoPy, TensorFlow 2 and its higher-level API Keras. The environment we used for the experiments is a Jupyter notebook of the Google Colaboratory. In order to help other scholars to replicate the same approach we provide the experiments' source code for any kind of reproduction in the second author's GitHub repository (Theodoros, 2021). Our research includes experiments for multiple time steps and ROIs with the twenty nine different methods described in Section 3 and the Attention based Encoder-Decoder presented in Section 4. In this part of the paper, we provide the figures and tables that summarize the most important outcomes. However, we have also uploaded the more detailed experimental outcomes in the GitHub repository.

For the evaluation of the proposed model we used two datasets. At first, we evaluated the performance of the SATT-HB-ED in a single-POI without using the geospatial transformation. Doing so, we focus our experiments on the ability of the SATT-HB-ED for sequence to sequence prediction. Next, we carried out extended experiments using a mobility simulator in an area of multiple ROIs/POIs. In this case we took into consideration the geospatial and crowd flow characteristics of the examined area and the human mobility behavior respectively. In the second set of experiments, we also used the proposed Geospatial transformation (Geo). In the Table. 1 we provide the data requirements for both data sets.

To evaluate the performance of our approach we use the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics, since they are the most popular metrics to be used in time series forecasting and ML regression models (Adhikari and Agrawal, 2013). MAE measures the average absolute deviation of forecasted values from original ones and it shows the magnitude of overall error, which occurred due to forecasting. The RMSE penalizes extreme errors occurring while forecasting. This emphasizes on the fact that the total forecast error is much more affected by large individual errors (i.e. large errors are much more expensive than small errors). For a good forecast, the obtained MAE and RMSE should be as small as possible.

5.1. Single-POI prediction outcomes and discussion

For the comparison and evaluation in a Single-POI task, we used the real-world dataset Crowdedness at the Campus Gym (Du et al., 2019). This dataset includes measurements of the number of people located in the campus gym of UC Berkeley. The measurements are taken every 10 minutes over more than one year and consists of more than 26,000 people counts. The reason we have selected this dataset is because going to a gym is a daily life activity for many people in urban areas. Additionally, with the current situation of the pandemic we have observed that many Covid-19 outbreaks were associated with the attendance of people at a gym. In particular, the visitors can stay in the POI of the gym for a significant amount of time, often more than one hour and there is a fluctuated flow of people during the parts of the day. The MHDP can provide the number of people crowded in the gym letting know the athletes and the coaches when they should go.

Many human out-of-home activities are characterized as stationary processes or can be transformed into stationary process by a difference transformation. Stationary means that the statistical properties of a time series do not change over time and it is the main assumption to guarantee the soundness of the model fit. We applied the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992) in the initial data values and it showed that the sequence values are not stationary. The non-stationarity is also visual perceived in Fig. 4, where we can see the density mean value to change in different parts of time. We continued with the differencing of the time series in order to eliminate the trend and seasonality. The new transformed time series was stabilized and the test showed that it is stationary.

In this first experiment, we have applied the ARIMA, linear regression (LR), and KNNR models. Additionally, we examined the three types of ED, namely, the LSTM ED, the Hybrid ED (HB ED) and the self-attention based Hybrid ED (SATT-HB-ED), which were described in Sections 2 and 4 respectively. The experimental outcomes are summarized in Table 2. The time-step was ten minutes and we predicted six steps ahead in a time window of one hour. Making experiments with different time granularities and number of steps in the look back window and look ahead window we derived the same conclusions regarding the applicability and performance of the methods. The multistep prediction in the ARIMA, LR, and KNNR took place with the recursive strategy while, in the ED with the direct strategy (Bontempi et al., 2013). The ED models have the intrinsic characteristic to provide multiple predictions directly by the output layer while the vanilla statistical, linear regression and machine learning models need multiple versions of the same model, one for each output. The latter approach increases the computational workload and we did not select it. For the sake of completeness, we mention that there are time series models that provide multi-outputs directly. This is the vector autoregression (Schimbinschi et al., 2017)

Table 1
Data requirements.

	Duration(days)	Time-step (min)	Regions (number)	People Counts
Campus Gym	430	10	1	26,000
Central Park	7	5	6	1,554,778

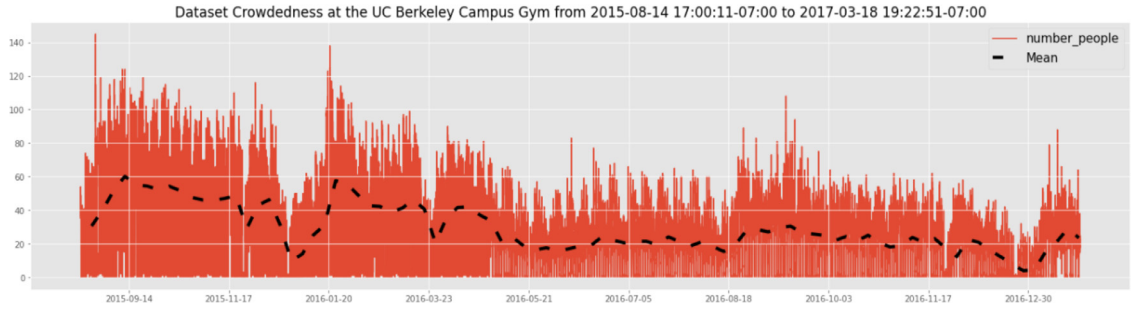

Fig. 4. Timeplot and vizual checking of stationarity. Different time periods have different statistical properties.

Table 2
Single POI Evaluation in the Dataset Crowdedness at the UC Berkeley Campus Gym.

	ARIMA	LR	KNNR	LSTM ED	HB ED	SATT-HB-ED
MAE	**4.008**	4.821	5.858	4.149	4.986	4.127
RMSE	6.225	6.647	8.640	6.096	7.11	**6.019**

model, but as we will see in the last experimental setup we selected the univariate because they had better performance.

From the Table 2 we see that in terms of RMSE the SATT-HB-ED has the best performance and in terms of MAE it has the second best performance. The ARIMA model has been optimized selecting the (p,d,q) values based on KPSS and Canova-Hansen tests (Canova and Hansen, 1995). The ED parameters have been learnt with the Adam optimizer but we didn't select the hyper-parameters i.e. number of layers, neurons, activation functions, etc. with a tuner such as Bayesian optimization or Hyperband. We just selected an ED topology based on our experience of previous time series tasks. This means that the ED performance can be further improved using a hyper-parameter optimization process.

The ED models seem promising even if they have not been hypertuned. The outcomes confirm that the ED approaches can achieve good performance in human density prediction. The RMSE metrics show that the SATT-HB-ED has the best performance in high-variability observations and when the sequential data have anomalous behaviors. Last but not least, it is obvious from the comparison of HB ED with SATT-HB-ED that the self-attention mechanism in the encoder and in the decoder improves the performance.

The above results corroborate that SATT-HB-ED has the following advantages in MHDP modeling compared to the others methods: i) It efficiently captures long-range dependencies and complex interactions. ii) It detects and focuses on the most relevant previous time steps against the target time-step. iii) It jointly leverages temporal dimensions (different time steps of a sequence), spatial dimensions (different regions of space) and different feature values. These characteristics are also present in the multi-ROIs experiments that follow. Specifically, in the following section, we will compare the performance of all the 6 categories mentioned in Section 2 and we will also apply the geospatial transformation.

5.2. Multi-ROIs prediction outcomes and discussion

For the evaluation and comparison of Geo SATT-HB-ED in multiple ROIs, we run a mobility simulation for seven days in the area of Central Park of New York. In every day of the simulation we examined approximately 200,000 to 230,000 pedestrians that roam around or stay in the same location. We partitioned the Central Park into six ROIs. We adopted a time step of five minutes, a look-back window of six steps and a look-ahead window also with six steps. We split the dataset into the training part with the first four days and the evaluation part with the last three days.

The Central Park covers a rectangular area of $3.41km^2$ with sides that are $4km$ in length and $0.8km$ width. In our simulation we also took into consideration many building blocks that surround it. The Central park has geospatial and smart city characteristics such as multiple attractions, every day activities, events, concerts, tours, the Central Park Zoo, the 21 official playgrounds and 8 lakes and ponds. All these make it an ideal area for study and experiment of mobility models. The visitors stay or walk through the central park, they remain, come in or go out in the different ROIs. The route of the visitors is affected by the geospatial characteristics of the terrain and the park attractions making some areas with low or zero concentration like the areas of lakes and some others with a high concentration like the Conservatory garden and the Rumsey Playfield. The lake of central park has also rowboats and gondolas but we limited this research only to pedestrians.

The area is partitioned into ROIs based on the Voronoi diagram and the k centroids generated by the k-means clustering algorithm (Du et al., 1999). Firstly, the k-means algorithm is applied in the training dataset in order to cluster all the recorded pedestrian geolocations. The clusters are created with the objective to minimize the intra-cluster distances of the pedestrians' geolocations and maximize the inter-cluster distance of the clusters centroids. Doing so, we have centroids with latitude and longitude being shaped coherently and by well separating the groups of pedestrians. Following, using the geolocations of the centroids and the Euclidean distance we draw the borders of the Voronoi diagrams as illustrated in Fig. 1. Each Voronoi cell defines the region of the corresponding ROI. The reason we follow this approach in order to select the ROIs instead of using predefined ROIs, is that the new Edge computing and wireless network infrastructures cover areas with this type of geospatial and number of connected users criteria (Chowdhury and De, 2021).

The trajectories of the pedestrians are generated with the software package named Simulation of Urban MObility (SUMO) (Lopez et al., 2018). SUMO is a highly portable, microscopic and continuous traffic simulation for handling large mobility networks. It has multiple types of transportations including pedestrians and comes with a large set of tools for different mobility scenarios creation. The simulation offers many realistic properties of the pedestrian mobility such as pedestrian-

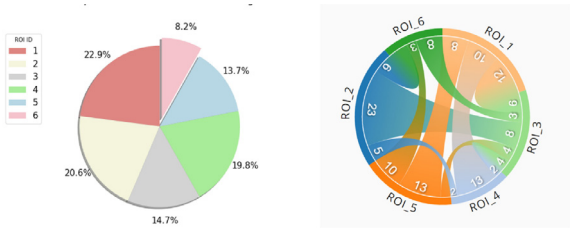

Fig. 5. Density of People (left) and mobility percentages among ROIs (right) in Central Park NYC.

pedestrian interactions when they are close, reasonable walking speeds and movement behavior.

Pedestrians interactions also include features such as the collision avoidance. In order to achieve this, the SUMO divides the lateral width of a lane into discrete stripes of fixed width, adding to the pedestrians the ability to overcome obstacles or slower pedestrians by moving to another adjacent stripe and proceeding. More complicated movement rules apply when moving on a walking area, where pedestrian paths cross in multiple directions. In that case, pedestrians follow a predetermined trajectory calculated at the beginning of the simulation. It is finally important to note that each pedestrian involves a list of rides, stops and walks. Rides are not implemented in our case, because no vehicles were included in the simulation. Stops correspond to non-traffic related activities such as working or shopping, while walks model trips taken by foot.

The ROIs cover different areas, number of people and mobility patterns. The size of the ROIs is constant in our experiments, defined by the Voronoi diagram and can be depicted in the Fig. 1. The distribution of people in the ROIs for three different timestamps is also depicted with different colors in Fig. 1. The total amount of people in the ROIs during the first four days of the simulation is depicted in the left of the Fig. 5. The chords in the right of the Fig. 5 represent the percentage of people that move from ROI-i to ROI-j during five minutes. These five minutes are selected randomly and it is obvious that the chord diagram changes over time. Every different use case and area of interest has different mobility patterns and statistical properties. Yet, this analysis is important in order to conceptually understand the challenges of the heterogeneity and dynamicity that a mobility model should tackle.

We compared the performance of the Geo SATT-HB-ED with the baseline and state-of-the-art models mentioned in the related work. We applied both the direct and recursive methods in order to develop the multistep forecasting as we did in the Section 5.1. The outcomes in terms of MAE and RMSE are summarized in Fig. 6. We see that the models are mostly grouped together based on the category they belong to. ML and linear regression models seem to have the worst performance with the highest MAE/RMSE. Next, ensemble methods and DL models follow with a significant improvement in the performance. Time series models and ED have the best performance. An unexpected outcome was the bad performance of Transformers (TRNF), since it is mentioned as the state-of-the-art method for many sequential problems. We made an error analysis and searched in the literature to reason the TRNF outcomes. The results of the error analysis can be explained by (Fan et al., 2021) who say that TRNF have inability to track long sequences, do not have access to higher level representations and cannot maintain a belief state. In contrast, SATT-HB-ED does not have these limitations. The long sequences can be tracked by the bidirectional and unidirectional LSTM layers and the belief state can be maintained in the encoder and the decoder.

Fig. 7 depicts that the Geo SATT-HB-ED has better performance compared to ARIMA in every ROI. We see that the ROI-1 and ROI-5 in both models have significantly lower accuracy than the other ROIs. This means that the historical data of the ROI-2, ROI-3, ROI-4 and ROI-6 are sufficient to train the prediction model compared to ROI-1 and ROI-5. In the error analysis, we saw the phenomenon that similar input patterns are mapped to different output sequences which also have high variability. This phenomenon was more intense in ROI-1 and ROI-5. The SATT-HB-ED can capture these mobility patterns better than ARIMA using a non-linear approach and giving the proper attention into the important parts of the long sequences.

In Fig. 6 we also see the performance of the geospatial scaling Geo SATT-HB-ED compared with not scaling SATT-HB-ED. We see a significant improvement in the RMSE and a slight deterioration in MAE. The geospatial transformation compress the information of the number of people in every ROI and the geospatial properties among the ROIs in one value. In case we disentangle these pieces of information and process them as different sequences of features we have a multivariate forecasting model. The related scientific literature contains many instances which showcased that the utilization of multiple features instead of a single one is beneficial to the model's efficiency to properly forecast future states. In the use case which is currently being examined, the application of multivariate forecasting would enable the model to simultaneously consume information related to all of the 6 regions. This is achieved via the use of a 6×6 matrix as the input sequence. Each column corresponds to a specific region and each row corresponds to a specific time-step. The output sequence is identical to the one produced by the univariate model.

In order to test the viability of a multivariate forecasting approach, we modified the SATT-HB-ED model in a manner which allowed it to leverage inputs from various regions. Table 3 shows these results and prove that the univariate approach is superior to the multivariate one both in terms of RMSE and MAE. This is due to the fact that the mobility-based correlations, which are formed between the various regions, are not strong enough to overcome the advanced complexity of the multivariate approach. This advanced complexity derives from the fact that the multivariate model has to digest a greater amount of information in order to form contextual relationships between the input and the output sequence.

In both experiments of Single-POI prediction and Multi-ROIs prediction we have seen a significant improvement in the accuracy using the SATT-HB-ED compared with the six categories of prediction models described in the related work. This happens because ED captures the long-range dependencies and complex interactions among the different time-steps of the data. Finally, the proposed geospatial feature representation improves further the accuracy leveraging the geometric properties of the areas of interest. Regarding the limitations of the proposed model, ED requires a significant amount of historical data. In addition the training process is computationally heavy, requiring the appropriate number of resources and training time.

6. Conclusion and future work

In this paper we have studied several multistep prediction mechanisms for the distribution of people in POIs/ROIs in an urban environment. We have seen that existing models cannot jointly process the geospatial and temporal properties of this challenging problem and cannot attain high accuracy. To this end, we tried to conceptualize what affects the density of people in a look ahead window and proposed a new ED-based mechanism called SATT-HB-ED. The experimental results with two different datasets confirm the applicability of our proposed approach.

Hence, transport planners and other related stakeholders can have the most accurate model regarding the density of people distributed in an urban area using the Geo SATT-HB-ED. Additionally, we believe that ED models can sufficiently capture long-range time dependencies and complex interactions among mobile entities, which constitute major mobility challenges. In addition, the modeling of physical characteristics of a mobility task using data transformation techniques can further improve the accuracy.

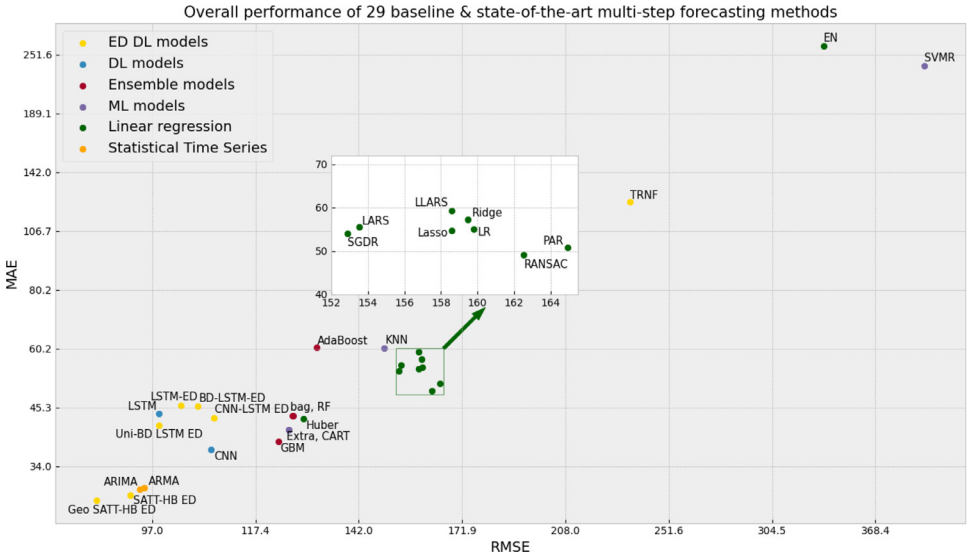

Fig. 6. Experimental results in terms of MAE & RMSE.

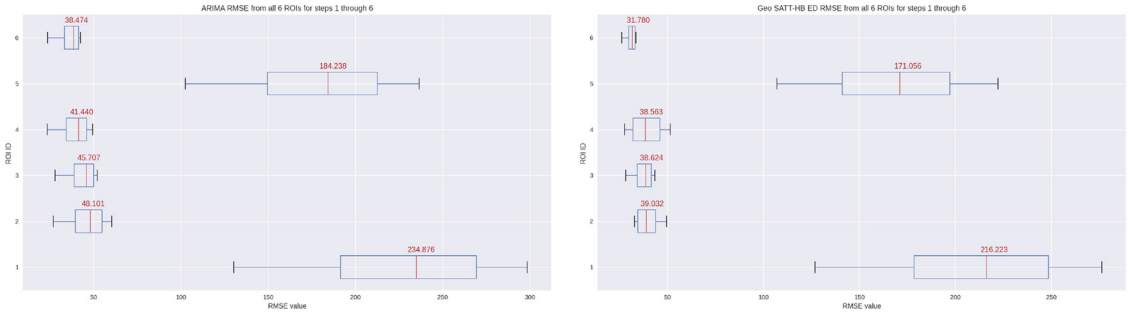

Fig. 7. ARIMA vs. Geo SATT-HB-ED.

Table 3
Comparison multivariate with univariate SATT-HB-ED.

		ROI-1	ROI-2	ROI-3	ROI-4	ROI-5	ROI-6
Multi- variate	MAE	143.145	90.573	103.262	179.017	179.839	53.088
	RMSE	277.001	119.649	124.315	225.338	316.104	76.303
Uni- variate	MAE	47.486	25.189	21.245	24.628	40.373	18.771
	RMSE	225.55	40.814	38.306	40.619	182.387	31.195

Our future work includes to adapt the proposed methodology in specific use cases and include additional covariants like parallel time series that have cross-correlation and semantic information of the ROIs. We believe that by modeling exogenous information we can find patterns from different knowledge domains that reason the mobility behavior of people. These patterns can be time dependent or time independent. This brings the challenge of how we can combine time series and batch data in a unified forecasting model. Lastly, we believe that the geospatial and mobility properties can be efficiently represented with graphs. Thus, in our future work we aim to also examine the graph neural networks for mobility prediction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

John Violos: Conceptualization, Methodology, Writing – original draft.

Acknowledgments

This work was supported in part by the CHIST-ERA-2018-DRUID-NET project ”Edge Computing Resource Allocation for Dynamic Networks”.

References

Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv:1302.6613 [cs, stat]*. http://arxiv.org/abs/1302.6613

Arias-Molinares, D., Julio, R., García-Palomares, J. C., & Gutiérrez, J. (2021). Exploring micromobility services: Characteristics of station-based bike-sharing users and their relationship with dockless services. *Journal of Urban Mobility, 1*, 100010. 10.1016/j.urbmob.2021.100010. https://www.sciencedirect.com/science/article/pii/S2667091721000108

Arman, A., Bellini, P., Nesi, P., & Paolucci, M. (2019). Analyzing Public Transportation Offer wrt Mobility Demand. In *Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities*. In *TESCA'19* (pp. 30–37). New York, NY, USA: Association for Computing Machinery. 10.1145/3364544.3364828.

Balcan, D., Gonçalves, B., Hu, H., Ramasco, J. J., Colizza, V., & Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of Computational Science, 1*(3), 132–145. 10.1016/j.jocs.2010.07.002. https://www.sciencedirect.com/science/article/pii/S1877750310000438

Barbosa, H., et al., (2018). Human mobility: Models and applications. *Physics Reports, 734*, 1–74. 10.1016/j.physrep.2018.01.001. https://www.sciencedirect.com/science/article/pii/S037015731830022X

Bontempi, G., Ben Taieb, S., & Le Borgne, Y.-A. (2013). Machine learning strategies for time series forecasting. In M.-A. Aufaure, & E. Zimányi (Eds.), *Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15–21, 2012, Tutorial Lectures*. In *Lecture Notes in Business Information Processing* (pp. 62–77). Berlin, Heidelberg: Springer. 10.1007/978-3-642-36318-4_3.

Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics, 13*(3), 237–252. 10.1080/07350015.1995.10524598. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524598

Cecaj, A., Lippi, M., Mamei, M., & Zambonelli, F. (2020). Comparing deep learning and statistical methods in forecasting crowd distribution from aggregated mobile phone data. *Applied Sciences, 10*(18), 6580. 10.3390/app10186580. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute

Chen, Q., et al., (2019). A survey on an emerging area: Deep learning for smart city data. *IEEE Transactions on Emerging Topics in Computational Intelligence, 3*(5), 392–410. 10.1109/TETCI.2019.2907718. Conference Name: IEEE Transactions on Emerging Topics in Computational Intelligence

Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. *CoRR, abs/1601.06733*. http://arxiv.org/abs/1601.06733

Chowdhury, A., & De, D. (2021). Energy-efficient coverage optimization in wireless sensor networks based on Voronoi-Glowworm Swarm Optimization-K-means algorithm. *Ad Hoc Networks, 122*, 102660. 10.1016/j.adhoc.2021.102660. https://www.sciencedirect.com/science/article/pii/S157087052100175X

De Saa, E., & Ranathunga, L. (2020). Comparison between ARIMA and deep learning models for temperature forecasting. *arXiv:2011.04452 [cs]*. http://arxiv.org/abs/2011.04452

Dechouniotis, D., Athanasopoulos, N., Leivadeas, A., Mitton, N., Jungers, R., & Papavassiliou, S. (2020). Edge computing resource allocation for dynamic networks: The druid-net vision and perspective. *Sensors, 20*(8). 10.3390/s20082191. https://www.mdpi.com/1424-8220/20/8/2191

Du, B., et al., (2020). Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. *IEEE Transactions on Intelligent Transportation Systems, 21*(3), 972–985. 10.1109/TITS.2019.2900481. Conference Name: IEEE Transactions on Intelligent Transportation Systems

Du, Q., Faber, V., & Gunzburger, M. (1999). Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Review, 41*(4), 637–676. 10.1137/S0036144599352836. Publisher: Society for Industrial and Applied Mathematics https://epubs.siam.org/doi/abs/10.1137/S0036144599352836

Du, Y., Gebremedhin, A. H., & Taylor, M. E. (2019). Analysis of university fitness center data uncovers interesting patterns, enables prediction. *IEEE Transactions on Knowledge and Data Engineering, 31*(8), 1478–1490. 10.1109/TKDE.2018.2863705. Conference Name: IEEE Transactions on Knowledge and Data Engineering

Faghih, S., Shah, A., Wang, Z., Safikhani, A., & Kamga, C. (2020). Taxi and mobility: Modeling taxi demand using ARMA and linear regression. *Procedia Computer Science, 177*, 186–195. 10.1016/j.procs.2020.10.027. https://www.sciencedirect.com/science/article/pii/S1877050920322948

Fan, A., Lavril, T., Grave, E., Joulin, A., & Sukhbaatar, S. (2021). Addressing some limitations of transformers with feedback memory. *arXiv:2002.09402 [cs, stat]*. http://arxiv.org/abs/2002.09402

Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S., & Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Networks, 111*, 11–34. 10.1016/j.neunet.2018.12.010. https://www.sciencedirect.com/science/article/pii/S0893608018303411

Gao, Q., Zhou, F., Trajcevski, G., Zhang, K., Zhong, T., & Zhang, F. (2019). Predicting human mobility via variational attention. In *The World Wide Web Conference*. In *WWW '19* (pp. 2750–2756). New York, NY, USA: Association for Computing Machinery. 10.1145/3308558.3313610.

Hecht-nielsen, R. (1992). III.3 - Theory of the backpropagation neural network**based on "nonindent" by Robert Hecht-Nielsen, which appeared in proceedings of the international joint conference on neural networks 1, 593–611, june 1989. ®1989 IEEE. In H. Wechsler (Ed.), *Neural Networks for Perception* (pp. 65–93). Academic Press. 10.1016/B978-0-12-741252-8.50010-8. https://www.sciencedirect.com/science/article/pii/B9780127412528500108

Hu, D. (2020). An introductory survey on attention mechanisms in NLP problems. In Y. Bi, R. Bhatia, & S. Kapoor (Eds.), *Intelligent Systems and Applications*. In *Advances in Intelligent Systems and Computing* (pp. 432–448). Cham: Springer International Publishing. 10.1007/978-3-030-29513-4_31.

Ilin, C., Annan-Phan, S., Tai, X. H., Mehra, S., Hsiang, S., & Blumenstock, J. E. (2021). Public mobility data enables COVID-19 forecasting and management at local and global scales. *Scientific Reports, 11*(1), 13531. 10.1038/s41598-021-92892-8. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Diseases;Health care Subject_term_id: diseases;health-care https://www.nature.com/articles/s41598-021-92892-8

Kapoor, S., Grace, D., & Clarke, T. (2017). A base station selection scheme for handover in a mobility-aware ultra-dense small cell urban vehicular environment. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 1–5). 10.1109/PIMRC.2017.8292760. ISSN: 2166–9589

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs]*. http://arxiv.org/abs/1412.6980

Kuo, C.-L., Chan, T.-C., Fan, I.-C., & Zipf, A. (2018). Efficient method for POI/ROI discovery using flickr geotagged photos. *ISPRS International Journal of Geo-Information, 7*(3), 121. 10.3390/ijgi7030121. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute https://www.mdpi.com/2220-9964/7/3/121

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1), 159–178. 10.1016/0304-4076(92)90104-Y. https://www.sciencedirect.com/science/article/pii/030440769290104Y

Lopez, P. A., et al., (2018). Microscopic traffic simulation using sumo. *The 21st ieee international conference on intelligent transportation systems*. IEEE. https://elib.dlr.de/124092/

Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2021). A survey on deep learning for human mobility. *arXiv:2012.02825 [cs]*. http://arxiv.org/abs/2012.02825

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS One, 13*(3), e0194889. 10.1371/journal.pone.0194889. Publisher: Public Library of Science https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194889

Raj S, S., & M, N. (2021). Ensemble human movement sequence prediction model with apriori based probability tree classifier (APTC) and bagged J48 on machine learning. *Journal of King Saud University - Computer and Information Sciences, 33*(4), 408–416. 10.1016/j.jksuci.2018.04.002. https://www.sciencedirect.com/science/article/pii/S1319157817303385

Saeik, F., et al., (2021). Task offloading in edge and cloud computing: a survey on mathematical, artificial intelligence and control theory solutions. *Computer Networks, 195*, 108177. 10.1016/j.comnet.2021.108177. https://www.sciencedirect.com/science/article/pii/S1389128621002322

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting, 36*(3), 1181–1191. 10.1016/j.ijforecast.2019.07.001. https://www.sciencedirect.com/science/article/pii/S0169207019301888

Sallah, K., et al., (2017). Mathematical models for predicting human mobility in the context of infectious disease spread: Introducing the impedance model. *International Journal of Health Geographics, 16*(1), 42. 10.1186/s12942-017-0115-7.

Schimbinschi, F., Moreira-Matias, L., Nguyen, V. X., & Bailey, J. (2017). Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. *Expert Systems with Applications, 82*, 301–316. 10.1016/j.eswa.2017.04.015. https://www.sciencedirect.com/science/article/pii/S0957417417302518

Singh, A., Baalsrud Hauge, J., Wiktorsson, M., & Upadhyay, U. (2022). Optimizing local and global objectives for sustainable mobility in urban areas. *Journal of Urban Mobility, 2*, 100012. 10.1016/j.urbmob.2021.100012. https://www.sciencedirect.com/science/article/pii/S2667091721000121

Theodoros, T. (2021). An innovative attention based encoder-decoder for multistep human density prediction. Original-date: 2021-10-06T15:09:31Zhttps://github.com/theodorosthd/An-Innovative-Attention-Based-Encoder-Decoder-for-Multistep-Human-Density-Prediction.

Triebe, O., Laptev, N., & Rajagopal, R. (2019). AR-Net: A simple auto-regressive neural network for time-series. *arXiv:1911.12436 [cs, stat]*. http://arxiv.org/abs/1911.12436

Trivedi, A., Silverstein, K., Strubell, E., Iyyer, M., & Shenoy, P. (2021). Wifi-mod: Transformer-based indoor human mobility modeling using passive sensing. *arXiv:2104.09835 [cs, eess]*. http://arxiv.org/abs/2104.09835

Verma, T., Sirenko, M., Kornecki, I., Cunningham, S., & Araújo, N. A. M. (2021). Extracting spatiotemporal commuting patterns from public transit data. *Journal of Urban Mobility, 1*, 100004. 10.1016/j.urbmob.2021.100004. https://www.sciencedirect.com/science/article/pii/S2667091721000042

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation, 1*(2), 270–280. 10.1162/neco.1989.1.2.270.

Xie, P., Li, T., Liu, J., Du, S., Yang, X., & Zhang, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion, 59*, 1–12. 10.1016/j.inffus.2020.01.002. https://www.sciencedirect.com/science/article/pii/S1566253519303094

Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*. In *ACAI 2019* (pp. 49–55). New York, NY, USA: Association for Computing Machinery. 10.1145/3377713.3377722.

Zheng, X., Han, J., & Sun, A. (2018). A survey of location prediction on Twitter. *IEEE Transactions on Knowledge and Data Engineering, 30*(9), 1652–1671. 10.1109/TKDE.2018.2807840. Conference Name: IEEE Transactions on Knowledge and Data Engineering

Zhou, F., Yue, X., Trajcevski, G., Zhong, T., & Zhang, K. (2019). Context-aware variational trajectory encoding and human mobility inference. In *The World Wide Web Conference*. In *WWW '19* (pp. 3469–3475). New York, NY, USA: Association for Computing Machinery. 10.1145/3308558.3313608.

John Violos is a researcher in the Dept. of Software Engineering and Information Technology at École de technologie supérieure. He was a member in the European Commission's Digital Single Market working group on the code of conduct for switching and porting data between cloud service providers. His research interests include Deep Learning, Machine Learning, Mobility Modeling.

Theodoros Theodoropoulos is a researcher in Department of Informatics and Telematics at Harokopio University of Athens. He is involved in several Eu-founded research projects such as Charity, Teaching and Accordion. His research interests include Deep Learning, Graph Neural Networks, Deep Reinforcement Learning, Cloud and Edge Computing.

Angelos-Christos Maroudis is currently a researcher in Department of Informatics and Telematics at Harokopio University of Athens. He is involved in the Eu-founded research project Charity. His research interests include Deep Learning, Graph Neural Networks and Time Series Analysis

Aris Leivadeas is currently an Associate Professor with the Dept. of Software and Information Technology Engineering at the Ecole de technologie Supérieure (ETS), Montreal, Canada. From 2015 to 2018 he was a postdoc in the Dept. of Systems and Computer Engineering, at Carleton University, Ottawa Canada. In parallel, Aris worked as an intern at Ericsson and then at Cisco in Ottawa, Canada. He received his diploma in Electrical and Computer Engineering from the University of Patras in 2008, the M.Sc. degree in Engineering from King's College London in 2009, and the Ph.D degree in Electrical and Computer Engineering from the National Technical University of Athens in 2015. His research interests include Edge Computing, IoT, and network automation and management. He received the best paper award in ACM ICPE'18 and IEEE iThings '21 and the best presentation award in IEEE HPSR'20.

Konstantinos Tserpes received the Ph.D. degree in the area of distributed systems from the School of Electrical and Computer Engineering, National Technical University of Athens, in 2008. He is currently an Associate Professor with the Department of Informatics and Telematics, Harokopio University of Athens. He has been involved in several EU and National funded projects leading research for solving issues related to scalability, interoperability, fault tolerance, and extensibility in application domains, such as multimedia, e-governance, post-production, finance, and e-health. His research interests include distributed systems, software and service engineering, big data analytics, and social systems. He is a member of the Editorial Board of Future Generation Computer Systems.