Human Trajectory Completion with Transformers

†§Junwei Ma, †Chao Yang, †Shiwen Mao, ‡Jian Zhang, §Senthilkumar CG Periaswamy, and §Justin Patton
†Dept. of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201
‡Dept. of Electrical and Computer Engineering, Kennesaw State University, Marietta, GA 30060
§The RFID Lab, Auburn University, Auburn, AL 36849
Email: {jzm0175, czy0017}@auburn.edu, {smao, jianzhang}@ieee.org, {szc0089, jbp0033}@auburn.edu

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838743

***Abstract*—With outbreak of the COVID-19 pandemic, contact tracing has become an important problem. It has been proven that maintaining social distance and isolating affected people are highly beneficial for curbing the spread of COVID-19, which all depend on identifying people's trajectories. However, the current interview-based approach is costly, and the existing mobile app-based schemes rely on complete and accurate data. In this paper, we propose a transformer encoder-based approach with spatial position embedding extracted using a graph Combinatorial Laplacian matrix to interpolate incomplete human trajectories. To model human trajectory, we propose a graphical embedded module to extract spatial features based on predefined location clusters. The incomplete trajectory sequences are first preprocessed into matrices and then used to train a deep transformer encoder network for trajectory completion. Our experiments using a real world Bluetooth Low Energy (BLE) dataset validate the efficacy of our proposed approach, which outperforms several baseline methods.**

***Index Terms*—COVID-19, Contact Tracing, Human Trajectory Completion, Graph Embedding, Transformers.**

I. Introduction

The outbreak of Coronavirus Disease 2019 (COVID-19) pandemic has become a serious threat since it spread rapidly worldwide. COVID-19 has been declared a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO) due to the dire situation. It has impacted everyone's daily life and forced governments to enforce strict administrative policies, including travel restrictions, city or district lockdown, quarantine, restraining order of work-from-home, and rapid responses for health emergencies. Worldwide the total number of cases reached 252M and 5.08M lives have lost by Nov. 12, 2021. The global economy has been contracted by 3.5 percent in 2020.

COVID-19 is more infectious than most other viruses and the carrier could be contagious without showing symptoms. Thus other people who have physical contact with, or even in close proximity to the carrier have a high risk of infection before the carrier has tested positive. Therefore, it is compelling to perform the so-called *Contact Tracing* procedure to prevent the virus transmission [1]. The current practice is to ask an individual who is tested positive about all the recently contacted people and identify who has a risk of being infected. Contact tracing [2] is usually accomplished by highly participant-depended manual interviews such as phone screening and questionnaires. With the detailed and accurate human mobility trajectory reported, individual contact information can be accessible and contact tracing will be easily performed. However, there exist many practical challenges in acquiring accurate and complete contact information. It is normally challenging for people to recall the person they have met, talked with, or stayed with in the last two or three weeks. For example, we can not identify everyone in the same restaurant a few weeks ago at a specific time, since most people there were unknown to you. In some cases, the situation is even worse; they cannot even identify where they have been. With the development of wireless technology and smartphone apps, localization-based methods are being used to extract human trajectories from the Global Positioning Systems (GPS) data, WiFi [3], Radio-frequency identification (RFID) [4], Bluetooth proximity detection [2], and fusion models [5]. However, such techniques' performance is heavily dependent on the completeness and accuracy of human trajectory data [1].

The human mobility *trajectory completion problem* is usually considered similar as the *trajectory prediction problem* in practice. Most human mobility trajectory studies focus on predicting a future location, which has been shown effective for human mobility [6]. Several classic approaches, such as matrix factorization [7] and Markov chain [8], have been first introduced. The Markov model approach is to built a transition matrix to model the action probability between locations from historical data. The authors in [9] proposed a model to predict locations and human movement dynamics by combing short mobility trajectory with social network structure. With the recent advances in deep learning, Recurrent Neural Networks (RNNs) [10], Long Short Term Memory (LSTM) [11], Gated Recurrent Unit (GRU) [12], Generative Adversarial Network (GAN) and Graph Convolutional Network (GCN) [13] have been applied to deal with the temporal sequence of trajectory prediction, with satisfactory performance demonstrated. In recent years, researchers have successfully exploited the popular attention mechanism [14] based transformer as a sequence-to-sequence model to learn the semantic contexts [15]. Graph has been used as a universal tool to describe complex networks, while the geographic features have been integrated with temporal trajectory sequence [16]–[18]. Some recent works have jointly explored the influence of temporal, semantic, social, and geographical contexts [19].

The human mobility trajectory completion problem bares certain resemblance to the trajectory prediction problem in practice. However, trajectory prediction only focuses on predicting a future location based on historical information. To solve the completion or recovery problem, it is usually necessary and helpful to learn the missing locations jointly from both past and future information. Note that, although in such an open world, people do have a wide range of possible locations, for this specific problem, i.e., the COVID tracing issue, we are only interested in collecting people's contacts during the COVID incubation period [2], which is the past 14-21 days. With the limited tracing period and restricted mobility, it is preferable to narrow down the tracing range and consider it as a problem with temporal and spatial constraints. In this paper, we propose a graph embedding transformer encoder based deep neural network to address the human mobility trajectory completion problem, which can be useful for contact tracing to curb the spread of COVID-19. The proposed approach consists of a data preprocessing module to extract segmented human mobility trajectory and cluster the locations based on pair similarity. We then exploit a graph embedding scheme to extract the geographical features to be integrated with the position encoded semantic. We further randomly mask one location in each trajectory segment to create incomplete human mobility trajectories. In the supervised training phase, we incorporate a five-layer transformer encoder based feature extractor, a five-layer deep convolution neural network (DCNN) decoder, and a softmax classifier, which employs the cross-entropy as loss function to measure the difference between true labeled data and the model output.

The main contributions of this work include the follows:

- To the best of our knowledge, this is the first attempt to tackle the human trajectory completion problem with a graph embedding transformer encoder approach.
- We develop the graph embedding scheme to extract the spatial features from the trajectory data to assist the training process.
- We implement the proposed approach and test it on a real world Bluetooth low energy (BLE) dataset for performance evaluation. The proposed scheme exhibits superior performance over several baseline schemes.

The reminder of this paper is organized as follows. In Section II, we discuss the motivation and challenges and provide the problem formualtion. In Section III, we present the proposed solution that is based on transformers. Our experimental study is provided in Section IV, and Section V concludes this paper.

II. Motivation, Challenges and Problem Formulation

In this section, we first examine the characteristics of the human trajectory completion problem, especially its significance for controlling COVID-19 spreading. Then the challenges in exacting complete human trajectories are discussed. Finally we formulate the trajectory completion problem.

A. Human Trajectory Tracing

As an important task in a variety of location-based social network (LBSNs) applications [9], human trajectory study has drawn increasing attention. It is widely used in applications such as personalized recommendations and contact tracing. Specifically, it is extremely helpful in the epidemic investigation related to the COVID-19 pandemic. Human mobility trajectories provide detailed information about locations visited together with the visiting time. With such temporal-spatial mobility trajectories of users, techniques such as spatial granularity can be applied to proximity analysis. Then a graph-liked proximity map could be derived to perform automatic contact tracing.

With the historical proximity map, if a person is reported to have COVID infection, we can use the proximity map to identify other people who have visited the same locations at the same time as the patient in the past one or two weeks The tracing time window could be determined by the corresponding health organization/authority based on the pandemic situation. We can find all the exposed individuals and prevent COVID from spreading to a larger population. Moreover, we can also use the human trajectory to find some locations that are frequently visited for further control actions such as sensitization or lockdown.

B. Technical Challenges

There are various types of data formats of human trajectories, such as GPS trajectory, urban camera monitoring system trajectory, and spatial-temporal check-in trajectory [5]. Many commonly collected human trajectories in the real world are based on GPS data, since it is a comprehensive satellite-based system that provides good accuracy. With the technological revolution of the Internet of things (IOT), the data collection of GPS trajectory becomes much easier by ubiquitous sensor devices. The more sensors such as smart-phones are deployed, the more diverse human trajectories could be recorded. Another widely deployed approach is utilizing the urban camera system, which is able to partially monitor human trajectory. It is convenient to recognize human identities by a camera and the entire trajectory could be reconstructed with abundant extracted snapshots. The state-of-the-art approach for human trajectory analysis is also based on the spatial-temporal check-in data. The advances in data collection have been driven by the developments in the worldwide cellular and WiFi networks.

In general, most human trajectory data suffers from the problem of incompleteness due to three main reasons. First, it is not every person that is willing to share his/her privacy, and therefore, only incomplete trajectory data is reported in many cases. Second, it is also common that people just forget to carry the GPS tracker or smartphone, resulting in partially measured trajectory data. Last is the technical limitation, such as poor GPS reception indoors and smartphones could temporally lost connection to the base station.

The problem of data incompleteness can exacerbate inference inaccuracy. For example, although many cities deploy large-scale traffic camera systems, not all roads are covered, resulting in incomplete historical trajectories. Considering the potentially poor quality of recorded images, it would be challenging to infer the correct trajectories. Such incomplete data introduces considerable uncertainties in recovering the missing trajectories over those uncovered areas.

In the contact tracing scenario, human trajectories are critical information for health officials to perform risk evaluation and decide whether further actions are to be taken. The existing approach for contact tracing is to manually interview the infected individuals. The data completeness and accuracy could be much lower compared with those approaches with automated recorded data (e.g., GPS or smartphones).

C. Problem Formulation

Considering the human mobility pattern as a spatial-temporal check-in sequence, we can define a trajectory *T* as:

$$T_{user} : (l_1, t_1) \rightarrow (l_2, t_2) \rightarrow \cdots \rightarrow (l_i, t_i), \tag{1}$$

where l_i contains the location information (e.g., longitude and latitude in GPS data) and t_i is the corresponding check-in timestamp. Unlike traditional interview-based methods, the locations here are determined by specific localization techniques and represent a dense and accurate map. Each tuple within the trajectory sequence refers to a single check detection.

In general, a continuous spatial-temporal human trajectory in chronological order is generated by each user during a time window from t_1 to t_i. More specifically, the trajectory sequence is usually temporally sparse [20] in real-world scenarios due to two main reasons. First, the user actually is not in the detection area, and she might leave the previously detected location. We cannot capture her trajectory until she comes back to locations that are covered by pre-installed sensors. The second reason is the data incompleteness issue that we already discussed. In this case, it would be desirable to recover the missing data segments, e.g., an unknown tuple (l_k, t_k).

A more challenging problem is what about multiple missing check-ins in the trajectory. One solution is to perform trajectory segmentation based on the holes in the long sequence. The assumption, in this case, is that the missing element is also sparse, meaning the majority of the trajectory has been captured. Moreover, the certain length of the trajectory prior to and after timestamp t_k are both known. Then it is possible to infer the unknown element in a segmented trajectory. Moreover, we transform a long sequence trajectory to multiple continuous subsequnces. This is to better feed the long sequence into the proposed transformer based neural network architecture. Details of such processing will be elaborated in the experimental study section. Thus, a trajectory sequence T can be represented in a matrix format, as:

$$T_{user} = \begin{bmatrix} (l_{1,1}, t_{1,1}) & (l_{1,2}, t_{1,2}) & \cdots & (l_{1,n}, t_{1,n}) \\ (l_{2,1}, t_{2,1}) & (l_{2,2}, t_{2,2}) & \cdots & (l_{2,n}, t_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ (l_{m,1}, t_{m,1}) & (l_{m,2}, t_{m,2}) & \cdots & (l_{m,n}, t_{m,n}) \end{bmatrix}.$$

where m is the index of the subsequences and n is the length of each subsequence.

To better understand the trajectory matrix, spatial grouping can be applied to map a trajectory from the high-dimensional location space to a coarser labeled cluster space [21]. It helps to mitigate the dimensionality problem, which causes considerable complexity in modeling and computation. Further, it is reasonable to assume that the adjacent locations could influence each other within a specific cluster due to the high infectiousness of COVID. Unlike traditional representation methods (e.g., one-hot), the spatial grouping or clustering aims to find the potential link between different locations. We perform clustering on the trajectory check-ins space, as:

$$T \in \mathbb{R}^{n \times m \times d} \Rrightarrow T \in \mathbb{R}^{n \times m \times c}, \tag{2}$$

where n is the trajectory segment length, m is the dimensions of segmentation in one trajectory, d is the location dimension and c is the cluster dimension. The location dimension c is pre-determined at the beginning based on the physical distribution of the location space. For location clustering, we employ a fused similarity estimation from spatial similarity and detection similarity.

For most human trajectories, it is reasonable to consider two adjacent locations as highly correlated [20]. Therefore, the spatial similarity can be modeled as:

$$S_{spa}(l_a, l_b) = \frac{1}{dist(l_a, l_b)}, \tag{3}$$

where $dis(l_a, l_b)$ is the Euclidean distance between the two different location identities a and b. The reason for using Euclidean distance is that the spatial clustering is performed in a closed and relatively small map in this study. We assume that the risks of infectiousness for healthy subjects are inversely proportional to their distance to the potential virus carriers.

The locations' reception similarity is modeled as the simultaneous check-ins at different locations for the same user. Due to overlapping coverage, it is possible to detect the same user in different locations (e.g., in the overlapping coverage of two WiFi access points). We model the reception similarity as:

$$S_{rec}(l_a, l_b) = \Pr(\exists (l_a, t_a) \in T_{user}, (l_b, t_b) \in T_{user} | t_a = t_b),$$

where $\Pr(\cdot)$ indicates the probability of mutual detection of two different location identities a and b. We can compute the inter-location similarity score $\sigma_{a,b}$ as:

$$\sigma_{a,b} = S_{spa}(l_a, l_b) * S_{rec}(l_a, l_b). \tag{4}$$

Based on the computed similarity score, clustering can be performed on the location space. Location clusters are generated as: $\alpha \in \mathbb{L}^{|c|}$. We obtain the cluster label represented trajectory matrix as follows.

$$\mathcal{T}_{user} = \begin{bmatrix} (\alpha_{1,1}, t_{1,1}) & (\alpha_{1,2}, t_{1,2}) & \cdots & (\alpha_{1,n}, t_{1,n}) \\ (\alpha_{2,1}, t_{2,1}) & (\alpha_{2,2}, t_{2,2}) & \cdots & (\alpha_{2,n}, t_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha_{m,1}, t_{m,1}) & (\alpha_{m,2}, t_{m,2}) & \cdots & (\alpha_{m,n}, t_{m,n}) \end{bmatrix}.$$

Therefore, our trajectory completion problem is to determine a missing element $(\alpha_{i,j}, t_{i,j})$ in the label represented trajectory matrix $\mathcal{T}_{user}$.

III. Transformer-based Solution

We propose a transformer-based deep neural network approach to tackle the human trajectory completion problem. This solution is composed of two key components, including (i) the graph embedding representation process, and (ii) the transformer-based deep learning model. Our transformer-based model for trajectory completion is illustrated in Fig. 1.

A. Graph Embedding

Graph embedding has attracted great attention recently due to its high efficiency for many tasks such as link prediction and community detection. Human trajectory sequence is usually formed in a high-dimensional space, which is difficult for us to extract useful features. Graph embedding model is able to map the node-set from a high-dimensional domain into a flat

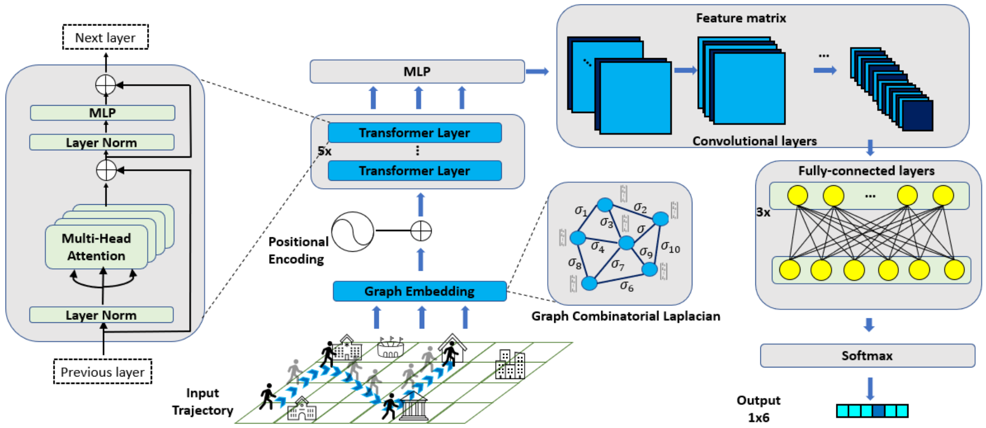

Fig. 1. The proposed graph embedding transformer model architecture.

low-dimension domain [7]. By establishing a cluster labeled trajectory representation, we will be able to leverage a graph embedding approach to extract the underlying spatial features in the vector space.

Consider a Graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ consisting of node set $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ and edge set $\mathcal{E} = \{e_{ij}\}_{i,j=1}^{n}$, where n denotes the number of vertices or nodes. Each edge e_{ij} describes the connection between two adjacent nodes v_i and v_j. Our model leverages an undirected binary static graph architecture to preserve structural properties. As normal graph embedding, our model is to encode nodes into a latent vector, which encodes a pre-determined location cluster into a label vector dimension [7].

Graph adjacency matrix $\mathbf{A}(\mathcal{G})_{n\times n}$ is a 2-D square matrix that indicates whether two nodes are connected in a graph. If two nodes v_i and v_j are connected, $\mathbf{A}(\mathcal{G})_{ij}$ is equal to 1; Otherwise, $\mathbf{A}(\mathcal{G})_{ij} = 0$. Graph degree matrix $\mathbf{D}(\mathcal{G})_{n\times n}$ is also a 2-D diagonal matrix that represents how many other nodes are attached to a node. Then the Graph Combinatorial Laplacian matrix $\mathbf{L}(\mathcal{G})_{n\times n}$ is computed as:

$$\mathbf{L}(\mathcal{G})_{n\times n} = \mathbf{D}(\mathcal{G})_{n\times n} - \mathbf{A}(\mathcal{G})_{n\times n}. \tag{5}$$

Similar to the conventional Laplacian operator, a normalized graph Laplacian matrix is modeled as a graph structural representation. We further model each row $\{x_1, x_2, ..., x_n\}$ of the graph Laplacian matrix as the graph node linear embedding vector. In general, two nodes close in the graph always have a smaller Cosine distance between the corresponding embedding vectors.

B. Deep Learning Model for Supervised Training

The structure of our transformer-based deep neural network is shown in Fig. 1. It consists of a transformer encoder [14] and a deep convolution neural network (CNN) [3] to perform a classification task. We leverage the transformer encoder as a feature extractor and the multi-layer CNN to estimate missing elements in the trajectory [15]. The major component in the transformer is the *Multi-head Self-attention* [14], given by:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{6}$$

where the $\mathbf{K}$ matrix are the keys, the $\mathbf{V}$ matrix are the values, and the $\mathbf{Q}$ matrix is the query that maps against the keys to the output.

The attention mechanisms come from the basic idea that the human brain always focuses on the most significant part when facing abundant information, while the other parts are processed as irrelevant components. It is widely used in the natural language processing tasks such as speech recognition and machine translation [15]. The self-attention mechanism aims at relating elements at different positions in a sequence to generate the sequence representation. Moreover, multi-head self-attention is to perform self-attention in parallel that combines multiple self-attention results to produce the final self-attention score [14]. Thus, the transformer is more powerful in dealing with multiple relationships and positions in a sequence. By leveraging the multi-head self-attention mechanism, the transformer has been proved to achieve a good performance as a sequence-to-sequence model.

Our transformer model is carefully designed for capturing the trajectory pattern of a specific user. We employ a multi-layer transformer encoder to encode the features in the graph-embedded segmented trajectory matrix, formulated as:

$$h_t = \text{TransformerEncoder}(\mathcal{T}_{user} \rightarrow \mathbf{L}(\mathcal{G})), \tag{7}$$

where $\mathbf{L}(\mathcal{G})$ is the graph-embedded trajectory matrix which is in three dimensions, and h_t is the transformer encoder layer's cell hidden state.

In the supervised training process, we randomly mask a single element in each embedded trajectory segment with a vector: $(\xi_1, \xi_2, ..., \xi_n)$, to create incomplete trajectories. This mask is to let the transformer learn the position of the missing part in the trajectory. Historical trajectories are important for estimating future mobility patterns [20]. However, the temporal location sequence is not sufficient to extract the entire mobility trajectory pattern and the spatial features of locations need to be determined. Together with the graph embedding, not only the spatial features are captured, but also we consider the time sequence of a human mobility trajectory to perform the complete task.

After feature extraction, we then utilize a deep CNN to train for optimal weights [3]. We design a four-layer CNN to obtain a strong learning and representation ability for the output vector.

Next, we transfer the output of the last CNN layer's final cell hidden state $h_{cnn}^{final} : (z_1, z_2, ..., z_n)$ to a softmax function and obtain the output vector $\hat{y} : (y_1, y_2, ..., y_n)$, as:

$$\hat{y_i} = \frac{e^{z_i}}{\sum_{i=1}^{n} e^{z_i}}, \text{ for } i = 1, 2, ..., n. \tag{8}$$

We adopt the *cross-entropy* to evaluate the difference between the estimated label $\hat{y}$ and the original embedded label vector y before masking. To obtain the optimal weights, we would like to minimize the following training loss function:

$$\min_{\epsilon} \mathit{\Gamma}(\epsilon) = -\sum_{i=1}^{n} y \log(\hat{y}). \tag{9}$$

IV. Experimental Study

A. Dataset Description

We evaluate the performance of our proposed approach with experiments using an open-source human mobility dataset [22]. To get closer to the reality of contact tracing, a relatively small scenario was preferred. It is a real world dataset created from a collection of BLE data packets generated by people-carried beacons to record human mobility. The participating human subjects carry the beacon device in their daily routine inside a building on campus. There are about 32 signal receiver deployed in this three-floor facility, continuously gathering the BLE packets sent by beacons [22]. Fig. 2 shows the layout of one of the three floors. When a packet is received by the edge device, the message is stored. The individual message contains the beacon ID, the Received Signal Strength Indicator (RSSI), the timestamp of receiving, and the receiver ID. Basically, the location information within the trajectory is dependent on the continuous detection of the receiver. Since the receiver locations are known and stationary, trajectories are formalized with sufficient detecting messages. The data was collected with 46 participants in one month from September to October 2016. The BLE beacon devices were carried by the participants all the time during their occupancy in the building. The trace data was utilized in our experiment to evaluate the performance of the proposed transformer-based approach.

B. Data Processing

To obtain clean data for training, we selected the appropriate subject's data that contains sufficient check-ins. Based on the trial scenario and characteristics of human mobility, we further performed down-sampling in the temporal domain. So the final results of the preprocessed trajectory data are segmented as 5-minute continuous check-ins sequence with 10 seconds temporal-even sampling. Specifically, we generated 19,195 human trajectory segments for the entire dataset, and each segment contains a sequence of 30 locations. We acknowledge that the possible noise in location data has not been taken into consideration in this case, but will be address in our future work. Based on the Bluetooth receiver deployment layout, we calculated the inter-receiver similarity scores and grouped all the receiver locations into six clusters. The last step is to mask the trajectory and perform a classification estimation task. Due to the difficulty of determining the ground truth of the missing trajectory, to balance the trade-off between uncertainty and accuracy, only one check-in per segment was masked with an extremely valued vector. The preprocessed dataset was further divided into train and test sets with ratio of 9:1.

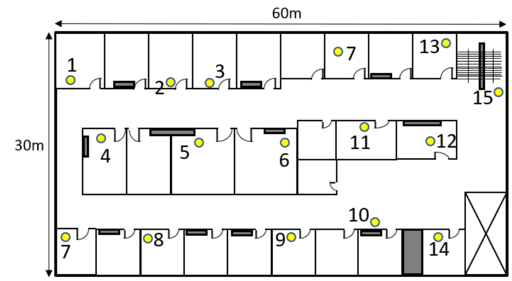

Fig. 2. Layout of one floor: signal receivers are marked by yellow dots.

TABLE I
ACCURACY PERFORMANCE COMPARISON OF DIFFERENT METHODS

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*	*ROC*
MLP	59.92%	0.478	0.599	0.481	0.557
LSTM	56.47%	0.433	0.565	0.420	0.832
Trans	64.43%	0.653	0.644	0.637	0.842
GE-Trans	**86.49%**	**0.868**	**0.865**	**0.865**	**0.960**

C. Results and Discussions

We use the metrics of model training loss and prediction accuracy to evaluate the proposed scheme. We also perform the same experiment with several baseline schemes for comparison, including (i) Transformer only (i.e, without graph embedding) termed as Trans, Multi-layer Perception (MLP) [5], and LSTM [11]. The proposed scheme is termed GE-Trans in the results. The exactly same dataset and data preprocessing procedure were applied for a fair comparison.

We find our proposed graph embedded transformer model achieves the highest training accuracy. Then, the trained models were executed on the test dataset for ten times and the averaged accuracy results are computed and presented in Table I. As the table shows, our proposed graph embedded transformer model achieves the highest accuracy of 86.49%, representing 26.57%, 30.02%, and 22.06% gains over MLP, LSTM, and Trans, respectively, which is also higher than the 75% accuracy of human mobility trajectory prediction reported in a recent work [19].

Fig. 3 presents the training accuracy curves for all the tested models over episodes. The curves show that transformer-based models, even without graph embedding, have a relatively better training accuracy performance compared with the traditional MLP and LSTM models. Our GE-Transformer model reaches 95% training accuracy after 100 episodes and all other baseline schemes cannot achieve an accuracy above 80% even after 300 episodes. Moreover, the LSTM and MLP models only achieve a stable train accuracy below 60%. On one hand, this is because these two models are not effective in learning the graphical feature of mobility trajectories. On the other hand, the attention mechanism based transformer models are more powerful when dealing with such spatial-temporal sequences, which bare certain resemblance to natural language with a

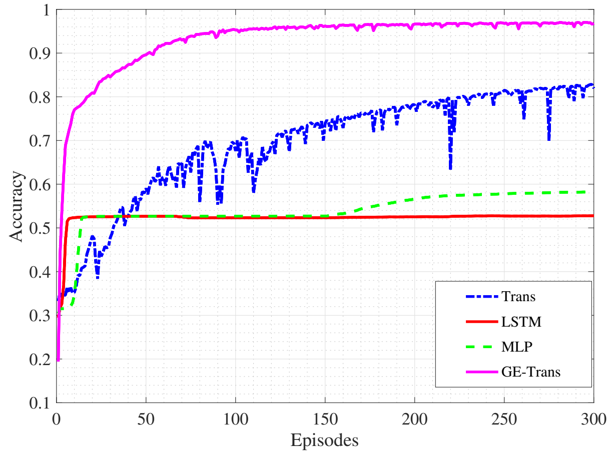

Fig. 3. Training accuracy of all the tested models.

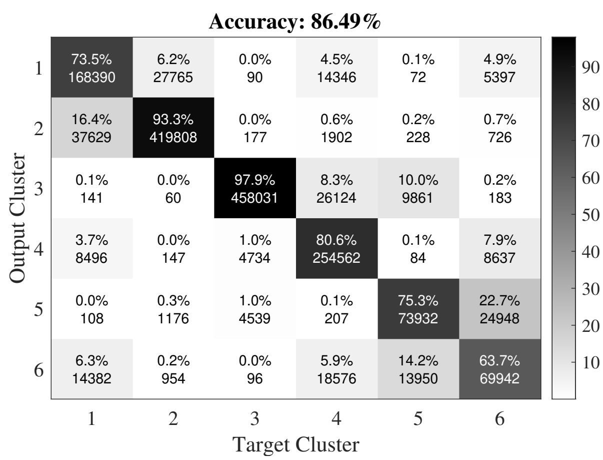

Fig. 4. Confusion matrix for estimation classification.

semantic context. In general, the graph embedded transformer model achieves the highest stable training accuracy in the first 300 episodes when compared with the other baseline models.

The detailed classification results based on our graph embedding transformer model are presented in the form of a confusion matrix in Fig. 4. It shows our model has a high positive rate overall in six location cluster categories.

V. Conclusions

In this paper, we proposed a graph embedding assisted transformer deep neural network model to address the human trajectory completion problem. We embedded the intrinsic spatial characteristics in the temporal location sequence to better understand the human mobility trajectory. Experimental study using real world BLE data was performed to validate our proposed deep learning model. The results demonstrated the superiority of the transformer-based approach. With the high efficiency and accurate classification, our proposed approach can be useful for contact tracing to combat the pandemic.

Acknowledgments

This work is supported in part by the National Science Foundation under Grants ECCS-1923163 and CNS-2107190, and through the Wireless Engineering Research and Education Center (WEREC) and the RFID Lab at Auburn University.

References

[1] Q. Hao, L. Chen, F. Xu, and Y. Li, "Understanding the urban pandemic spreading of covid-19 with real world mobility data," in *Proc. ACM KDD'20*, Virtual Event, July 2020, pp. 3485–3492.

[2] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, "A survey of COVID-19 contact tracing apps," *IEEE Access*, vol. 8, pp. 134 577–134 601, July 2020.

[3] W. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Network Science and Engineering*, vol. 7, no. 1, pp. 316–327, Jan./Mar. 2020.

[4] C. Yang, X. Wang, and S. Mao, "SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays," in *Proc. IEEE SECON'19*, Boston, MA, June 2019, pp. 1–9.

[5] X. Wang, Z. Yu, and S. Mao, "DeepML: Deep LSTM for indoor localization with smartphone magnetic and light sensors," in *Proc. IEEE ICC'18*, Kansas City, MO, July 2018, pp. 1–6.

[6] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.

[7] C. Cheng, H. Yang, I. King, and M. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *Proc. AAAI'12*, Toronto, Canada, July 2012, pp. 17–23.

[8] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Next place prediction using mobility Markov chains," in *Proc. Workshop Measurement, Privacy, and Mobility*, Bern, Switzerland, Apr. 2012, pp. 1–6.

[9] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. ACM KDD'11*, San Diego, CA, Aug. 2011, pp. 1082–1090.

[10] D. Yang, B. Fankhauser, P. Rosso, and P. Cudre-Mauroux, "Location prediction over sparse user mobility traces using RNNs: Flashback in hidden states!" in *Proc. IJCAI'20*, Yokohama, Japan, Jan. 2020, pp. 2184–2190.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint arXiv:1412.3555, Dec. 2014. [Online]. Available: https://arxiv.org/abs/1412.3555

[13] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, Sept. 2016. [Online]. Available: https://arxiv.org/abs/1609.02907

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. NIPS'17*, Long Beach, CA, Dec. 2017, pp. 5998–6008.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, Oct. 2018. [Online]. Available: https://arxiv.org/abs/1810.04805

[16] Q. Liu, S. Wu, L. Wang, and T. Tan, "Predicting the next location: A recurrent model with spatial and temporal contexts," in *Proc. AAAI'16*, Phoenix, AZ, Feb. 2016, pp. 194–200.

[17] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation," in *Proc. ACM SIGIR'15*, Santiago, Chile, Aug. 2015, pp. 433–442.

[18] Q. Guo, Z. Sun, J. Zhang, and Y.-L. Theng, "An attentional recurrent neural network for personalized next location recommendation," in *Proc. AAAI'20*, New York, NY, Feb 2020, pp. 83–90.

[19] H. Xue, F. D. Salim, Y. Ren, and N. Oliver, "MobTCast: Leveraging auxiliary trajectory forecasting for human mobility prediction," arXiv preprint arXiv:2110.01401, Sept. 2021. [Online]. Available: https://arxiv.org/abs/2110.01401

[20] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin, "Deepmove: Predicting human mobility with attentional recurrent networks," in *Proc. ACM WWW'18*, Lyon, France, Apr. 2018, pp. 1459–1468.

[21] S. Li, Z. Qin, and H. Song, "A temporal-spatial method for group detection, locating and tracking," *IEEE Access*, vol. 4, pp. 4484–4494, Aug. 2016.

[22] D. Sikeridis, I. Papapanagiotou, and M. Devetsikiotis, "CRAWDAD dataset unm/blebeacon (v. 2019-03-12)," [online] Available: https://crawdad.org/unm/blebeacon/20190312, Mar. 2019.