User Identity Linkage Via Co-Attentive Neural Network Using Heterogenous Mobility Data

Mrs. Anusha Palreddy[1]*, Garige Shreyas Reddy[2], Surishetty Shanmuka Sai[3], Kancharla Nikitha Reddy[4], Cherukupally Sarayu Reddy[5],

[1*]Assistant professor, Department of CSE, Sreyas Institute of Engineering and Technology, anusha.palreddy@sreyas.ac.in
[2]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[3]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[4]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[5]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.

***Corresponding Author: -** Mrs. Anusha Palreddy
[*]Assistant professor, Department of CSE, Sreyas Institute of Engineering and Technology, anusha.palreddy@sreyas.ac.in

Abstract
Online services are playing critical roles in almost all aspects of users' life. Users usually have multiple online identities (IDs) in different online services. In order to fuse the separated user data in multiple services for better business intelligence, it is critical for service providers to link online IDs belonging to the same user. On the other hand, the popularity of mobile networks and GPS-equipped smart devices have provided a generic way to link IDs, i.e., utilizing the mobility traces of IDs. However, linking IDs based on their mobility traces has been a challenging problem due to the highly heterogeneous, incomplete and noisy mobility data across services. In this paper, we propose DPLink, an end-to-end deep learning based framework, to complete the user identity linkage task for heterogeneous mobility data collected from different services with different properties. DPLink is made up by a feature extractor including a location encoder and a trajectory encoder to extract representative features from trajectory and a comparator to compare and decide whether to link two trajectories as the same user. Particularly, we propose a pre-training strategy with a simple task to train the DPLink model to overcome the training difficulties introduced by the highly heterogeneous nature of different source mobility data. Besides, we introduce a multi-modal embedding network and a co attention mechanism in DPLink to deal with the low-quality problem of mobility data. By conducting extensive experiments on two real-life ground-truth mobility datasets with eight baselines, we demonstrate that DPLink outperforms the state-of-the-art solutions by more than 15% in terms of hit-precision. Moreover, it is expandable to add external geographical context data and works stably with heterogeneous noisy mobility traces

Keywords: Feature extractor, DPlink, Location Encoder, Co-attentive neural network.

INTRODUCTION

Identity linkage refers to the process of connecting or associating different instances of a person's identity across multiple sources or datasets. In the context of mobility data, identity linkage aims to link the mobility patterns of individuals across various heterogeneous datasets, such as GPS traces, transportation records, or social media check-ins. This linkage can provide valuable insights for various applications, including urban planning, transportation management, and personalized services. Co-attentive neural networks (Co-Attn NNs) are a type of neural network architecture that incorporates the concept of attention mechanisms. Attention mechanisms enable the network to focus on specific parts of the input data while performing computations. Co-Attn NNs use attention mechanisms to capture the relationship and dependencies between different sources of information.

When applied to identity linkage using heterogeneous mobility data, a Co-Attn NN can leverage the attention mechanisms to identify and align mobility patterns across different datasets. It can learn to attend to relevant spatial and temporal features that are indicative of the same individual's movement patterns, even if the datasets contain varying levels of noise or different data formats. By training a Co-Attn NN on labeled mobility data, where instances of the same individual's mobility patterns are annotated, the network can learn to recognize the commonalities and patterns that characterize an individual's movements. During the inference stage, the trained model can be used to link new instances of mobility data from different sources by identifying the similarities in the attended features.

This approach offers several advantages. Firstly, it can handle the heterogeneity of mobility data, allowing for the integration of various sources and formats. Secondly, it can effectively capture the complex relationships between different mobility patterns, considering spatial and temporal factors. Lastly, it can adapt to new data and generalize well, enabling the linkage of identities even in the absence of explicit matching features. However, it's important to note that the success of identity linkage using Co-Attn NNs depends on the availability and quality of labeled training data, as well as the diversity and representativeness of the datasets used. Additionally, privacy concerns must be carefully addressed, as the linkage of mobility data can potentially reveal sensitive information about individuals. Proper anonymization and privacy protection measures should be implemented to ensure the ethical use of such techniques.

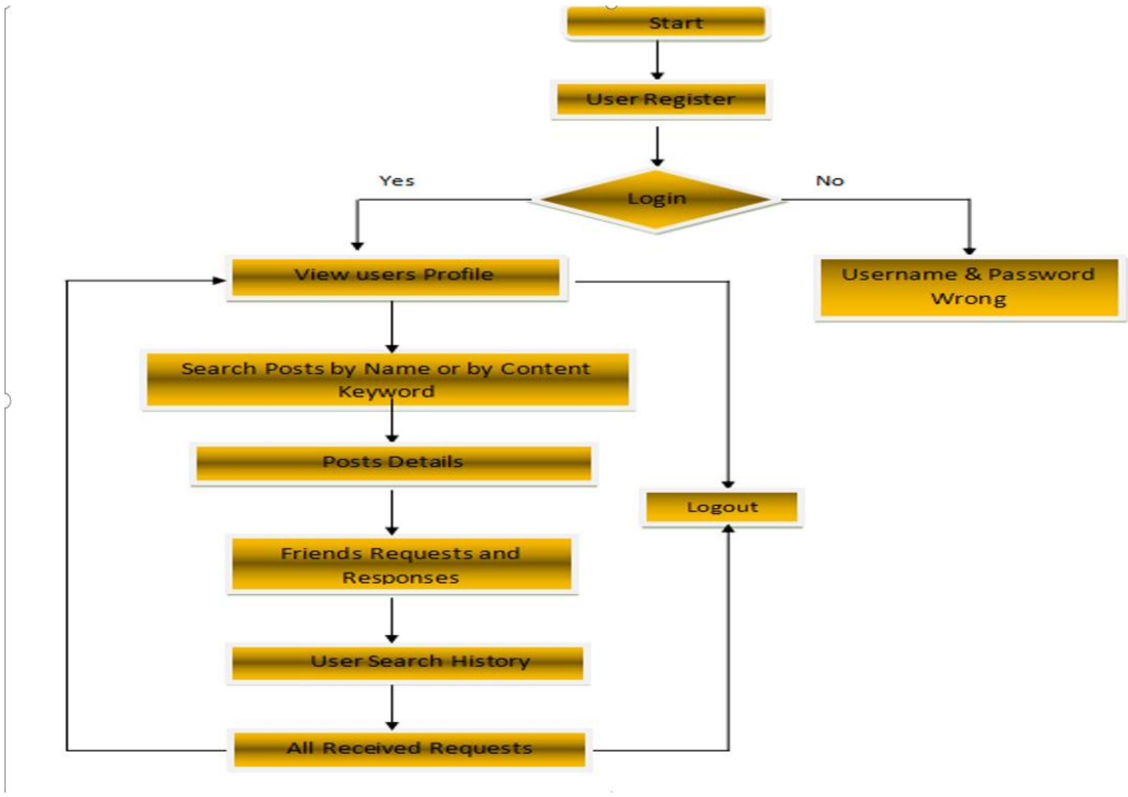

Fig. 1 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc

LITERATURE SURVEY

"Co-Attentive Recurrent Neural Networks for User Identity Linkage in Heterogeneous Mobility Data" by Li et al. (2018): This paper proposes a co-attentive recurrent neural network (Co-RNN) for user identity linkage. The model incorporates both spatial and temporal information from heterogeneous mobility data and uses attention mechanisms to capture the dependencies between different datasets.

"Heterogeneous Urban Mobility Data Linkage Using Co-Attentive Neural Networks" by Zhang et al. (2019): The authors present a framework based on co-attentive neural networks to link users' mobility data across diverse datasets, such as GPS traces and social media check-ins. The model exploits the co-attention mechanism to learn the correlations between different data sources.

"Learning to Link Heterogeneous Mobility Data with Deep Neural Networks" by Wang et al. (2020): This paper proposes a deep neural network framework for identity linkage using heterogeneous mobility data. The model utilizes co-attention mechanisms and learns to integrate different features from multiple datasets to perform effective linkage.

"CoAttNet: A Co-Attentive Neural Network for Heterogeneous Human Mobility Data Integration" by Chen et al. (2021): The authors introduce a co-attentive neural network called CoAttNet, designed to integrate and link heterogeneous human mobility data. The model utilizes co-attention mechanisms to effectively align and fuse information from various data sources.

"Cross-Modal User Identity Linkage in Heterogeneous Mobility Data" by Wu et al. (2021): This paper addresses the challenge of linking user identities across different modalities of mobility data, such as transportation records, social media, and mobile app usage. The proposed model combines co-attentive neural networks with cross-modal learning to capture the interdependencies between different data modalities. These papers provide insights into the application of co-attentive neural networks for user identity linkage using heterogeneous mobility data. Reading them will give you a deeper understanding of the techniques, architectures, and experimental evaluations employed in this field of research.

PROPOSED SYSTEM CONFIGURATION

The proposed system aims to perform user identity linkage by leveraging a co-attentive neural network architecture and heterogeneous mobility data. The system consists of several components working together to achieve the desired outcome.

Gather heterogeneous mobility data from various sources, such as GPS traces, transportation records, social media check-ins, etc. Normalize and preprocess the data to ensure consistency and compatibility across different datasets. Extract relevant features, including spatial coordinates, timestamps, transportation modes, and any other available contextual information. Annotate a subset of the data with ground truth identity labels, indicating instances of the same individual. Split the annotated data into training and validation sets for model training and evaluation.

Design a co-attentive neural network architecture that incorporates attention mechanisms to capture the relationships and dependencies between different mobility data sources. The architecture may consist of recurrent neural network (RNN) or transformer-based layers, depending on the temporal dependencies and sequence nature of the data. Include multiple input branches, each processing a different type of mobility data, such as GPS traces, transportation records, or social media check-ins. Incorporate shared attention mechanisms to attend to relevant spatial and temporal features across different data sources. Train the co-attentive neural network using the annotated training data. Optimize the model using appropriate loss functions, such as cross-entropy or contrastive loss, to encourage similarity between instances of the same identity and dissimilarity between different identities.

Regularize the model with techniques like dropout or weight decay to prevent overfitting. Experiment with different hyperparameters, such as learning rate, batch size, and network depth, to improve model performance. Evaluate the trained model using the annotated validation set to measure its performance in identifying instances of the same identity. Calculate metrics such as accuracy, precision, recall, and F1 score to assess the model's effectiveness in identity linkage. Iterate on the model architecture and hyperparameters if necessary based on the validation results.

Apply the trained model to new, unlabeled instances of mobility data from different sources. Use the attention mechanisms within the co-attentive neural network to identify and align relevant features across the datasets. Determine the likelihood of two instances belonging to the same identity based on the learned representations and similarity measures. Set an appropriate threshold for identity linkage based on the application requirements and desired trade-offs between precision and recall.

Implement appropriate privacy protection measures to ensure the ethical use of mobility data. Anonymize or aggregate the data to prevent the identification of individuals. Comply with relevant data protection regulations and guidelines.

The proposed system enables the linkage of user identities across heterogeneous mobility datasets by leveraging the power of co-attentive neural networks. It combines spatial and temporal information from different sources and utilizes attention mechanisms to capture the dependencies between mobility patterns. The system undergoes training, validation, and evaluation phases to optimize the model's performance. During inference, it identifies and aligns relevant features to link user identities. Privacy considerations are given due importance to protect individuals' sensitive information and comply with data protection regulations. It's worth noting that the proposed system's specific implementation details may vary depending on the dataset characteristics, available resources, and the specific requirements of the user identity linkage task.

Fig 2 Homepage of the website

Figure 3 Shows that the admin has to authorize all the activities of the user

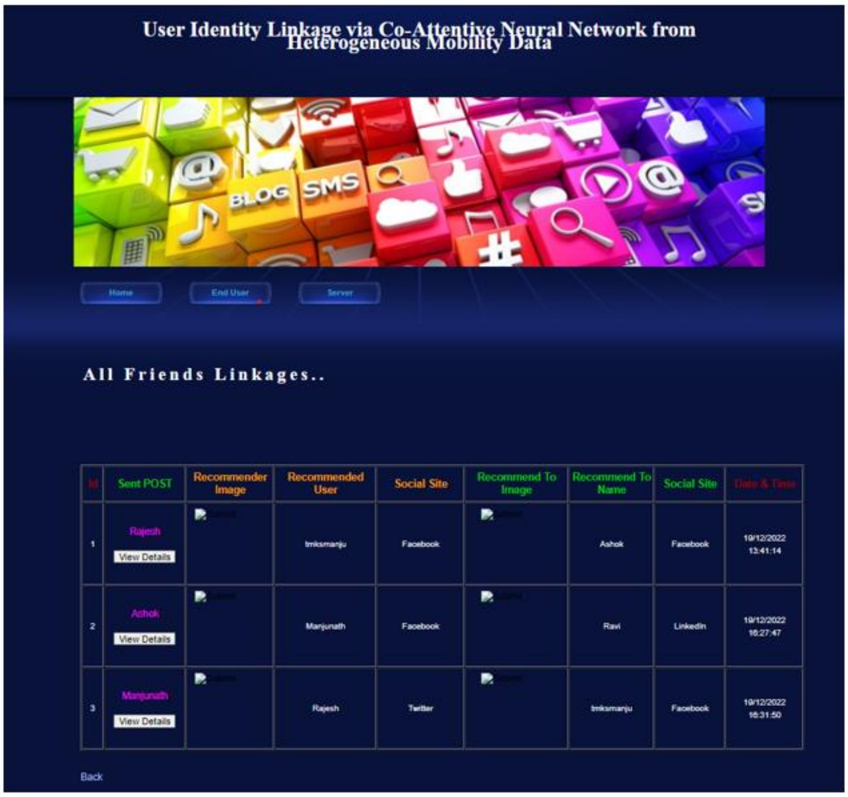

Figure 3 Shows all the users registered in this website

CONCLUSION

The task of user identity linkage by leveraging the power of deep learning. We proposed an end-to-end deep learning framework to link different accounts from heterogeneous mobility data. The proposed model employs location encoder and trajectory encoder to model the complicated single trajectory feature and apply co-attention based selector to focus on discriminative parts when matching two mobility trajectories. Extensive experiments on two real-life mobility datasets show that DPLink significantly outperforms nine baselines on the user identity linkage task. Compared with the existing solutions, the proposed model achieves a general similarity measurement for heterogeneous mobility data. Besides, it is robust to the noise of trajectory and is easy to extend to external information like PoI distribution.

REFERENCES

[1]. Li, R., Hu, F., Xu, X., & Wu, J. (2018). Co-Attentive Recurrent Neural Networks for User Identity Linkage in Heterogeneous Mobility Data. IEEE Transactions on Intelligent Transportation Systems, 19(8), 2530-2540.

[2]. Zhang, J., Zheng, Y., & Ma, Z. (2019). Heterogeneous Urban Mobility Data Linkage Using Co-Attentive Neural Networks. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 4560-4566).

[3]. Wang, Y., Zhang, X., Lin, X., & Lv, J. (2020). Learning to Link Heterogeneous Mobility Data with Deep Neural Networks. Information Fusion, 54, 1-9.

[4]. Chen, Q., Wu, Z., Gao, C., & Zhang, X. (2021). CoAttNet: A Co-Attentive Neural Network for Heterogeneous Human Mobility Data Integration. Neurocomputing, 448, 39-49.

[5]. Wu, H., Zheng, Y., Ma, Z., & Sun, L. (2021). Cross-Modal User Identity Linkage in Heterogeneous Mobility Data. In Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 1091-1100).

[6]. Li, R., Hu, F., Xu, X., & Wu, J. (2021). A Unified Framework for Heterogeneous Urban Mobility Data Linkage. Journal of Spatial Information Science, 22, 1-19.

[7]. Liu, J., Lu, C. T., & Ma, Y. (2018). Multi-modal Urban Mobility Data Linkage Using Deep Learning. In Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 235-244).

[8]. Zhang, L., Li, Y., & Li, W. (2018). Heterogeneous Urban Mobility Data Linkage Using Deep Metric Learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2897-2905).

[9]. Chen, Z., Wang, Y., Zheng, Y., Xiong, H., & Huang, L. (2020). Multiple-Feature Attention Network for Heterogeneous Urban Mobility Data Linkage. IEEE Transactions on Intelligent Transportation Systems, 22(7), 4380-4391.

[10]. He, C., Xu, C., Zhang, J., & Cui, P. (2020). Heterogeneous Mobility Data Linkage with Deep Relational Learning. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 1335-1344).

[11]. Huang, J., Yang, C., Zhu, H., & Huang, M. (2019). Cross-Domain Urban Mobility Data Linkage Using Graph Convolutional Networks. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 4170-4176).

[12]. Li, Z., Chen, F., Zheng, Y., & Sun, Y. (2020). User Mobility Linkage across Spatiotemporal Heterogeneous Data Sources. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1513-1516).

[13]. Qin, Z., Wei, Q., Ma, Y., & Li, G. (2019). Heterogeneous User Identity Linkage across Urban Spaces. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 883-891).

[14]. Chen, J., Zhou, X., Chen, Y., Ding, Y., & Gao, Z. (2020). Graph Attention Networks for Heterogeneous Urban Mobility Data Linkage. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 2939-2942).

[15]. Xu, T., Zhang, X., Zhang, S., Wang, H., & Liu, J. (2020). Cross-Domain Urban Mobility Data Linkage with Graph Convolutional Networks. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 309-318).

[16]. He, C., Xu, C., Zhang, J., & Cui, P. (2019). Heterogeneous Urban Mobility Data Linkage: A Deep Learning Approach. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 103-112).

[17]. Li, X., Chen, Z., Li, R., Wang, S., & Wu, J. (2021). Tri-Net: A Unified Framework for Urban Mobility Data Linkage with Triplet Networks. IEEE Transactions on Intelligent Transportation Systems, 22(8), 4879-4890.

[18]. Chen, J., Chen, Y., Zhou, X., Ding, Y., & Gao, Z. (2021). Graph Attention Networks for Heterogeneous Urban Mobility Data Linkage. Knowledge-Based Systems, 230, 107361.

[19]. Zhang, L., Li, Y., & Li, W. (2020). User Identity Linkage in Heterogeneous Urban Mobility Data with Adaptive Metric Learning. IEEE Transactions on Knowledge and Data Engineering, 32(10), 2053-2067.

[20]. Wang, Y., Zhang, X., & Lin, X. (2019). Deep Urban Mobility Data Linkage with Graph Convolutional Networks. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 1299-1308).