Cross-Domain Trajectory Association Based on Hierarchical Spatiotemporal Enhanced Attention Hypergraph

Chenlong Wu[1,2], Ze Wang[1,2*], Keqing Cen[1,2], Yude Bai[1,2], Jin Hao[3]

[1]School of Software, Tiangong University, Tianjin, China
[2]Tianjin Key Laboratory of Autonomous Intelligence Technology and Systems, Tiangong University, Tianjin, China
[3]Boya Triz (Tianjin) Technology Co., Ltd., Tianjin, China
{wuchenlong, wangze, cenkeqing, baiyude}@tiangong.edu.cn, hao.jin@live.cn

Abstract

Identifying and linking the same users across different social platforms is crucial for understanding user behavior and preferences. However, cross-domain datasets exhibit diverse characteristics, such as varying check-in frequencies, significant disparities in data precision, and distinct distributions. Existing trajectory representations rely on recurrent neural network, which fails to dynamically learn multi-dimensional feature relations and capture high-order associations. Furthermore, current methods for integrating trajectory information fails to capture the complex relations and dynamic variations among cross-domain mobility trajectories. To this end, we propose the Hierarchical **S**patio-**T**emporal Enhanced **A**ttention Hype**r**graph **Net**work (StarNet). This model dynamically regulates the multi-dimensional features of trajectories through a locally enhanced spatiotemporal graph neural network. Meanwhile, StarNet employs a hypergraph network enhanced by a global spatiotemporal to capture high-order associations between cross-domain trajectories. The fusion enhancement association integrates local and global information, which enables this model to link user identities. Extensive experiments on two well-known LBSN cross-domain datasets reveal that StarNet outperforms state-of-the-art baselines in the accuracy of user identity linkage.

Introduction

Location-Based Social Networks (LBSNs) encompass comprehensive data on user interactions, spatiotemporal tags, and contextual semantics (Cen et al. 2024). Analyzing this heterogeneous data is pivotal for understanding user preferences. Through cross-domain user identity linking, mobile app providers can fully analyze user behavior to achieve precise personalized recommendations, such as suggesting services and products of interest to users. However, differences in user behavior and characteristics across various platforms preclude a comprehensive understanding of user individuality and interests through a separate social medium. Therefore, cross-dataset user identity linkage is essential for gaining a deeper understanding of user behavior while enabling personalized services.

The key step in cross-domain user identity linkage is learning representations of mobility trajectories. Effective trajectory representation should describe users' spatiotemporal movement patterns and capture high-order correlations between cross-domain trajectories. Trajectory representation learning has long been a prominent topic in spatiotemporal data mining and natural language processing (Wang et al. 2024) (Bai et al. 2020). This study summaries several trajectory representation models, noting that while these models improve user trajectory representation by capturing spatiotemporal relations, they exhibit three limitations.

① Current methods rely on frameworks such as Recurrent Neural Networks (Medsker, Jain et al. 2001) and Long Short-Term Memory networks (Graves and Graves 2012) to define trajectory sequences, such as TULVAE (Zhou et al. 2018), TULER (Gao et al. 2017), DeepTUL (Miao et al. 2020), DPLink (Feng et al. 2019), and TULAM (Li et al. 2024). Although RNNs and LSTMs are advantageous in handling sequential data, they are inefficient in capturing long-term dependencies, especially with lengthy or irregularly spaced trajectories. Additionally, they struggle to gain higher-order implicit relations in dynamic trajectories.

② Existing user trajectory analysis often employs linear layers to merge local and global information, as seen in methods such as AttnTUL (Chen et al. 2024), MVMN (Zhang, Lai, and Wang 2020), CAT-ART (Li et al. 2023), and MSTHN (Lai et al. 2023). However, these linear layers fail to fully capture the complex and dynamic interactions inherent in trajectories and cannot combine the macro- and micro-semantics of user behavior. The mechanisms for information fusion need to be further explored.

③ LBSN datasets contain noisy records, leading to spatial and temporal inconsistencies across different services (Chen et al. 2018). The behavioral patterns and features of the same user's trajectories differ across service providers, influenced by factors such as sampling rates and time intervals. Such variations affects the accuracy of user identity linkage.

To address these challenges, we propose a Hierarchical Spatio-Temporal Enhanced Attention Hypergraph Network (StarNet) for enhancing accuracy in user identity linkage. StarNet comprises a local spatiotemporal enhanced graph neural network module, a global spatiotemporal enhanced hypergraph network module, and a fusion-enhanced association module. The local module learns features within trajectories and multi-dimensional data, enhancing the model's understanding of local trajectory char-

*Corresponding author.

acteristics. The global module captures higher-order features between cross-domain trajectories, improving the model's handling of cross-domain trajectories. The fusion module integrates local and global cross-domain trajectory information for association and prediction. The experimental results on two LBSN cross-domain datasets demonstrate that the proposed StarNet outperforms baselines (DPlink and AttnTUL), achieving an average improvement of 9.98% in ACC@1 and 12.45% in Macro-F1.

In summary, this work makes the following contributions:

- We pioneer the use of spatio-temporal enhanced hypergraph networks in cross-domain trajectory association, incorporating cross-attention mechanisms and contrastive normalization to effectively capture higher-order implicit relations between cross-domain trajectories.
- To dynamically learn multi-dimensional trajectory features, we integrate dynamically composable multi-head attention with graph convolutional networks (GCNs). We employ the Kolmogorov-Arnold Networks (KANs) to dynamically fuse both local and global trajectory information, which encourages a comprehensive understanding of cross-domain mobility trajectories.
- Extensive experiments on two real-world cross-domain LBSN datasets were conducted to compare the performance of StarNet against five baselines. The results show that StarNet achieved a 7.16%$\sim$12.8% improvement in ACC@1 and a 7.68%$\sim$17.22% increase in Macro-F1, underscoring the superior performance of StarNet.

Related Work

Trajectory Association Methods Based on Temporal Networks In recent years, as research has progressed, many studies have begun to utilize deep learning techniques for trajectory association. These methods have not only improved the accuracy of trajectory matching and user identification but have also better captured the semantic and spatiotemporal features within trajectory data. For instance, TULVAE (Zhou et al. 2018) captures complex long-term dependencies and trajectory semantics using variational autoencoders and hierarchical RNNs. DPLink (Feng et al. 2020) represents paired trajectories through multi-layer joint embeddings and then performs matching analysis. CPlink (Ma et al. 2022) extracts spatiotemporal periodic behaviors in specific geographic regions to calculate similarity pairs of cross-domain users, thus achieving cross-platform user identity matching.

Although these temporal networks methods have made significant progress in trajectory representation and association, challenges remain in capturing high-order relations between trajectories and dynamically learning intra-trajectory features. Efficient association of multi-source datasets and cross-domain trajectory matching still pose challenges, necessitating further exploration and solutions.

Trajectory Association Methods Based on Graph Neural Networks Given the multifaceted nature of data in Location-Based Social Networks (LBSNs), Graph Neural Network (GNN) methods (Jin et al. 2023) (Bai et al. 2023) are inherently adept at capturing complex and diverse relations, making them suitable for exploring user preferences, including user relations, temporal factors, and geographical factors. GNNTUL (Zhou et al. 2021) learns human mobility through GNNs and classifiers, effectively associating trajectories with users. S2TUL (Deng et al. 2023) captures more complex movement relations by constructing different homogeneous and heterogeneous graphs and integrating them into a unified semi-supervised learning framework. ANES (Byun, Choi, and Kim 2023) utilizes user trajectory data and bipartite graphs to learn aspect-oriented relations between users and Points of Interest (POIs), thereby associating users. AttnTUL (Chen et al. 2024) combines GNNs with a hierarchical attention mechanism to capture local and global spatiotemporal relations in trajectory data. EgoMUIL (Huang et al. 2023) captures fused user spatiotemporal features and cross-platform relations by aggregating surrounding node information and combining topological similarity, stay location similarity, and co-occurrence frequency.

However, most of the aforementioned methods focus on specific types of relations or single-dimensional analyses and are designed only for homogeneous datasets. In this study, considering these limitations, we construct corresponding hypergraph networks modules for trajectory information at different granularities. Additionally, we design multiple attention mechanisms tailored to different feature extractions, ultimately achieving the fusion of different levels of trajectory information through the KANs mechanism.

Preliminaries

In this section, we first introduce some preliminary concepts and then formally define the problem of cross-domain trajectory user association. Let $\mathcal{U} = \{u_1, u_2, \ldots, u_i\}$ represent the set of users that exist across different datasets.

- **Spatio-Temporal Check-in Point**: A check-in p is a tuple consisting of a location $l \in L$, a check-in category $c \in \mathbb{C}$, and a timestamp $t(0 <= t <= 24)$, where location l represents the geographical coordinates (i.e., longitude and latitude). Formally, $p = (l, c, t)$, representing a check-in at location l of category c at time t.

- **Mobility Trajectory**: A Mobility trajectory is a spatio-temporal sequence of check-ins generated by user u_i in chronological order. Formally, $Tr_{u_i} = \{p_1, p_2, \ldots\ldots, p_m\}$.

- **Cross-Domain Trajectory User Association Task**: Given a user set $\mathcal{U} = \{u_1, u_2, ..., u_i\}$ and a set of data sources from different domains $\mathcal{D} = \{D_1, D_2, ..., D_n\}$, for each data source D_k, with an anonymous trajectory set $Tr^{(k)}$ and known user trajectory set $Tr_u^{(k)}$, the objective of the cross-domain trajectory user association task is to learn a mapping function f:$Tr^{(1)} \cup Tr^{(2)} \cup \cdots \cup Tr^{(n)} \rightarrow U$ such that anonymous trajectories from different data sources can be linked to the corresponding user set $\mathcal{U}$.

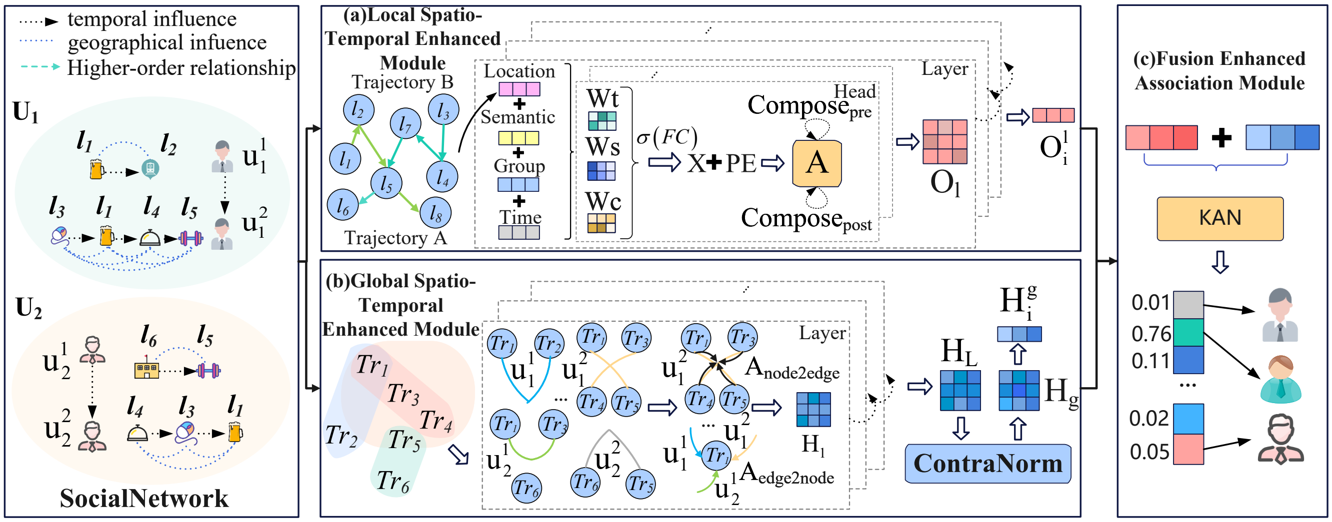

Figure 1: The overview of StarNet. (a) The process of capturing correlations between spatial positions in trajectories and dynamically learning multi-dimensional feature relations. (b) Learning complex connections between trajectories and between users and trajectories, while capturing high-order features of cross-domain trajectories. (c) Utilizing an associator to achieve precise user identity linkage.

Methodology

In this section, we introduce the details of our model, StarNet (as shown in Figure 1). It comprises three key components: (1) Local Spatio-Temporal Enhanced Graph Neural Network Module, (2) Global Spatio-Temporal Enhanced Hypergraph Network Module, and (3) Fusion Enhanced Association Module.

Local Spatio-Temporal Enhanced Graph Neural Network Module

Spatio-Temporal Convolutional Network and Multi-Dimensional Feature Encoding To capture the correlations between all spatial positions in trajectories, we construct a local spatio-temporal graph $\mathcal{G}_l = (\mathcal{V}_l, \mathcal{E}_l)$, where each grid cell is a node in $\mathcal{G}_l$. If two consecutive trajectory points map to different grid cells, an edge is created between these two grid cells (nodes). We employ a graph convolutional network on this local spatio-temporal graph to learn the local embedding representation of trajectories.

To jointly capture the topological structure and spatial location information between the grid cells, we first perform convolution operations on the local spatio-temporal graph $\mathcal{G}_l$. The multi-layer spatio-temporal convolutional network executes the following hierarchical propagation rules:

$$\mathbf{H}_l^{(i+1)} = \mathrm{ReLU}\left(\widetilde{\mathbf{D}}_l^{-\frac{1}{2}}\widetilde{\mathbf{A}}_l\widetilde{\mathbf{D}}_l^{-\frac{1}{2}}\mathbf{H}_l^{(i)}\mathbf{W}_l^{(i)}\right) \tag{1}$$

where $\mathbf{w}_l^{(i)}$ is a trainable weight matrix, $\mathbf{A}_l$ is the adjacency matrix, $\tilde{\mathbf{A}}_l = A_l + I$, and $\widetilde{\mathbf{D}}_{l_{ii}} = \sum_j \widetilde{\mathbf{A}}_{l_{ij}} \cdot \mathbf{H}_l^{(0)} = \mathbf{X}_l$. $\mathbf{X}_l$ represents the feature matrix of $\mathcal{G}_l$. $\mathbf{H}_l^{(i)} \in \mathbb{R}^{n\times d}$ is the output of the i-th layer.

When deeply analyzing human mobility data, relying solely on spatial information is often insufficient to capture the complexity of user behavior. Therefore, we integrate temporal features, semantic features, and group features of trajectory data into the trajectory feature enhancement modeling process.

We design three sparse linear embedding layers to encode the time window t_i, semantic information s_i, and group label c_i, forming a multi-dimensional feature representation vector x_i:

$$x_i = \sigma\Big(FC([W_t t_i + b_t; W_s s_i + b_s; W_c c_i + b_c; H_l(g_i)])\Big) \tag{2}$$

where σ is the activation function, $W_t, W_s, W_c, b_t, b_s, b_c$ are learnable parameters in the embedding layers, [;;] denotes the concatenation function, and $FC(\cdot)$ denotes the fully connected layer. $H_l(g_i)$ means the local embedding of grid g_i.

Dynamic Composable Multi-Head Attention Encoder To better integrate multi-dimensional information features such as spatiotemporal features, semantic features, and group features, and to fully capture the long-term complex dependencies within sequences, we propose a dynamic composable multi-head attention mechanism. By introducing learnable compositional functions and dynamically adjusting attention weights, it enhances the model's responsiveness and adaptability to different trajectory features.

To reinforce the time-awareness capability of sequence data, positional encoding is first added at the input stage:

$$Z_i = X_i + PE \leftrightarrow \tag{3}$$

$$\mathrm{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d}}\right) \tag{4}$$

$$\mathrm{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d}}\right) \tag{5}$$

In this context, $X_i = \{x_1, x_2, x_3 ..., x_m\}$ denotes the positional embeddings of all grids within the trajectory Tr_{u_i}, with positional encoding applied to distinguish the positions in the sequence, where pos represents the respective position indices. The dynamic composable multi-head attention mechanism dynamically adjusts its internal representations based on the specific characteristics of the trajectory, thereby optimizing the information fusion process. The detailed implementation of this mechanism is as follows:

First, the input tensor Z_i is multiplied by the projection matrices for the queries W^Q and keys W^K, generating the initial attention score matrix A_i^s:

$$A_i^S = \frac{(Z_i W^Q)(Z_i W^K)^T}{\sqrt{D_h}} \tag{6}$$

This step captures the initial correlations between elements and normalizes them by the square root of the dimension D_h of each attention head, stabilizing the training process and preventing gradient vanishing. Next, a learnable pre-composition function $Compose$ is used to dynamically adjust the attention distribution based on the input trajectory features:

$$A^S = \text{Compose}\big(A^S, Q, K; \theta_{\text{pre}}\big) \tag{7}$$

Here, θ_{pre} are trainable parameters that adjust the preliminary attention scores before applying $Softmax$, ensuring that the attention mechanism can accurately identify and respond to key information within the trajectory data. The processed scores A^S are then normalized using the $Softmax$ function to generate standardized attention weights A^W:

$$A^W = \text{Compose}\big(\text{Softmax}(A^S, dim = -1), Q, K; \theta_{\text{post}}\big) \tag{8}$$

The parameters θ_{post} further fine-tune and optimize the model's responses, allowing it to refine the weight distribution based on the already adjusted attention scores, thus optimizing the understanding of different temporal points, semantic layers, and group features within the trajectory data. Finally, the output of each attention head O_i is computed:

$$O_i = A_i^W (V W^V) \tag{9}$$

$$O_l = \text{Concat}(O_i^1, O_i^2 ..., O_i^h) W^O \tag{10}$$

O_l represents the result after integrating information from all attention heads. By leveraging its unique learnable and adaptive properties, the dynamic composable multi-head attention mechanism optimizes the processing of multi-source trajectory data, enhancing the model's flexibility and efficiency in capturing complex data features.

Global Spatio-Temporal Enhanced Hypergraph Network Module

Construction of the Global Spatio-Temporal Enhanced Hypergraph To more effectively model and uncover the complex connections between trajectories and the relations between users and trajectories, this study constructs a global spatio-temporal hypergraph $\mathcal{G}_H = (\mathcal{V}_H, \mathcal{E}_H)$. In this spatio-temporal hypergraph, we define two types of nodes: trajectory nodes Tr_{u_i} and user nodes u_j We connect trajectory nodes and user nodes through hyperedges to represent their complex relations.

With the definition and construction of the aforementioned structure, the spatio-temporal hypergraph neural network can effectively process and learn these complex node relations. The hypergraph convolutional layer aggregates and propagates information through the following formula:

$$H^{(l+1)} = \sigma\big(D_v^{-1/2} H D_e^{-1} H^T D_v^{-1/2} H^{(l)} W^{(l)}\big) \tag{11}$$

where H is the incidence matrix, D_v and D_e are the degree matrices of the nodes and hyperedges, respectively, and $W(l)$ is the weight matrix of the layer. σ is an activation function, such as the $ReLU$ activation function. This spatio-temporal hypergraph neural network model can not only capture the spatial relations between trajectories in depth but also reveal the deep-level connections between users and their trajectories.

Cross-Attention Mechanism To deeply learn and capture the features of cross-domain trajectories, we incorporate the cross-attention within the spatio-temporal hypergraph neural network. This mechanism aims to extract high-order relations and dynamic interactions from trajectory data through a refined multi-layer structure. Additionally, to address the issues of node weight averaging and information dilution, we implement a cross-domain trajectory weight distribution mechanism to ensure more rational attention allocation.

In each layer of the hypergraph, we configure two specialized aggregation functions responsible for analyzing and integrating information among nodes from different perspectives to capture and express complex spatial and temporal dependencies.

$$f_j^l = A_{n2e}^l(\{h_k^{l-1} \mid \forall n_k \in e_j\}) = \sigma\left(\sum_{n_k \in e_j} \alpha_{jk} W_1 h_k^{l-1}\right) \tag{12}$$

$$h_i^l = A_{e2n}^l\big(h_i^{l-1}, \{f_j^l | \forall e_j \in \mathcal{E}_i\}\big) = \sigma\left(\sum_{e_j \in \mathcal{E}_i} \beta_{ij} W_2 f_j^l\right) \tag{13}$$

where ε_i represents the set of hyperedges connected to node v_i, and f_j is the hyperedge representation. l denotes the layer number. $A_{edge2node}$ (A_{e2n}) and $A_{node2edge}$ (A_{n2e}) are two aggregation functions representing updates from hyperedges to nodes and from nodes to hyperedges, respectively. α and β are the corresponding attention weights, while W_1 and W_2 are trainable weight matrices. The hypergraph network with cross-attention can capture high-order relations and extract key information from the association between trajectories and users.

The key focus of the spatio-temporal hypergraph is to enhance the representation of each hyperedge and hypernode. However, when a hyperedge connects multiple nodes, a challenge arises as it can lead to the averaging of these nodes' weights and dilution of critical information. To address this, we employ a cross-domain trajectory weight distribution mechanism, ensuring more rational attention allocation and improving the quality of analysis. The attention matrix α for nodes to hyperedges is computed as follows:

$$l_k = \frac{l - d_k}{l \cdot 2^{d_k}} \tag{14}$$

$$u_k = \mathrm{MLP}\left(h_k^{l-1}\right) \tag{15}$$

$$\alpha_{jk} = \mathrm{Softmax}\left(l_k \frac{\exp(u_k)}{\sum_{n_a \in e_j} \exp(u_a)}\right) \tag{16}$$

where d_k represents the trajectory association degree of node k with other nodes, including shared geographic locations, similar time stamps, or user behavior patterns. u_k is a weight factor adjusted based on trajectory association degree to strengthen the connection with key nodes or paths. Similarly, the attention matrix β for hyperedges to nodes is calculated as follows:

$$\mathrm{e_j} \ = \ \mathrm{MLP}(\mathrm{f_j^l}, \ \mathrm{h_i^{l-1}}) \tag{17}$$

$$\beta_{ij} = \mathrm{Softmax}\left(l_j \frac{\exp(e_j)}{\sum_{e_a \in e_j} \exp(e_a)}\right) \tag{18}$$

The final layer output is $\mathbf{H}_L = \{\mathbf{h}_1^L, ..., \mathbf{h}_n^L\}$. Through the above mechanisms, we ensure that in the complex structure of multiple nodes and hyperedges, the attention weights of each node and hyperedge are reasonably allocated, thus avoiding information dilution and enhancing the model's expressive capability and analytical quality.

Contrastive Normalization Layer Inspired by the principles of contrastive learning, we enhance feature uniformity in the embedding space through a contrastive normalization layer, mitigating the phenomenon of dimensional collapse. The primary formula is as follows:

$$H_g = H_L - \frac{s}{\tau} \cdot \mathrm{softmax}(H_L H_L^T) H_L \tag{19}$$

where H_L and H_g represent the representations before and after the update, respectively, and s is the gradient descent step size. By updating after a certain representation layer, we can reduce the uniformity loss, thereby helping to alleviate dimensional collapse. The application of the contrastive normalization layer allows us to improve the performance and generalization ability of deep models without significantly increasing the number of parameters.

Fusion Enhancement Correlation Module

To overcome the challenges of high computational complexity, insufficient generalization ability, and suboptimal handling of non-linear data during information fusion, we employ a fusion enhancement method based on the Kolmogorov-Arnold Networks (Liu et al. 2024).

We define the shape of a KAN as $[n_1, \ldots, n_{L+1}]$, where L denotes the number of KAN layers. Notably, the Kolmogorov-Arnold theorem is defined by a KAN of shape $[n, 2n+1, 1]$. A more general, deeper KAN can be represented by composing L layers:

$$\mathbf{y} = \mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_L \circ \mathbf{\Phi}_{L-1} \circ ... \circ \mathbf{\Phi}_1)\mathbf{x}. \tag{20}$$

To fully utilize the fused cross-domain trajectory information, we design an association mechanism whose core objective is precise user identity linkage. Specifically, we first perform high-level feature extraction on the local representation O_i^l and global representation H_i^g of each trajectory to capture the complex relations of user behavior across different spatial and temporal scales. To effectively integrate these features, we employ the KAN layer, which can adaptively fuse local and global information, ensuring the comprehensiveness and consistency of cross-domain trajectory representation.

Under the processing of the KAN layer, the advanced representation of trajectories is mapped to a vector space of dimension $|\mathcal{U}|$, thereby generating feature vectors that reflect user identity associations. The specific calculation process is as follows:

$$\mathbf{y}_i = \mathrm{W}_c \left(\mathrm{KAN}\left(O_i^l, \mathrm{H}_i^g\right)\right) + b_c \tag{21}$$

where $\mathrm{W}_c \in \mathbb{R}^{|\mathcal{U}| \times 2d}$ and $b_c \in \mathbb{R}^{|u|}$ are learnable weight matrices and biases, respectively, and the vector $\mathbf{y}_i$ is the estimated probability of associating trajectory Tr_i with a user.

To train our model, we apply cross-entropy as the loss function and optimize our model using the backpropagation algorithm. We define the cross-entropy-based loss function as follows:

$$\mathcal{L}(\Theta) = -\frac{1}{\zeta} \sum_{i=1}^{\zeta} c_i \log\left(\sigma(\mathrm{y}_i)\right) + \frac{\lambda}{2} \parallel \Theta \parallel^2, \tag{22}$$

where c_i is the true label of the cross-domain trajectory Tr_i, σ is the softmax function, ζ is the number of training trajectories, Θ is the set of all trainable parameters, and λ is an L2 regularization hyperparameter to mitigate overfitting.

Experiments

We evaluated the performance of StarNet in cross-domain trajectory association tasks, which involve associating cross-domain users based on their historical trajectories. Our aim was to address the following four primary research questions:

- RQ1: How does StarNet perform in cross-domain trajectory association tasks compared to existing methods?
- RQ2: What is the relative importance of semantic features, temporal features, and group features in trajectory feature representation?
- RQ3: Is the attention mechanism design of StarNet effective in capturing intra-trajectory dependencies and the significance of cross-domain trajectory features?
- RQ4: How does the design of the contrastive normalization layer and the KAN layer influence the efficacy of StarNet?

Settings

Datasets We utilize real-world cross-domain LBSN datasets, Foursquare-Twitter(Zhang, Kong, and Yu 2014) and Instagram-Twitter(Riederer et al. 2016), to validate our proposed model. The statistical information of these datasets is provided in Table 1.

We use the baseline source code published by the authors, employing the parameter settings recommended in the original papers and fine-tuning for each dataset to achieve optimal performance. In our experiments, we set the embedding

Dataset	User numbers	Location numbers	Check-ins
Instagram	2505	59639	428292
Twitter	1721	46312	447972
Foursquare	2970	16689	44915
Twitter	1228	10472	53337

Table 1: Basic information of the datasets

dimension d=128, the number of multi-head attention heads to 8, and the regularization parameter to 5e-4. For fair comparison, we set the number of epochs to 80, batch size to 16, and dropout to 0.5 for all learning methods. The learning rate was adjusted from 0.0001 to 0.01, using an early stopping mechanism with patience set to 10 to avoid overfitting. Each experiment was repeated 10 times, and we report the average results. All experiments for model efficiency evaluation are conducted on a machine with Intel Xeon(R) Gold 6348@2.60GHz 24 core CPU, 100GB memory, and NVIDIA Tesla A800 (80GB) GPU.

Baselines We compare StarNet with seven baselines algorithms, including (1)**TULER-GRU** (Gao et al. 2017), (2)**TULER-LSTM** (Gao et al. 2017), (3)**TULER-BiLSTM** (Gao et al. 2017), (4)**T3S** (Yang et al. 2021), (5)**DPLink** (Feng et al. 2020), (6)**TULAM** (Li et al. 2024), (7)**AttnTUL** (Chen et al. 2024).

Overall Performance Comparison (RQ1)

We first evaluated the overall performance of StarNet on two datasets and compared it with state-of-the-art baselines. The overall performance of each model is reported in Table 2, where the best performance is highlighted in bold and the second best is underlined.

From the results in Table 2, we can see that our model achieved the best performance across all metrics on both cross-domain datasets. This is because our designed local graph neural network and global enhanced hypergraph network model effectively capture the micro and macro spatial features of spatiotemporal trajectories. Furthermore, our dynamic composable multi-head attention can dynamically learn and integrate dependencies on trajectory sequences, temporal dimensions, group features, and semantic information. Through KAN, it can adaptively fuse local and global representations. This is the primary reason our model performs best on both sparse datasets (i.e., Foursquare-Twitter) and dense datasets (i.e., Ins-Twitter).

The results in Table 2 show that on sparse datasets, our model outperforms the state-of-the-art baseline AttnTUL on metrics such as ACC@1, ACC@5, and Macro-F1. Specifically, our model improved ACC@1 by 12.80%, ACC@5 by 8.30%, and Macro-F1 by 17.22% compared to AttnTUL. These significant performance improvements demonstrate the effectiveness of our model in handling trajectory data with complex spatiotemporal dependencies.

On dense datasets, our model also exhibits excellent performance. Compared to AttnTUL, our model achieved significant improvements in ACC@1, ACC@5, and Macro-F1 metrics. Notably, our model improved ACC@1 by 7.16%, highlighting its strong capability in integrating multi-dimensional spatiotemporal information.

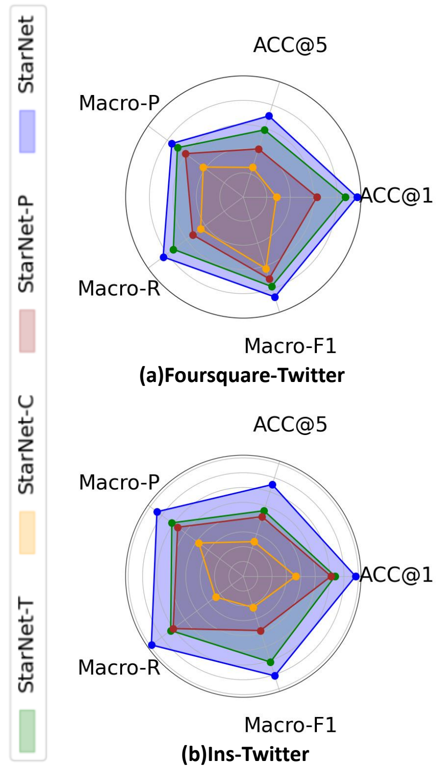

Figure 2: An ablation study on Feature Factor.

Feature Factor Analysis (RQ2)

To investigate the contributions of different sources to trajectory feature representation, we introduced three variants of StarNet: StarNet-T, StarNet-C, and StarNet-P. These variants respectively represent models where temporal, group, and semantic features are removed from the trajectory features. We compared the performance of these three variants with StarNet, and the results are shown in Figure 2. Notably, the inclusion of temporal, semantic, and group features all contribute to performance improvement. Among these, the inclusion of group features is particularly effective in enhancing the ability to identify trajectories.

Study of Attention Mechanism in StarNet (RQ3)

We conducted ablation experiments to verify the necessity of the attention mechanism. Specifically, we designed two variants: StarNet-STCAttn (S-S) and StarNet-CrossAttn (S-C), which represent the removal of the dynamic composable multi-head attention mechanism in the local graph and the cross-attention mechanism in the global graph, respectively. As shown in Figure 3, the ablation study highlights the importance of attention mechanisms in capturing intra-trajectory dependencies and cross-domain trajectory features. The comparison results of StarNet-STCAttn indicate that the dynamic multi-head composable attention in the lo-

Methods	Foursquare-Twitter					Ins-Twitter				
	ACC@1	ACC@5	Ma-P	Ma-R	Ma-F1	ACC@1	ACC@5	Ma-P	Ma-R	Ma-F1
TULER-GRU	0.4078	0.4777	0.3056	0.3316	0.2995	0.5970	0.6164	0.4837	0.4908	0.4675
TULER-LSTM	0.4082	0.4921	0.3121	0.3491	0.3088	0.5884	0.6228	0.5108	0.5150	0.4960
TULER-BiLSTM	0.4196	0.5197	0.3117	0.3470	0.3085	0.6024	0.6312	0.5112	0.5221	0.5056
T3S	0.4413	0.5441	0.3483	0.3664	0.3388	0.6315	0.6681	0.5212	0.5380	0.5139
DPLink	0.4512	0.5520	0.3518	0.3754	0.3455	0.6365	0.6721	0.5694	0.5232	0.5267
TULAM	0.4865	0.5763	0.3672	0.3929	0.3623	0.6446	0.6885	0.5717	0.5441	0.5382
AttnTUL	0.5485	0.6458	0.4538	0.4206	0.4006	0.6896	0.7614	0.6103	0.5820	0.5715
StarNet	**0.6187**	**0.6994**	**0.5127**	**0.4675**	**0.4696**	**0.7390**	**0.8026**	**0.6571**	**0.6195**	**0.6154**
Improve	**12.80%**	**8.30%**	**12.98%**	**11.15%**	**17.22%**	**7.16%**	**5.41%**	**7.67%**	**6.44%**	**7.68%**

Table 2: The overall performance of StarNet

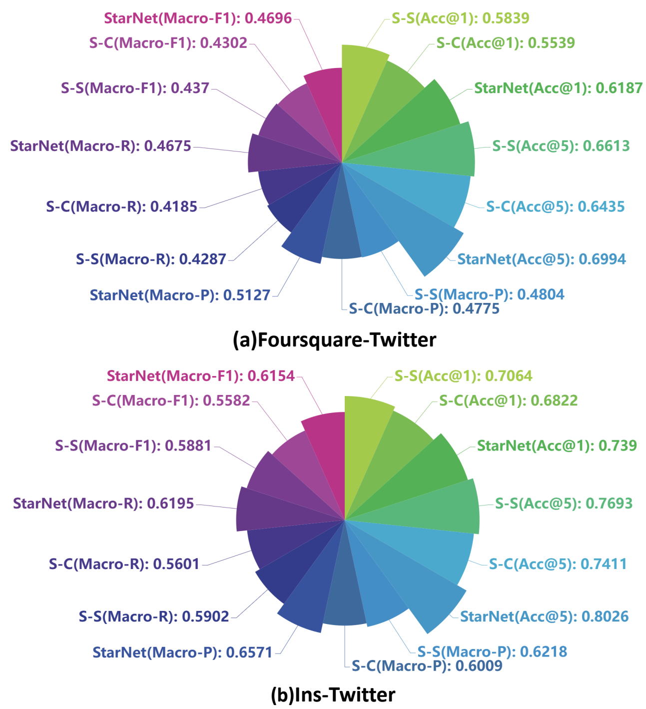

Figure 3: An ablation study on Attention Mechanism.

cal graph is highly effective for capturing complex dynamics within trajectories. Notably, StarNet-CrossAttn's ACC@1 performance is significantly lower than that of StarNet, indicating that cross-attention in the global graph is crucial for extracting key information among different trajectories.

Comparison of Normalization Layer and KAN Layer Design Study (RQ4)

We conducted ablation experiments to evaluate the impact of the normalization layer and the KAN layer on StarNet. Two variants were considered: one where the normalization layer was removed (S-C), and another where the KAN layer was replaced with an MLP layer (S-K). As shown in Figure 4, the experiments with StarNet-ContraNorm demonstrate that

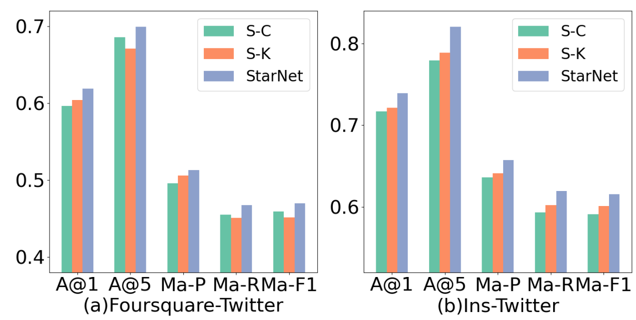

Figure 4: An ablation study on Comparison of Normalization Layer and KAN Layer.

the proposed contrastive normalization layer plays a positive role in addressing the problem of oversmoothing in spatiotemporal enhanced hypergraph networks. The experimental results of StarNet-KAN, compared to StarNet, show the advantages of using the KAN network over MLP to integrate local and global information, highlighting the effectiveness of KAN in handling complex spatio-temporal data.

Conclusion

In this paper, we propose a model StarNet for identifying and linking user trajectories across domain datasets. The local spatio-temporal enhanced graph neural network module integrates graph convolutional networks with a dynamic composable multi-head attention mechanism to adaptively adjust multi-dimensional data features. To capture higher-order relations between cross-domain trajectories and mitigate trajectory sparsity, the global spatio-temporal enhanced hypergraph network module incorporates cross-attention and contrastive normalization layers into the spatio-temporal hypergraph network. Finally, the fusion-enhanced association module employs the KANs to adaptively merge local and global trajectory representations, achieving precise user identity linkage. Extensive comparative experiments on two real-world LBSN cross-domain datasets demonstrate that StarNet significantly outperforms state-of-the-art baseline models across all evaluation metrics.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62402338), the Tianjin Technical Innovation Guidance Special Project under Grant 22YDT-PJC00140 and Key Project Foundation of Tianjin under Grant 20212053.

References

Bai, Y.; Chen, S.; Xing, Z.; and Li, X. 2023. ArgusDroid: detecting Android malware variants by mining permission-API knowledge graph. *Science China Information Sciences*, 66(9): 192101.

Bai, Y.; Xing, Z.; Li, X.; Feng, Z.; and Ma, D. 2020. Unsuccessful story about few shot malware family classification and siamese network to the rescue. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 1560–1571.

Byun, H.; Choi, Y.; and Kim, C.-K. 2023. Aspect-oriented unsupervised social link inference on user trajectory data. *Information Sciences*, 626: 249–261.

Cen, K.; Yang, Z.; Wang, Z.; and Dong, M. 2024. A cross-domain user association scheme based on graph attention networks with trajectory embedding. *Machine Learning*, 113(10): 7905–7930.

Chen, W.; Huang, C.; Yu, Y.; Jiang, Y.; and Dong, J. 2024. Trajectory-User Linking via Hierarchical Spatio-Temporal Attention Networks. *ACM Transactions on Knowledge Discovery from Data*, 18(4): 1–22.

Chen, W.; Yin, H.; Wang, W.; Zhao, L.; and Zhou, X. 2018. Effective and efficient user account linkage across location based social networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 1085–1096. IEEE.

Deng, L.; Sun, H.; Zhao, Y.; Liu, S.; and Zheng, K. 2023. S2tul: A semi-supervised framework for trajectory-user linking. In *Proceedings of the sixteenth ACM international conference on web search and data mining*, 375–383.

Feng, J.; Li, Y.; Yang, Z.; Zhang, M.; Wang, H.; Cao, H.; and Jin, D. 2020. User identity linkage via co-attentive neural network from heterogeneous mobility data. *IEEE Transactions on Knowledge and Data Engineering*, 34(2): 954–968.

Feng, J.; Zhang, M.; Wang, H.; Yang, Z.; Zhang, C.; Li, Y.; and Jin, D. 2019. Dplink: User identity linkage via deep neural network from heterogeneous mobility data. In *The world wide web conference*, 459–469.

Gao, Q.; Zhou, F.; Zhang, K.; Trajcevski, G.; Luo, X.; and Zhang, F. 2017. Identifying Human Mobility via Trajectory Embeddings. In *IJCAI*, volume 17, 1689–1695.

Graves, A.; and Graves, A. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37–45.

Huang, H.; Ding, F.; Yin, H.; Liu, G.; Wang, C.; and Wu, D. O. 2023. EgoMUIL: Enhancing Spatio-temporal User Identity Linkage in Location-Based Social Networks with Ego-Mo Hypergraph. *IEEE Transactions on Mobile Computing*.

Jin, D.; Feng, B.; Guo, S.; Wang, X.; Wei, J.; and Wang, Z. 2023. Local-global defense against unsupervised adversarial attacks on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8105–8113.

Lai, Y.; Su, Y.; Wei, L.; Chen, G.; Wang, T.; and Zha, D. 2023. Multi-view spatial-temporal enhanced hypergraph network for next poi recommendation. In *International Conference on Database Systems for Advanced Applications*, 237–252. Springer.

Li, C.; Xie, Y.; Yu, C.; Hu, B.; Li, Z.; Shu, G.; Qie, X.; and Niu, D. 2023. One for all, all for one: Learning and transferring user embeddings for cross-domain recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 366–374.

Li, H.; Cao, S.; Chen, Y.; Zhang, M.; and Feng, D. 2024. TULAM: trajectory-user linking via attention mechanism. *Science China Information Sciences*, 67(1): 112103.

Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024. KAN: Kolmogorov-Arnold Networks. arXiv:2404.19756.

Ma, X.; Ding, F.; Peng, K.; Yang, Y.; and Wang, C. 2022. CP-Link: Exploiting continuous spatio-temporal check-in patterns for user identity linkage. *IEEE Transactions on Mobile Computing*, 22(8): 4594–4606.

Medsker, L. R.; Jain, L.; et al. 2001. Recurrent neural networks. *Design and Applications*, 5(64-67): 2.

Miao, C.; Wang, J.; Yu, H.; Zhang, W.; and Qi, Y. 2020. Trajectory-user linking with attentive recurrent network. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems*, 878–886.

Riederer, C.; Kim, Y.; Chaintreau, A.; Korula, N.; and Lattanzi, S. 2016. Linking users across domains with location data: Theory and validation. In *Proceedings of the 25th international conference on world wide web*, 707–719.

Wang, C.; Yu, H.; Li, X.; Ma, F.; Wang, X.; Taleb, T.; and Leung, V. C. 2024. Dependency-Aware Microservice Deployment for Edge Computing: A Deep Reinforcement Learning Approach with Network Representation. *IEEE Transactions on Mobile Computing*.

Yang, P.; Wang, H.; Zhang, Y.; Qin, L.; Zhang, W.; and Lin, X. 2021. T3s: Effective representation learning for trajectory similarity computation. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2183–2188. IEEE.

Zhang, J.; Kong, X.; and Yu, P. S. 2014. Transferring heterogeneous links across location-based social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*, 303–312.

Zhang, W.; Lai, X.; and Wang, J. 2020. Social link inference via multiview matching network from spatiotemporal trajectories. *IEEE transactions on neural networks and learning systems*, 34(4): 1720–1731.

Zhou, F.; Chen, S.; Wu, J.; Cao, C.; and Zhang, S. 2021. Trajectory-user linking via graph neural network. In *ICC 2021-IEEE International Conference on Communications*, 1–6. IEEE.

Zhou, F.; Gao, Q.; Trajcevski, G.; Zhang, K.; Zhong, T.; and Zhang, F. 2018. Trajectory-User Linking via Variational AutoEncoder. In *IJCAI*, 3212–3218.