



User-station attention inference using smart card data: a knowledge graph assisted matrix decomposition model

Qi Zhang¹ · Zhenliang Ma¹ · Pengfei Zhang² · Erik Jenelius¹ · Xiaolei Ma³ · Yuanqiao Wen⁴

Accepted: 29 April 2023 / Published online: 18 June 2023
© The Author(s) 2023

Abstract

Understanding human mobility in urban areas is important for transportation, from planning to operations and online control. This paper proposes the concept of user-station attention, which describes the user's (or user group's) interest in or dependency on specific stations. The concept contributes to a better understanding of human mobility (e.g., travel purposes) and facilitates downstream applications, such as individual mobility prediction and location recommendation. However, intrinsic unsupervised learning characteristics and untrustworthy observation data make it challenging to estimate the real user-station attention. We introduce the user-station attention inference problem using station visit counts data in public transport and develop a matrix decomposition method capturing simultaneously user similarity and station-station relationships using knowledge graphs. Specifically, it captures the user similarity information from the user-station visit counts matrix. It extracts the stations' latent representation and hidden relations (activities) between stations to construct the mobility knowledge graph (MKG) from smart card data. We develop a neural network (NN)-based nonlinear decomposition approach to extract the MKG relations capturing the latent spatiotemporal travel dependencies. The case study uses both synthetic and real-world data to validate the proposed approach by comparing it with benchmark models. The results illustrate the significant value of the knowledge graph in contributing to the user-station attention inference. The model with MKG improves the estimation accuracy by 35% in MAE and 16% in RMSE. Also, the model is not sensitive to sparse data provided only positive observations are used.

Keywords User-station attention · Knowledge graph · Smart card data · Public transport

1 Introduction

Understanding human mobility and behavior patterns is the prerequisite for effective decision-making and interventions in urban mobility systems under normal and abnormal situations (e.g., congestions, pandemics, and natural disasters). Nowadays, with the rapid development of sensing techniques, large-volume and multi-source mobility data (e.g. smart card data in public transport, GPS trajectory data, and geotagged social media data) can be accessed and applied in human mobility pattern analysis. Mobility data may provide human travel behavior records at a disaggregated level, which significantly benefits mobility-related studies, such as individual travel patterns and behavior analysis [1–3].

In order to understand user behavior and optimizing public transport systems, the level of interest or dependence of individual users or groups of users on specific stations in a public transport system can be an important aspect. Numerous factors influence this measure, including station location, user preferences, travel patterns, and the availability of alternative transportation options. For instance, users may show a higher level of attention towards stations that are close to their home or workplace, or that offer convenient transfer options. Through the analysis of user-station attention, transport providers can gain valuable insights into user preferences and travel patterns, which can assist them in making data-driven decisions to improve services. Moreover, studying user-station attention leads to a better comprehension of human mobility, such as travel purpose and underlying motivations, and facilitates downstream applications in transport, such as mobility prediction and location recommendation.

Although many studies have explored visiting patterns of passengers to stations, we have found no studies on

✉ Zhenliang Ma
zhema@kth.se

Extended author information available on the last page of the article

user-station attention in the literature. We introduce the user-station attention problem as the problem of inferring the ‘real’ hidden (unobserved) user-station attention from the observed station visit counts data (e.g., from smart card data). The problem is challenging given its intrinsic unsupervised learning nature. In addition, it is elusive to associate the visit counts with attention for cases with zero or relatively small (e.g., 1–2 visits per week) visit counts. For example, a user having no historical visiting record to a station does not necessarily mean that the user has no interest in that station. It could be because the user is unaware of or has not yet visited that station, but could be potentially interested to go (e.g., based on nearby points of interest or restaurants). Furthermore, missing station visit count data for new users will lead to the ‘cold start’ problem when analyzing their behaviors for proactive interventions.

Mathematically, the studied user-station attention problem is similar to the user-item rating problem in recommendation systems (e.g., items, restaurants, movies, and music). Generally, the proposed models in user-item rating studies aim to capture the user similarity (e.g., sociodemographic characteristics) and item-item relationships (e.g., laptop and mouse) to make effective recommendations [4–6]. Analogous to that, an effective user-station attention inference method should model the user similarity and station-station relationships. In our context, user similarity is measured by mobility patterns extracted from users’ travel histories (approximating users’ sociodemographic). In the studied problem, we consider the user-station attention as the attention of a user to a destination station assuming the user’s origin station is known or inferred from the residential location, and we also assume the time-dependent activities are a prior. Therefore, station-station relationships should capture the travel purposes/activities in order to facilitate the user-(destination) station attention inference. In other words, we assume that users with similar travel patterns and traveling from the same station at the same time are more likely to visit the same destination station (detailed discussions in Section 3.1).

The knowledge graph (KG) is a potential approach to model relations between stations. KG is a widely studied method in recent years, which is a graph-based knowledge representation and organization method. KG can be regarded as a large knowledge base with a directed network structure (nodes representing entities and edges are relationships between entities) [7]. The main advantage of KG is that it can capture and present not only the directly observed information, but also the intricate relations among knowledge, and connect the fragmented pieces of knowledge in various information systems. The KG has been successfully applied in a wide range of domains such as searching, journalism, healthy, entertainment, network security, and pharmaceuticals [8–11]. In the urban mobility area, the temporal aspects of mobility events (e.g., departures, stations) and their depen-

dencies are essential elements in the modeling analysis. Despite several studies attempting to develop KGs for transportation considering times [12–14], they are in essence static KGs since they use times as entity attributes but the relationships do not change over time.

Another challenge for constructing KG in the mobility area is that the knowledge is usually hidden (e.g., trip purpose) in the observed travel trajectories. Developing automatic methods to extract the hidden knowledge is needed but challenging in practice without ground-truth observations.

In this paper, we introduce the user-station attention inference problem using observed station visit counts data in public transport and develop a matrix decomposition method capturing both user similarity and station-station relationships modeled using KG. Specifically, we capture the user similarity information from the user-station visit counts matrix. It extracts the stations’ latent representation and hidden relations (activities) between stations to construct the mobility knowledge graph (MKG) from smart card data. To extract the hidden relations between stations, a neural network (NN)-based nonlinear decomposition approach is introduced. The main contributions are summarized as follows:

- Develop a methodology to infer the ‘real’ user-station attention from partially observed user-station visit counts data. It captures both user similarity and station-station activity relations using MKG;
- Propose MKG construction framework from smart card data. It captures the spatiotemporal travel pattern correlations between stations using a neural network-based nonlinear decomposition method;
- Conduct systematic case studies to validate the proposed method using both synthetic and real-world datasets, and comparing with benchmark models.

The remaining paper is organized as follows. Section 2 reviews the related works of user-station attention problem and knowledge graphs. Section 3 describes the research problem and framework, and proposes the user-station attention inference method. Section 4 presents the case study to validate the proposed method and explore its performance in user-station attention matrix inference. Section 5 summarizes the main findings and outlines future research directions.

2 Related work

User-station attention has not been previously studied. However, the concept is related to the problem of suggesting relevant items to users in recommender systems (or recommendation systems). In recommender systems, algorithms

aim at predicting a user's 'rating' or 'preference' for an item (e.g., movies, books, restaurants, etc.). In recent years, the KG has been widely used in recommender systems and achieved fruitful results [15–19]. Concretely, a KG-based recommender system exploits the connection between entities (nodes) representing users, items to be recommended, and their interactions (relations) [4]. The relations, explicit or implicit, are used to identify items that may be interesting to the target user [17]. Thus, compared to other approaches (e.g., collaborative filtering, and content-based filtering), the relations in KG can provide additional valuable information to facilitate inference between nodes to discover new connections. It leads to better prediction performance, especially when the dataset is sparse [19]. In general, the KG exhibits rich semantic correlations that can be used to explore hidden connections, help expand users' interests to increase the diversity of recommended items, and enhance the interpretability of the system [20]. Therefore, the KG has great advantages in mining relationships and establishing connections between entities, which lays an important foundation for user-station attention inference.

KGs as a form of structured knowledge representation and extraction method have drawn great research attention from both academia and industry. The KG can be classified into general knowledge graphs and domain knowledge graphs according to the application field [16]. The domain KGs usually require more comprehensive background knowledge and specific datasets than the general ones. Given the studied problems, this paper focuses on reviewing the domain KGs.

KGs have been extensively studied in many fields, especially in the medical field. Many studies constructed a medicine/health knowledge graph by extracting medical entities, relations, and properties from different sources of medical knowledge, respectively [21–23]. Yang et al. [24] developed a corporate risk knowledge graph using basic corporate information from encyclopedia data through information extraction, knowledge fusion, ontology construction, and dynamic knowledge reasoning. It introduced a time dimension to describe the evolutionary characteristics of enterprise risk events and focused on characters and entities in the corporate field. Du et al. [25] built a GIS knowledge graph (GIS-KG) by combining various GIS knowledge sources to form a hierarchical ontology and using deep-learning techniques to map GIS publications onto the ontology. By capturing and integrating GIS knowledge from various primary sources, it improved domain-specific information retrieval for the wide-ranging field of geographic information science and its technologies. Zhang et al. [26] systematically analyzed the maritime dangerous goods related knowledge and defined concepts and relations to build an ontology framework and then filled entities in this framework to construct a knowledge graph for maritime dangerous goods. It was used for the efficient retrieval of dangerous

goods knowledge, as well as the automatic judgment of segregation requirements. Mao et al. [27] proposed a process safety knowledge graph and developed a domain ontology on the delayed coking process to handle the process safety knowledge and support risk causes and consequences analysis. It enhanced the knowledge-based analysis abilities in discovering the hidden relationships between possible risk causes and consequences in an emergency situation.

The knowledge sources of the above studies are all based on domain textual knowledge, and different methods are used to extract knowledge. In addition, most studies used the triple form structure (entity-relation-entity) with a few using the quadruple form (entity-relation-entity-property) to extract information from static text data.

Compared with other domain knowledge graphs, research in transportation is still emerging. For example, Tan et al. [14] constructed a knowledge graph of an urban traffic system and extracted entities, relations, and attributes from Automatic Fare Collection (AFC) data, static basic data of metro lines and stations, urban road traffic data, and other data to construct a traffic field ontology. It was used for traffic knowledge discovery and intelligent question answering of urban traffic services (traveler discovery). Shan et al. [28] proposed the notion of an urban knowledge graph, which composes multiple traffic topics, entities, attributes, and features extracted from traffic text knowledge. It performs better than traditional statistical methods for reasoning pedestrian volume based on a hybrid reasoning algorithm. Chen et al. [12] proposed a knowledge graph-based framework for digital contact tracing in public transportation systems during the COVID-19 pandemic. The framework fuses multi-source data and uses the trip-chaining model to construct a knowledge graph, from which a contact network is extracted and a breadth-first search algorithm is developed to efficiently trace infected passengers in the contact network. Wang et al. [29] studied the integration of reinforcement learning and spatial knowledge graph for incremental mobile user profiling. The spatial knowledge graph extracted entities from the environment and established the semantic connectivity among spatial entities and characterized the semantics of user visits over connected locations, activities, and zones. Li et al. [30] presented a knowledge graph-based framework for discovering potential destinations of low predictability travelers. The framework considers trip association relationships and constructs a trip knowledge graph to model the trip scenario. With a knowledge graph embedding model, it can obtain the possible ranking of unobserved destinations by calculating triples' distance.

Despite efforts made in numerous studies to develop KGs for transportation that incorporate time as an entity attribute, they remain static in nature as the relationships do not dynamically change over time. Furthermore, constructing a KG in the mobility area is particularly challenging in comparison

to other fields due to the fact that the knowledge, such as trip purpose, is typically hidden within the observed travel trajectories. Therefore, it is essential but challenging to develop methods for extracting this hidden knowledge in practical applications without ground-truth observations.

3 Methodology

This section introduces the user-station attention inference problem. It proposes the inference methodology assisted by MKG and also correspondingly the MKG construction models from smart card data. The notation used throughout the paper is shown in Table 1.

3.1 Problem description and research framework

In practice, the attention of a user to stations and the station-station relations (activities) in MKG could vary for different time periods. Therefore, we infer the user-station attention matrix and develop a set of sub-graphs corresponding to different time periods, such as morning peak and off-peak. For ease of presentation, we define key terminologies used in this paper.

Definition 1 (Individual mobility trajectories) The individual mobility trajectories are a sequence of trips of users, i.e., $T_r = \{tr_1, tr_2, tr_3, \dots, tr_n\}$, with a trip $tr_i = \{user\ id, origin, destination, entry\ time, exit\ time\}$.

Definition 2 (User-station visit counts matrix) The user-station visit counts matrix $\mathbf{X} = [x_{ij}] \in \mathbb{R}^{m \times n}$ describes the visiting counts of users to stations, where m is the number of users, n the number of stations. For each entry x_{ij} in matrix \mathbf{X} , it is the visiting counts of i^{th} user to j^{th} station.

Definition 3 (User-station attention matrix) The user-station attention matrix $\mathbf{Y} = [y_{ij}] \in \mathbb{R}^{m \times n}$ is defined as the whole attention of the transport network, where m is the number of users, n the number of stations. For each entry $y \in (0, 1)$ in matrix \mathbf{Y} , it is the attention of i^{th} user to j^{th} station.

Definition 4 (Mobility knowledge graph) The mobility knowledge graph is defined as a time-dependent knowledge graph, which is composed of sub-graphs for different time intervals. Each sub-graph is a quadruple form knowledge graph composed of nodes, edges, and time period, in which nodes represent stations and edges represent the relations (travel activities) between stations during a certain time period. It can be represented as $\mathbf{G} = \{(\mathbf{l}_o, \mathbf{r}, \mathbf{l}_d, t) \mid \mathbf{l}_o \in \mathbf{L}_o, \mathbf{r} \in \mathbf{R}, \mathbf{l}_d \in \mathbf{L}_d\}$, where \mathbf{l}_o is the origin stations' latent vector; \mathbf{r} is the relationship's latent vector; \mathbf{l}_d is the destination stations' latent vector; and t is the time period. At each time period t , the time element in \mathbf{G} becomes irrelevant in the sub-graph.

Note that the user-station visit counts matrix provides a simple count of the number of times a user has visited a particular station, without indicating the importance or dependence of the user on that station. In contrast, the user-station attention matrix can quantify the importance of a station to a user or group, providing insights into the level of interest or dependency of users on different stations. Also, the user-station visit counts may be biased towards frequently visited stations, potentially overlooking less frequently visited stations and resulting in an incomplete understanding of the user-station relationship. Therefore, while the user-station visit counts matrix provides basic information about the users and stations, the user-station attention matrix offers a more comprehensive, intuitive, and nuanced understanding of their relationships.

Figure 1 shows the studied problem in this paper. The problem is to infer the full and 'true' user-station attention matrix \mathbf{Y} given a partially observed user-station visit counts matrix \mathbf{X}_p . Due to mechanical faults, new users, or ticket users in the network, matrix \mathbf{X} may not be fully observed. Moreover, the zero and small values (e.g., 1-2 visits per week) in matrix \mathbf{X} are regarded as 'untrusted' observations. Therefore, only certain observations in matrix \mathbf{X} (i.e., partially observed user-station visit counts matrix \mathbf{X}_p) are used to infer the user-station attention matrix. Similar to the user-item rating prediction problem in the recommender system, we represent users and stations in a latent low-dimensional space of dimension k in Eq. (1). The i^{th} user is represented by a latent vector $\mathbf{u}_i \in \mathbb{R}^{1 \times k}$ and j^{th} station by a latent vector $\mathbf{l}_j \in \mathbb{R}^{1 \times k}$. The attention of i^{th} user to j^{th} station is expressed as the inner product of their latent representations [31] as follows:

$$\mathbf{y}_{ij} = \mathbf{u}_i \mathbf{l}_j^T. \quad (1)$$

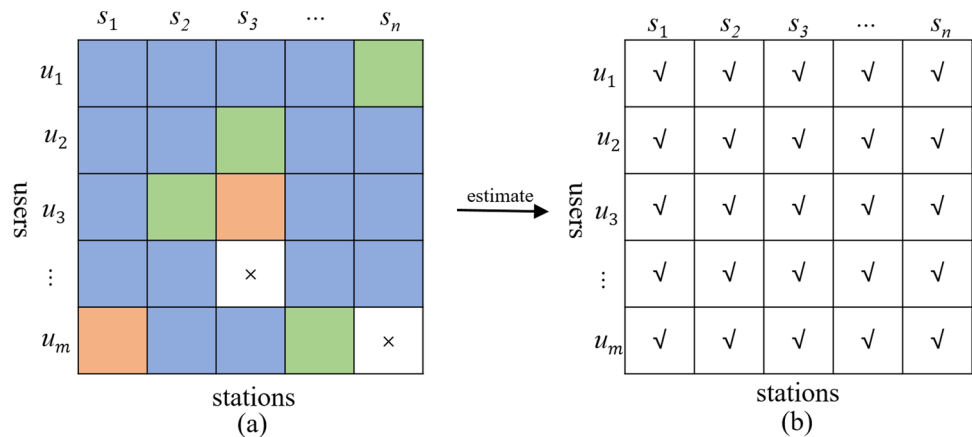
However, there are two challenges to solving this problem. Firstly, as discussed previously, the zero and extremely small values in matrix \mathbf{X} may not reflect the user's real attention to a station, which should be considered noisy data. Secondly, the lack of historical trajectory data (missed values) can lead to cold start problems. To tackle this issue, the user similarity and station-station relations (activities) are exploited to assist in inferring the fully user-station attention matrix.

A common implementation in recommender systems is to recommend new items by identifying other users with similar tastes. Similarly, the user's attention to a destination that has not been visited or not frequently visited can be inferred based on similar users. In addition, a person's travel demand is derived from their daily activity patterns. Therefore, we assume that people with the same travel activity and traveling from the same origin station at the same time period would more likely to visit the same destination. People with the same travel activity refer to those who have the same

Table 1 Notation

Variable	Interpretation
l	Vector of station, $l \in L$
l_o	Vector of origin station
l_d	Vector of destination station
l_t	Station's temporal latent vector
l_s	Station's spatial latent vector
L	Stations' latent representation
L_t	Stations' temporal latent representation
L_s	Stations' spatial latent representation
L_{ts}	Stations' spatiotemporal latent representation
r	Vector of relation, $r \in R$
R	Relation's latent representation
Tr	Individual mobility trajectories, $Tr = \{tr_1, tr_2, \dots\}$
t	Time period
m	Number of users
n	Number of stations
p	Number of relations
w	Number of hidden neurons
T	temporal station demand matrix
S	OD flow matrix
Q	Coefficient matrix
D	Feature diagonal matrix
H	Latent indexing matrix
W_1, W_2	Weights of the first layers and second layers
k, k_t, k_s, k_{ts}, k_r	Number of latent features
u	Vector of user, $u \in U$
U	User's latent representation matrix
X	Observed user-station visit counts matrix
X_p	Partially observed user-station visit counts matrix
X_o	Pre-set 'real' user-station attention matrix
X_s	Simulated user-station visit counts matrix
Y	The 'real' user-station attention matrix
M	Maximum iteration
N	Number of user-station pairs in the network
λ	Regularization rate

Fig. 1 Problem description. (a) Observed user-station visit counts matrix (Green grid: 0; Orange grid: observations with extremely small values; Blue grid: observations with normal values; \times : missed observation). (b) The user-station attention matrix



travel purpose, such as work, shopping, school, etc. Such an assumption is important for accurate and robust user-station attention estimation. By analyzing the travel patterns of users with similar travel activities and factoring in the origin station and time period, it may be possible to infer the attention of a particular user towards a destination. For instance, if a destination has not been visited by a particular user but has been frequently visited by users with similar travel activities during the same time period, it may be inferred that the user has a certain level of interest in that destination.

For missing values, the activities (relations) between OD pairs contribute to the user-station (destination) attention inference. For example, suppose the relation between stations A and B is shopping during off-peak and a new user travels from station A at off-peak, then we can infer the new user may possibly have attention to station B. In order to assist the studied problem, we thus define the station-station relations in MKG as travel purposes/activities. Given these, we formally define our studied problems as below.

Problem 1 (User-station attention matrix inference) *Given the partially observed user-station visit counts matrix \mathbf{X}_p to infer the fully user-station attention matrix \mathbf{Y} capturing both user similarity and station-station activity relations using MKG.*

Problem 2 (MKG construction) *Given the individual mobility trajectories T_r , extract the stations' latent representation \mathbf{L} and the hidden relationships (activities) \mathbf{R} capturing the spatiotemporal correlations of travel patterns during time interval t to construct the MKG $\mathbf{G}(\mathbf{L}_o, \mathbf{R}, \mathbf{L}_d, t)$ in a form of $\langle \text{origin, relationship, destination, time period} \rangle$.*

Figure 2 shows the framework of the proposed method. The input data are individual mobility trajectories T_r . It contains three modules: user similarity, MKG construction, and user-station attention matrix inference.

1. **User Similarity** This step constructs the observed user-station visit count matrix \mathbf{X} from mobility trajectory data and obtained the users' latent representation \mathbf{U} based on a matrix decomposition approach. In matrix \mathbf{X} , each row records the visiting counts of a user at different stations, which reflects the user's travel pattern. The user similarity is reflected by different rows of the matrix. By matrix factorization, the users' latent representation can be obtained, which captures the user similarity (implicit) features.
2. **MKG Construction** This step constructs the temporal station demand matrix \mathbf{T} and OD flow matrix \mathbf{S} from mobility trajectory data. An NN-based nonlinear decomposition approach is proposed to extract the stations' latent representation, and station-station relations (activities) \mathbf{R} to construct the MKG. The extracted relations capture the spatiotemporal travel pattern correlations between stations.
3. **User-Station Attention Matrix Inference** It infers the full and 'real' user-station attention matrix \mathbf{Y} capturing both user similarity and station-station relations using the constructed MKG.

3.2 User similarity

By collecting the user's check-in and check-out data from the mobility trajectories T_r during time period t , we can construct the observed user-station visit counts matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, where m is the number of users, and n the number of stations. Each entry x_{ij} in matrix \mathbf{X} represents the visiting numbers of i^{th} user to j^{th} station. Each row aggregates the user's number of visits to each station, which reflect the user's travel pattern. Therefore, we can capture the users' hidden characteristics (user similarity) by decomposing matrix \mathbf{X} . Similar

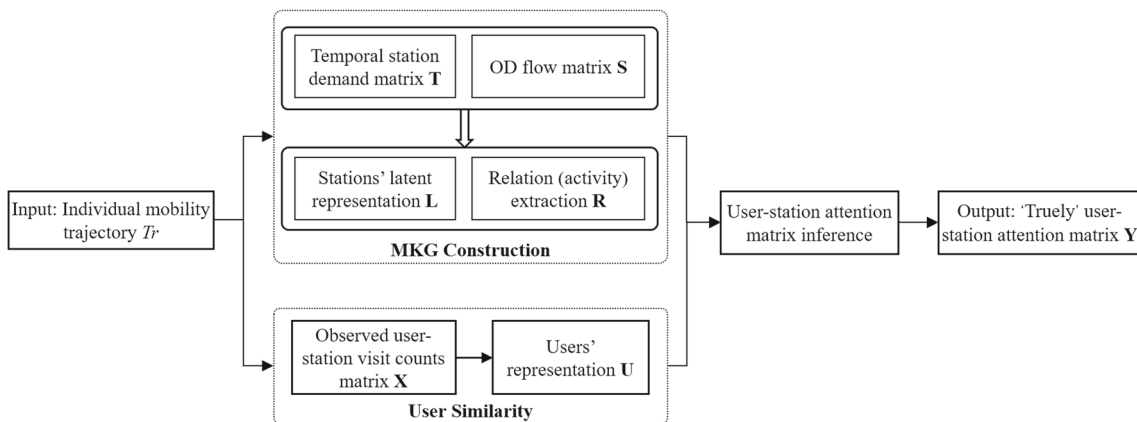


Fig. 2 Research framework

to Eq. (1), based on the matrix factorization, the matrix \mathbf{X} can be decomposed into:

$$\mathbf{X} \approx \mathbf{U}\mathbf{L}^T. \tag{2}$$

$\mathbf{U} \in \mathbb{R}^{m \times k}$ is the users' latent representation capturing the user similarity (implicit); $\mathbf{L} \in \mathbb{R}^{n \times k}$ is the stations' latent representation. Therefore, by optimizing the following objective function, we can obtain \mathbf{U} and \mathbf{L} .

$$\min_{\mathbf{U}, \mathbf{L} \geq 0} \|\mathbf{X} - \mathbf{U}\mathbf{L}^T\|^2 + \lambda [\|\mathbf{U}\|^2 + \|\mathbf{L}\|^2]. \tag{3}$$

$\|\cdot\|$ denotes the Frobenius norm and λ the regularization rate. The last term is the regularization for penalizing the norm of \mathbf{U} and \mathbf{L} . The Frobenius norm is the Euclidean norm applied to the vectorized version of a matrix. It is commonly used to regularize the objective function in optimization problems [32].

After this step, we can obtain \mathbf{U} and \mathbf{L} to infer the user-station attention matrix \mathbf{Y} . However, it has two major drawbacks. First, the learned stations' latent representation \mathbf{L} is not easy to interpret; second, it does not capture the station-station relations and other constraint information (e.g., temporal, spatial, spatiotemporal features). To overcome these limitations, the MKG is constructed to enrich stations' latent representation.

3.3 Mobility knowledge graph construction

The construction of MKG consists of two parts: entity extraction and relation extraction. Entity extraction refers to extracting different station entities (including subway stations, train stations, or bus stops) from mobility data. Unlike GPS trajectory data, the passenger's boarding and alighting stations are usually recorded in the smart card data, which makes it not difficult to identify the passengers' travel station. Since all variables in this study are calculated in the latent space, we need to obtain the stations' latent representation. Relation extraction is to extract the implicit relations (activities) between the stations. In the context of MKG in public transport, relations represent the spatiotemporal travel pattern correlations between two stations within a certain time period, which is in an implicit form ('hidden' mechanism). Therefore, relation extraction is a core part of MKG construction.

3.3.1 Stations' latent representation

For each station, we construct the station's latent vector $\mathbf{l} = [\mathbf{l}_t; \mathbf{l}_s]$ capturing both temporal and spatial features of station demand patterns, where \mathbf{l}_t and \mathbf{l}_s are the station's temporal and spatial latent vectors, respectively. The travel

characteristics of each station are time-varying. For example, the number of visitors at a station may be quite different between peak and off-peak hours. Therefore, it is important to extract the visiting feature of stations based on time. We first select the time period of interest t (e.g., morning peak or off-peak), then construct the temporal station demand matrix $\mathbf{T} \in \mathbb{R}^{n \times h}$, where h is the number of hours within time period t . An entry T_{ij} in matrix \mathbf{T} represents the visit counts of i^{th} station in the j^{th} hour. According to the Non-negative Matrix Factorization (NMF) theory [33], the non-negative matrix \mathbf{T} can be decomposed into two non-negative matrices, which is utilized to extract the visiting feature of each station. The time series matrix \mathbf{T} can be decomposed into:

$$\mathbf{T} \approx \mathbf{L}_t\mathbf{Q}. \tag{4}$$

Matrix $\mathbf{L}_t \in \mathbb{R}^{n \times k_t}$ is the stations' temporal latent representation, $\mathbf{Q} \in \mathbb{R}^{k_t \times h}$ is the coefficient matrix and k_t is the number of stations' temporal latent features.

To obtain the spatial latent features between OD-pairs, the OD flow matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ is constructed, in which an entry S_{ij} represents the visit counts from i^{th} station to j^{th} station. Given the matrix \mathbf{S} , it can be decomposed into:

$$\mathbf{S} \approx \mathbf{L}_s\mathbf{D}\mathbf{L}_s^T. \tag{5}$$

Matrix $\mathbf{L}_s \in \mathbb{R}^{n \times k_s}$ is the stations' spatial latent representation; $\mathbf{D} \in \mathbb{R}^{k_s \times k_s}$ is the feature diagonal matrix and k_s is the number of spatial latent features. The OD flow matrix \mathbf{S} records the relation from one station to another, which can also be seen as spatial dependencies. Therefore, the OD flow matrix \mathbf{S} can be regarded as the embedding of the system's topology. Similarly, the matrix \mathbf{L}_s and \mathbf{D} are the stations' spatial embedding and the interactions in the latent component space, respectively.

Combining Eqs. (4) and (5), the stations' latent representation \mathbf{L} is expressed as $\mathbf{L} = [\mathbf{L}_t; \mathbf{L}_s]$, which is formed by concatenating stations' temporal latent representation \mathbf{L}_t and spatial latent representation \mathbf{L}_s . Therefore, we can get the stations' latent representation \mathbf{L} by optimizing the following objective function

$$\min_{\mathbf{L}_t, \mathbf{L}_s, \mathbf{Q}, \mathbf{D} \geq 0} \|\mathbf{S} - \mathbf{L}_s\mathbf{D}\mathbf{L}_s^T\|^2 + \|\mathbf{T} - \mathbf{L}_t\mathbf{Q}\|^2 + \lambda [\|\mathbf{L}_t\|^2 + \|\mathbf{L}_s\|^2 + \|\mathbf{Q}\|^2 + \|\mathbf{D}\|^2]. \tag{6}$$

The first and second term capture the spatial and temporal latent patterns, respectively; the last term is the regularization for penalizing the norm of \mathbf{L}_t , \mathbf{L}_s , \mathbf{Q} , and \mathbf{D} ; λ is the regularization rate. Actually, Eqs. (4) and (5) can be seen as stations' temporal and spatial-based information extraction respectively. After this step, the stations' temporal and

spatial related latent representation $\mathbf{L} = [\mathbf{L}_t; \mathbf{L}_s]$ is extracted and can be used to enrich \mathbf{L} in Eq. (3).

3.3.2 Relation extraction

Relation extraction for KG is an extensive and profound research topic and most studies focus on supervised learning methods [34–37]. In contrast to the conventional supervised relation extraction learning task, the relation extraction problem in this paper is an unsupervised learning task since the ‘real’ relations between stations in MKG are unknown in the training data. Moreover, compared to linear matrix decomposition, the non-linear matrix decomposition method can extract more information and simultaneously capture spatiotemporal travel patterns of OD pairs (rather than arbitrarily concatenating spatial and temporal patterns). In the OD flow matrix \mathbf{S} , each entry records a specific flow from the origin to the destination, which is determined by characteristics of origin and destination stations, as well as their relations. Therefore, the relations can be extracted from the OD flow matrix \mathbf{S} . We propose a neural network (NN)-based model, based on the model used in [38], to formulate the problem to extract the relations’ and stations’ latent representation with the objective of recovering the observed OD flow matrix (i.e., a supervised learning model).

To illustrate that, Fig. 3 shows the feedforward NN framework for the MKG relation extraction learning task. The input data $[\mathbf{l}_o; \mathbf{r}_{od}; \mathbf{l}_d]$ is the triple form of MKG in a certain time period, which is horizontally concatenated by origin’s latent vector \mathbf{l}_o , relation’s latent vector \mathbf{r}_{od} and destination’s latent vector \mathbf{l}_d . The estimated OD flow is generated through the hidden layer and the initial weights $\mathbf{W}_1, \mathbf{W}_2$ of the NN. Then

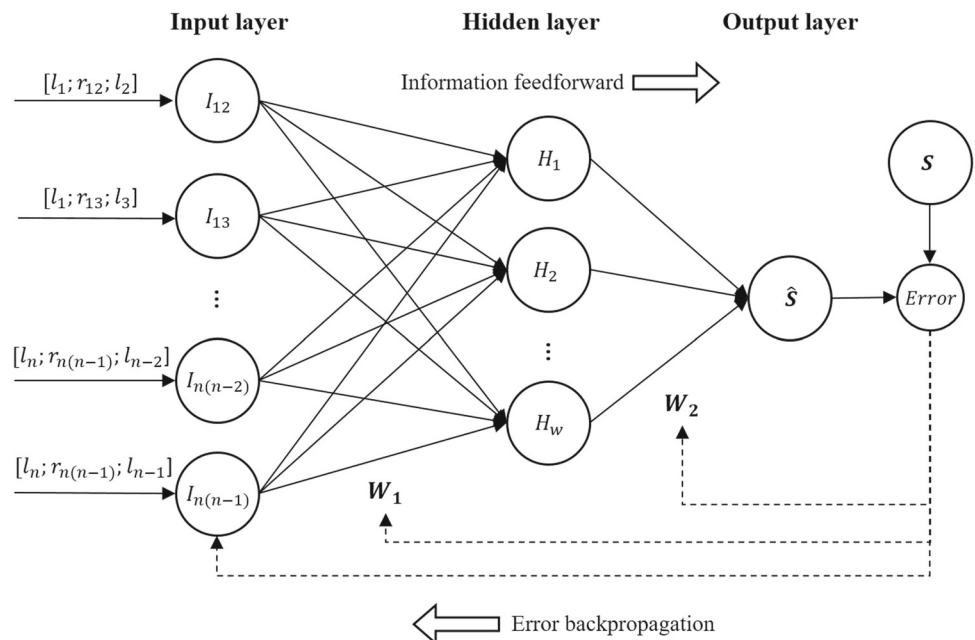
the error between the estimated and observed OD flows is calculated and used to update the weights and the input triple form via the backpropagation algorithm. The objective is to minimize the error in order to obtain the relations and stations’ latent representation to construct MKG. Adopting a supervised learning approach for this unsupervised task yields several advantages, including the network’s ability to acquire more intricate information (i.e., comprehending spatiotemporal OD travel patterns) and enhancing the inference performance.

In this method, the stations’ spatiotemporal latent representation and the relations’ latent representation are defined as $\mathbf{L}_{ts} \in \mathbb{R}^{n \times k_{ts}}$ and $\mathbf{R} \in \mathbb{R}^{p \times k_r}$, respectively, where p is the numbers of relations; k_{ts} and k_r are the number of latent features of stations and relations. To specify a relationship between two stations, a latent indexing vector $\mathbf{H} \in \mathbb{R}^{1 \times p}$ is introduced, and the latent vector of a specific relation is $\mathbf{r}_{od} = \mathbf{H}\mathbf{R}$. Therefore, the objective of the relation extraction is to map each input vector $[\mathbf{l}_o; \mathbf{r}_{od}; \mathbf{l}_d]$ to an output entry S_{od} . The relation extraction problem is defined as: Given a training set $\{([\mathbf{l}_1; \mathbf{r}_{12}; \mathbf{l}_2], S_{12}), ([\mathbf{l}_1; \mathbf{r}_{13}; \mathbf{l}_3], S_{13}), \dots, ([\mathbf{l}_n; \mathbf{r}_{n(n-2)}; \mathbf{l}_{n-2}], S_{n(n-2)}), ([\mathbf{l}_n; \mathbf{r}_{n(n-1)}; \mathbf{l}_{n-1}], S_{n(n-1)})\}$, for each input $\mathbf{z} = [\mathbf{l}_o; \mathbf{r}_{od}; \mathbf{l}_d] \in \mathbf{Z}$ and corresponding output entry $S_{od} \in \mathbf{S}$, the goal is to learn a function $f: \mathbf{Z} \rightarrow \mathbf{S}$ mapping inputs to outputs. For each input vector \mathbf{z} , the feedforward NN-based $f_{NN}(\mathbf{z})$ is defined as:

$$f_{NN}(\mathbf{z}) = \mathbf{W}_2 \text{Sigmoid}((\mathbf{z}\mathbf{W}_1)^T). \tag{7}$$

Matrix $\mathbf{W}_1 \in \mathbb{R}^{(2k_{ts}+k_r) \times w}$ and $\mathbf{W}_2 \in \mathbb{R}^{1 \times w}$ represent the weights of the first layers and second layers; w is the number of hidden neurons and $\text{Sigmoid}()$ is the activation function.

Fig. 3 Framework of feedforward neural network



Similar to Eq. (6), by considering both temporal and spatial features, i.e., Eqs. (4) and (7), the stations' latent representation \mathbf{L}_{ts} , and relation \mathbf{R} can be learned by minimizing the following objective function:

$$\min_{\mathbf{L}_{ts}, \mathbf{R}, \mathbf{H}, \mathbf{Q} \geq 0} \sum_{\forall (l_o, r_{od}, l_d) \in \mathcal{S}} (S_{od} - f_{NN}(\mathbf{z}))^2 + \|\mathbf{T} - \mathbf{L}_{ts}\mathbf{Q}\|^2 + \lambda \left[\|\mathbf{L}_{ts}\|^2 + \|\mathbf{H}\|^2 + \|\mathbf{R}\|^2 + \|\mathbf{Q}\|^2 \right]. \quad (8)$$

The stations' latent spatiotemporal representation \mathbf{L}_{ts} is shared in the first and second terms and constrained by the two terms simultaneously. The parameter λ is the regularization rate. The last term is the regularization for penalizing the norm of \mathbf{L}_{ts} , \mathbf{H} , \mathbf{Q} , and \mathbf{R} . After this step, the MKG can be represented as $\mathbf{G}(\mathbf{L}_{ts}, \mathbf{R}, \mathbf{L}_{ts}, t)$.

3.4 User-station attention matrix inference

As mentioned previously, the user-station attention inference problem should comprehensively consider both user similarity and station-station relationships, i.e., Eqs. (3) and (8). Therefore, the user-station attention matrix can be inferred by optimizing the following objective function:

$$\min_{\mathbf{U}, \mathbf{L}_{ts}, \mathbf{R}, \mathbf{H}, \mathbf{Q} \geq 0} \left\| \mathbf{X} - \mathbf{U}\mathbf{L}_{ts}^T \right\|^2 + \sum_{\forall (l_o, r_{od}, l_d) \in \mathcal{S}} (S_{od} - f_{NN}(\mathbf{z}))^2 + \|\mathbf{T} - \mathbf{L}_{ts}\mathbf{Q}\|^2 + \lambda \left[\|\mathbf{U}\|^2 + \|\mathbf{L}_{ts}\|^2 + \|\mathbf{H}\|^2 + \|\mathbf{R}\|^2 + \|\mathbf{Q}\|^2 \right]. \quad (9)$$

The first term is to obtain the users' latent representation capturing the user similarity (implicit) features. The second and third terms capture the stations' spatiotemporal latent patterns, and station-station relations capture the spatiotemporal travel pattern correlations between the two stations. The parameter λ is the regularization rate. The last term is the regularization for penalizing the norm of \mathbf{U} , \mathbf{L}_{ts} , \mathbf{H} , \mathbf{Q} , and \mathbf{R} . Therefore, the fully user-station attention matrix \mathbf{Y} is inferred by capturing both user similarity and station-station relations using MKG.

3.5 Solution algorithm

To solve the optimization problem in Eq. (9), we develop an iterative user-station attention matrix inference algorithm based on the gradient descent algorithm. Algorithm 1 presents the pseudo-code of the solution algorithm.

Algorithm 1 Generative process for user-station attention matrix inference.

Input: the temporal station demand matrix T , the OD flow matrix S , learning rate α , maximum iteration N , regularization rate λ , convergence error ε

Output: latent variables V^{ts} , R^{ts} , H , Q

```

1: Randomly initialize latent variables  $V^{ts}$ ,  $R^{ts}$ ,  $H$ ,  $Q$ 
2: Normalize matrix  $T, S, X$ 
3:  $Obj = Eq. (9)$ 
4: compute  $cost_0 = Obj$ 
5: for  $i = 1 : N$  do
6:   for each variable  $v$  in  $[U, L_{ts}, R, H, Q]$  do
7:     compute gradient  $\nabla_v = \frac{\partial(Obj)}{\partial v}$ 
8:     update variable  $v_{i+1} = v_i - \alpha \cdot \nabla_v$ 
9:     update  $cost_i = Obj$ 
10:   end for
11:   if  $|\frac{cost_i - cost_{i-1}}{cost_{i-1}}| \leq \varepsilon$  then
12:     break
13:   end if
14: end for
15: return latent variables  $U, L_{ts}$ 

```

4 Case study

4.1 Data preparation

To validate the proposed method, both synthetic data and real-world smart card data from Hong Kong Mass Transit Railway (MTR) are used. MTR is the major public transport network of Hong Kong, serving the urbanized areas of Hong Kong Island, Kowloon, and the New Territories. The smart card data records trip transaction information including tap-in/out stations and times. The data used was from January 1st to March 31st, 2018 (86476 users and 19 million trips). According to the characteristics of actual data, the study considers the time interval of morning peak (MP, 07:00-10:00) and off-peak (OP, 10:00-15:00) to construct MKG and randomly selected 1000 individuals to infer the user-station attention, respectively.

The information utilized in our approach comprises time-dependent demand, OD flows, and user-station visit counts, all of which can be obtained from various mobility data sources such as smartcard data, GPS data, and geotagged data. None of these data sources involve sensitive or private information. The user-station visit counts data are identified solely by trips (without or with anonymized card ID) and do not contain any personally identifiable information, such as name, address, or email. Table 2 shows the structure of the AFC database used in the study.

Since the observed user-station visit counts data does not reflect the real user-station attention, the real user-station attention cannot be obtained from historical observations. To validate the proposed method, we use the synthetic user-station visit counts data generated from a pre-set 'real' user-station attention matrix. Algorithm 2 illustrates the pro-

Table 2 Example of AFC data structure in database

Card ID	Origin	Entry time	Destination	Exit time
xxxxxx_1	Tai Po Market	2018-01-02 20:03:49	University	2018-01-02 20:08:27
xxxxxx_2	Wong Tai Sin	2018-01-02 10:53:54	KowloonTong	2018-01-02 11:01:23
xxxxxx_3	KowloonTong	2018-01-02 21:22:57	Lok Fu	2018-01-02 21:31:43
xxxxxx_4	South Horizons	2018-01-02 09:44:11	Causeway Bay	2018-01-02 10:08:59
xxxxxx_5	Tai Wo Hau	2018-01-02 07:57:54	Causeway Bay	2018-01-02 08:33:51
xxxxxx_6	Hung Hom	2018-01-02 08:49:00	Tsuen Wan West	2018-01-02 09:11:54
xxxxxx_7	Tsuen Wan West	2018-01-02 17:59:25	Hung Hom	2018-01-02 18:25:38

cess of generating the synthetic user-station visit counts matrix. It takes inputs of pre-set ‘real’ user-station attention matrix \mathbf{Y} , number of stations n , number of users m , number of iterations (days) d , deviation σ , and outputs the simulated ‘observed’ user-station visit counts matrix \mathbf{X}^s . To keep the same travel pattern with the real-world dataset, the pre-set ‘real’ user-station attention matrix \mathbf{Y} is set as the normalized observed user-station visit counts matrix \mathbf{X} . Each user is simulated for 1000 days (i.e., $d = 1000$) and the deviation σ (e.g., $\sigma = 0.01$) is used to simulate the random variation of users’ visits on different days.

Algorithm 2 Synthetic data generation process.

Input: the pre-set user-station attention matrix Y , number of stations n

Output: the simulated ‘observed’ user-station visit counts matrix X^s

```

1: Initialize matrix  $X^s \leftarrow zeros(m, n)$ 
2: Normalize the matrix  $Y$  (the sum of each row is 1)
3: for  $i = 1 : m$  do
4:   for  $j = 1 : n$  do
5:     calculate the probability interval of each nonzero attention
        $p_{ij} = \frac{\sum_0^{j-1} y_{ij}}{\sum_0^j y_{ij}}$ 
6:   end for
7:   for  $k = 1 : d$  do
8:     randomly generate variable  $\sigma$  in the range of  $[0, 1)$ 
9:      $\theta^* = \theta + \sigma$ 
10:    for  $j = 1 : n$  do
11:      if  $\theta^* \in p_{ij}$  then
12:         $X_{ij}^s = X_{ij}^s + 1$ 
13:      end if
14:    end for
15:  end for
16: end for
17: return latent variables  $U, L_{ts}$ 

```

To visually check the quality of the synthetic data, we normalize the simulated visit counts matrix as the simulated attention matrix. We expect that the simulated attention matrix retains the structure of the pre-set real attention matrix but also has certain deviations. Fig. 4 shows the comparison of the pre-set (Fig. 4(a)) and the simulated (Fig. 4(b)) user-station attention matrix, and their differences (Fig. 4(c)) for 200 users. The difference is calculated by subtracting the pre-set from the simulated matrix. The result shows that

the output simulated user-station visit counts matrix retains the travel pattern of the real-world dataset and also contains certain variations which reasonably represent the studied user-station attention inference problem in practice.

4.2 Validation using synthetic data

The pre-set user-station attention matrix \mathbf{Y} is considered the ground-truth attention. The problem is to infer the user-station attention matrix $\hat{\mathbf{Y}}$ given partial simulated user-station visit counts matrix based on MKG. The model performance is validated by comparing the inferred user-station attention matrix with the pre-set one. Two performance metrics are used for synthetic data validation: 1) mean absolute error (MAE), and 2) root mean squared error (RMSE) metrics are used. The MAE and RMSE can be defined as:

$$MAE = \frac{\sum_{i,j} |y_{ij} - \hat{y}_{ij}|}{N} \tag{10}$$

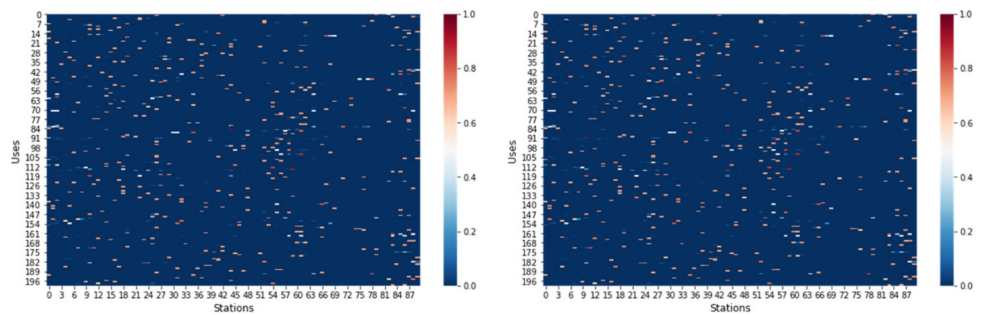
$$RMSE = \sqrt{\frac{\sum_{i,j} (y_{ij} - \hat{y}_{ij})^2}{N}} \tag{11}$$

y_{ij} is the ground-truth user-station attention. \hat{y}_{ij} is the inferred user-station attention, and N denotes the number of user-station pairs in the network.

The proposed method is validated under both sparse and dense training data settings, i.e., 10% and 30% of the positive value cells (trustworthy data) are selected as the training dataset. Moreover, the study compares the model performance with its variants (MKG capturing different information). The baseline models are as follows:

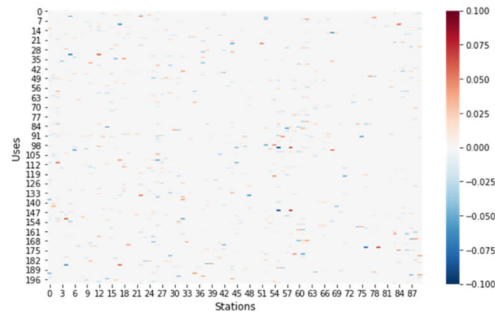
- **US_None.** The model captures the user similarity information (i.e., Eq. (3)) without MKG (not capturing station relations).
- **US_Lt.** The model captures user similarity (i.e., Eq. (2)) and stations’ temporal latent features (i.e., Eq. (4)) in representing MKG relations.

Fig. 4 Heat map comparison of pre-set ‘real’ and simulated user-station attention matrix



(a) Pre-set ‘real’ user-station attention matrix (normalized observed user-station visit counts matrix)

(b) Simulated user-station attention matrix (normalized simulated user-station visit counts matrix)



(c) Difference between the pre-set and simulated user-station attention matrix.

- **US_ $[L_t;L_s]$** . The model captures the user similarity (i.e., Eq. (2)) and stations’ temporal and spatial information in representing MKG relations using a rule-based approach (i.e., Eq. (6)). That is the stations’ latent representation is formed by concatenating the temporal latent matrix L_t and spatial latent matrix L_s .
- **US_ $[MKG]$** . The model considers both user similarity and station-station relations modeled using MKG (i.e., Eq. (9)). The stations’ latent representation simultaneously captures the spatiotemporal latent features using the NN-based nonlinear matrix decomposition.

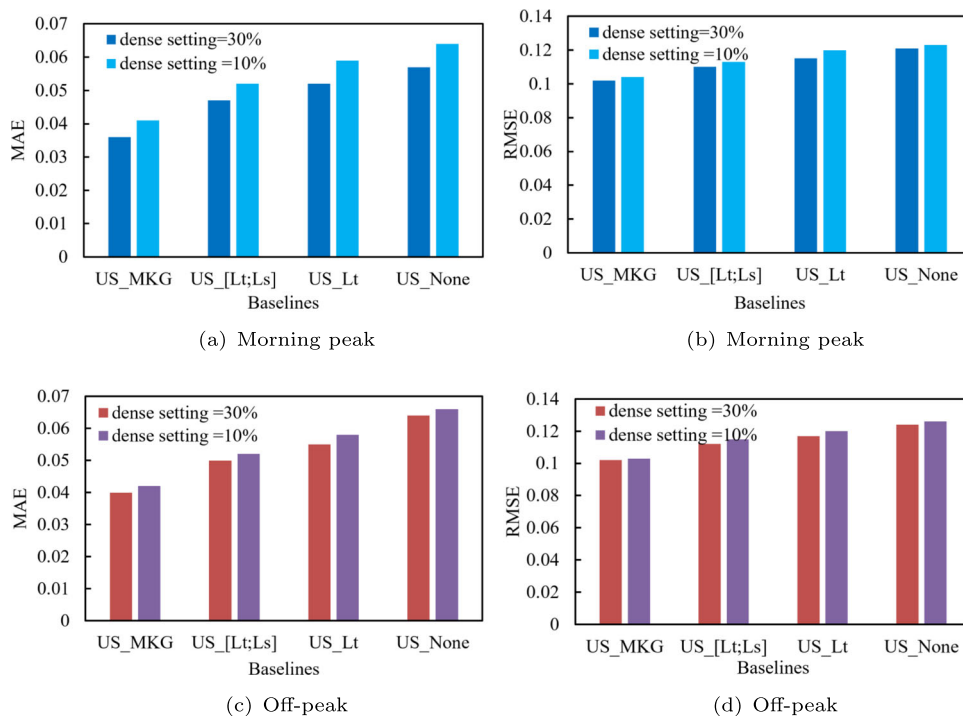
We use a grid-search method to find the optimal parameter settings in Algorithm 1 for the experimental analysis. The optimal hyper-parameter values in Algorithm 1 are: $k = 20$, $k_t = 10$, $k_s = 10$, $k_{ts} = 20$, $p = 30$, $w = 20$, $\lambda = 0.01$, $\alpha = 1e - 5$, $N = 100$.

Figure 5 shows the models comparison of baselines based on MAE and RMSE. In general, incorporating MKG significantly improves the model inference performance over the **US_None** model for both morning peak and off-peak periods. The **US_MKG** model has the best performance, followed by the **US_ $[L_t;L_s]$** and **US_ L_t** models. Compared with **US_None**, **US_MKG** reduces the MAE and RMSE by about 35% and 16% at morning peak and off-peak,

respectively. Compared to the morning peak, the performance in off-peak is worse for all models. This is because morning peak travels tend to be more regular for activities. Furthermore, it shows that the inference performance is very close under the two training data settings with different density levels, which indicates its potential in dealing with sparse user-station attention visit count matrix in inferring the true attentions.

To further illustrate the impact of different data densities and different training datasets on the model, the sensitivity analysis is performed for the optimal model **US_MKG**. We compare the model performance trained using only positive entries (PG, without zero entries) and the full entries (FG, including zero entries) with different density settings for training (i.e., 10%, 30%, and 100%). Figure 6 shows the performance comparison. Compared with using the full entries as the training set, the models using only positive entries perform better in both the morning peak and off-peak scenarios. It indicates that reliable user-station visit counts information could benefit the inference of true user-station attention. In addition, it shows that the model performance is positively correlated to the number of training samples. The model performance of the off-peak is slightly worse than that of the morning peak, as the travel characteristics of users are more regular in the morning peak.

Fig. 5 Performance comparison of baselines. (a) MAE in the morning peak; (b) RMSE in the morning peak; (c) MAE in the off-peak; (d) RMSE in the off-peak



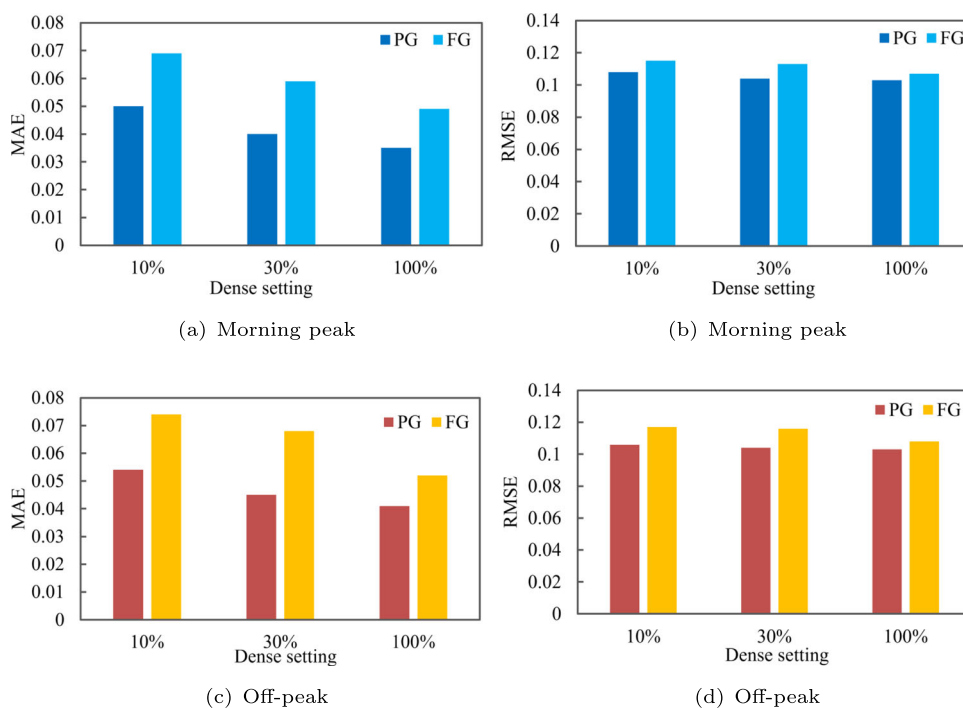
4.3 Validation using actual data

We use actual data to further validate the proposed model performance by comparing it with baseline models. The ‘observed’ user-station attention \mathbf{Y} (normalized observed user-station visit counts matrix \mathbf{X}) is considered the ground-

truth attention. For the actual data, the recall is used to compare the performance. Since zeros in the observed user-station visit counts matrix \mathbf{X} represents unreliable data, the recall only considers the positive values. For each user, the $recall@K$ was defined as:

$$recall@K = \frac{TP@K}{TP@K + FN@K} \tag{12}$$

Fig. 6 Performance comparison of different training sets in the morning peak and off-peak. (a) MAE in the morning peak; (b) RMSE in the morning peak; (c) MAE in the off-peak; (d) RMSE in the off-peak



where $TP@K$ is the number of stations that the user pays attention to (positive value) in the top K stations, and $TP@K + FN@K$ is the total number of stations that the user pays attention to. For example, for a user i , we can get the inferred user-station attention list $\{\hat{y}_{i1}, \hat{y}_{i2}, \hat{y}_{i3}, \dots, \hat{y}_{in}\}$ and the corresponding ground-truth user-station attention list is $\{y_{i1}, y_{i2}, y_{i3}, \dots, y_{in}\}$. Then we can get a candidate set $\{(\hat{y}_{i1}, y_{i1}), (\hat{y}_{i2}, y_{i2}), \dots, (\hat{y}_{in}, y_{in})\}$, where $\hat{y}_{ij} \geq 0$ and $y_{ij} \geq 0$. When the length of the candidate set is greater than K , $TP@K$ is K . Otherwise, K is the length of the candidate set. $TP@K + FN@K$ is the number of ground-truth user-station attention greater than zero. For example, suppose $K = 2$, the length of the candidate set is 5, and there are 20 ground-truth user-station attention greater than zero, then $recall@K$ is 0.1. The recall for the entire matrix is the average recall of all users.

The experiment compares the proposed model performance with baseline models introduced in the above section. Similarly, the experiment compares the performance of the framework under sparse (insufficient) and dense (sufficient) training settings, i.e., 10% and 30% positive entries are selected as the training dataset. Figure 7 shows the experimental results. Figure 7(a) and (c) show the performance of morning peak and off-peak under a dense setting (i.e., 30%),

whereas Fig. 7(b) and (d) show the results under a sparse training setting (i.e., 10%). In general, the results indicate that the MKG greatly improves the model’s performance in inferring true attention. The **US_MKG** model consistently performs better than others under different settings. Since **US_None** does not introduce any MKG information, it performs worst in both morning peak and off-peak, especially under a sparse setting. Compared with **US_Lt** and **US_[Lt; Ls]** which separately consider the temporal and spatial information, **US_MKG** simultaneously considers the stations’ latent representations and their relations which can more effectively capture hidden information. Similarly, we found that the performance in the morning peak is higher than that in the off-peak as the travel is more concentrated and regular in the morning peak.

4.4 Implications and discussions

Figure 8 visualizes a user’s attention to different stations inferred using the proposed method. The orange squares represent visited stations from the user’s historical trajectories. The green squares denote the stations (unobserved previously) that the user may also be interested in visiting. The user’s origin is ‘Lo Wu’ connecting the railway station

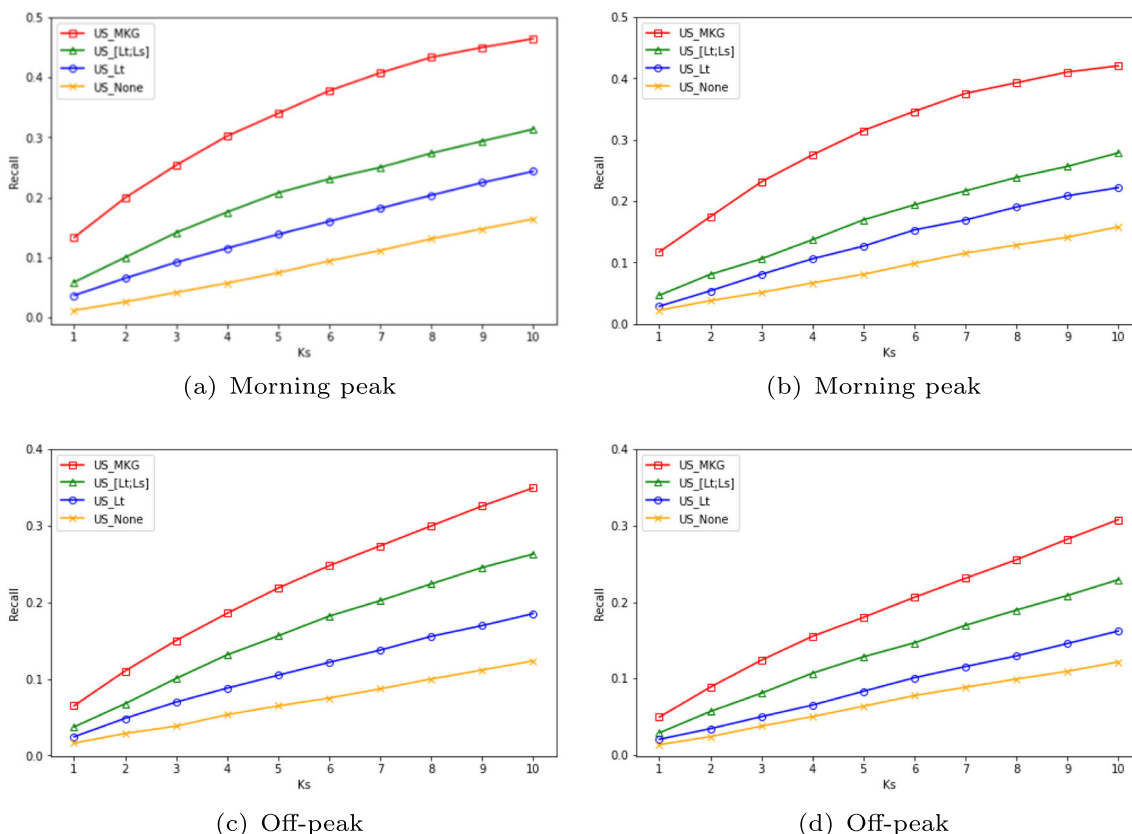
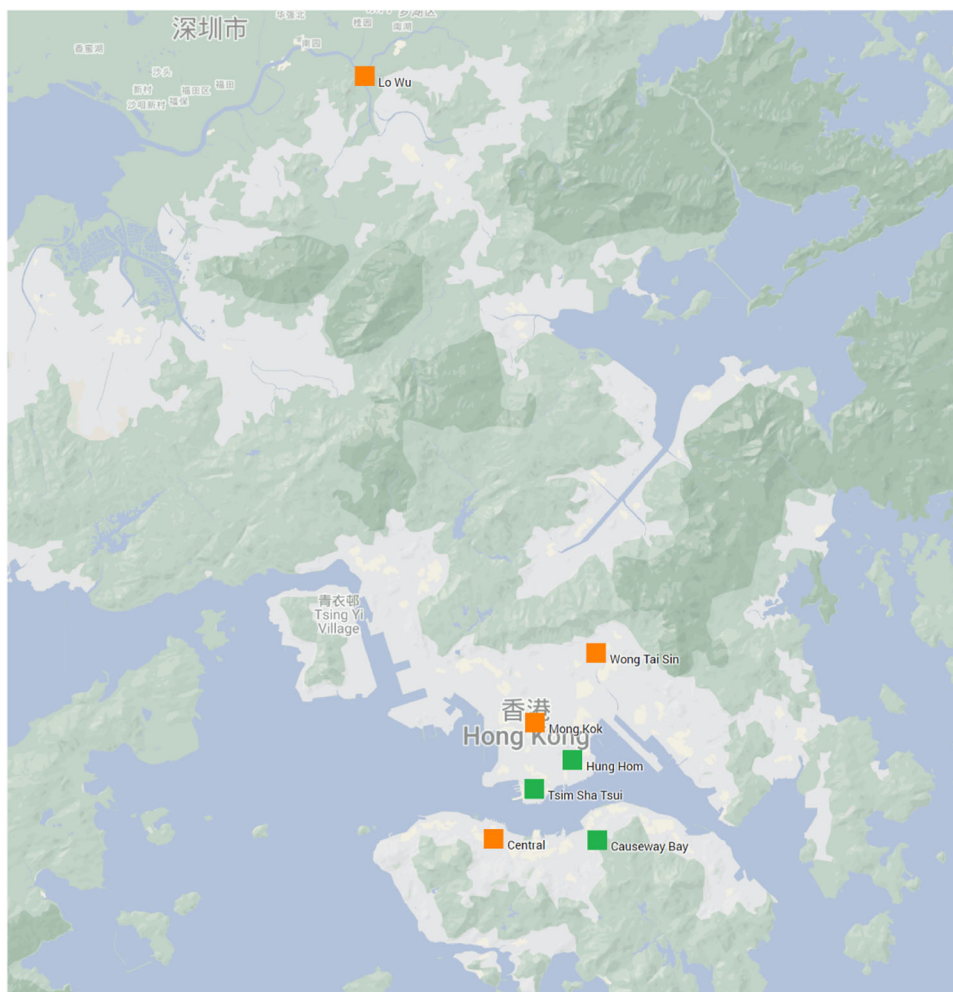


Fig. 7 Performance comparison of baselines based on recall@K ($K=1, 2, \dots, 10$). (a) Morning peak and dense setting = 30%; (b) Morning peak and dense setting = 10%; (c) Off-peak and dense setting = 30%; (d) Off-peak and dense setting = 10%

Fig. 8 Example of a user's attention to stations (Orange square: visited stations; green square: inferred stations with attention)



'Luohu' in Shenzhen. The user visited stations 'Wong-TaiSin', 'MongKok', and 'Central' in the central business areas with places of interest and shopping malls. So the user may be a buyer and his/her travel activity (station-station relationship) is likely to be 'shopping' or 'purchasing'. The attention stations inferred from the model are 'Hung Hom', 'Tsim Sha Tsui', and 'CausewayBay' which are close to these observed visiting stations and are also places of interest and shopping areas. Therefore, the inferred attention stations are reasonable and the user may have interest to go to these areas for shopping activities.

Note that the studied user-station attention problem is mathematically similar to the user-item rating problem in recommendation systems. The two problems share a commonality in that they both utilize user similarity and relationships to infer items or stations that have no prior interaction history. However, there are also differences between the two problems. For instance, the context of travel behavior in transportation is different from that of item recommendations. For user-station attention, the user's travel pattern remains relatively stable within a certain period and geographic area. In

contrast, the user's choice of items is highly changeable and unpredictable, as it is influenced by various factors, including personal preferences, current needs, and external stimuli. For example, while users may frequently travel to the same stations for work or school, their shopping habits may vary significantly from day to day, with some days focused on groceries, while others may involve purchases of clothing, electronics, or other goods.

Analyzing user-station attention can provide insights into the underlying factors driving user behavior within public transport systems, which can be leveraged to develop predictive models of user behavior and demand. For instance, historical data on user-station attention can be used to forecast the most heavily used stations during different times of the day or week, allowing for optimized scheduling and routing of public transport services. Additionally, user-station attention analysis can facilitate the development of personalized location recommendation systems based on a user's travel history and preferences, such as recommending the best station to start a journey based on the destination, travel time, and other relevant factors. Moreover, the station-station

relationship in MKG is also of great significance for other application scenarios. For example, in precision marketing, it can assist to locate more target groups and place targeted advertisements based on the main travel activities of stations/locations. The knowledge obtained from the investigation of user-station attention can be utilized to enhance the efficiency, efficacy, and overall user experience of public transportation systems, rendering them more appealing and user-friendly while contributing to promoting sustainable urban mobility.

For the practical implementation of our approach, the absence of privacy or sensitive data previously discussed is a prerequisite. Furthermore, the accuracy and reliability of the data are critical to ensure the validity and practicality of our model in real-world scenarios. Therefore, data should be collected and processed in a way that mitigates the risk of bias. Moreover, for greater generalizability, our proposed approach should be applied to datasets and contexts beyond those used in the initial analysis. Additionally, the involvement of stakeholders such as transportation authorities, policymakers, and users is essential in the design and implementation to ensure that our approach addresses real-world needs and is implemented effectively.

5 Conclusion

This paper introduces the user-station attention inference problem from partially observed user-station visit counts data in public transport. We develop a matrix decomposition method capturing simultaneously user similarity and station-stations relationships using the knowledge graph. Also, we propose a neural network-based nonlinear decomposition approach to effectively extract the MKG relations from smart card data capturing the latent spatiotemporal travel dependencies.

The experiments using both synthetic data and smart card data verify the validity of the proposed model by comparing the with benchmark approaches. The results illustrate the value of the knowledge graph in contributing to the inference of user-station attention. Compared with the model without MKG, the model with MKG will improve the model performance by 35% in MAE and 16% in RMSE. The results also show that it is important to select the positive observations in training the model (rather than using all observations with zero values). The model is less sensitive to the dense settings (10%, 30%, and 100% of samples used for training) provided only positive observations are used. Future work will focus on explaining the inferred user-station attentions by exploring more context information (e.g., points of interest, station features), and also examine its value in contributing to prediction applications, such as individual mobility prediction.

Acknowledgements This work was supported by China Scholarship Council under Grant 202006950007 and KTH Digital Futures (cAIM-BER).

Author Contributions The authors confirm contributions to the paper as follows: study conception and design: Zhenliang Ma, Pengfei Zhang, Qi Zhang; methodology and data collection: Qi Zhang, Pengfei Zhang, Zhenliang Ma; analysis and interpretation of results: Qi Zhang, Zhenliang Ma, Erik Jenelius, Xiaolei Ma, Yuanqiao Wen; draft manuscript preparation: Qi Zhang and Zhenliang Ma; manuscript revision: Zhenliang Ma, Pengfei Zhang, Erik Jenelius, Xiaolei Ma, Yuanqiao Wen; All authors reviewed the results and approved the final version of the manuscript.

Funding Open access funding provided by Royal Institute of Technology.

Data Availability The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics and informed consent for data used This article does not contain any studies with human participants or animals performed by the author.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Zhao Z, Koutsopoulos HN, Zhao J (2020) Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. *Transportation Research Part C: Emerging Technologies* 116:102627
2. Gao Z, Janssens D, Jia B, Wets G, Yang Y et al.: Identifying business activity-travel patterns based on gps data. *Transportation Research Part C: Emerging Technologies* 128:103136 (2021)
3. Hu L, Li Z, Ye X (2020) Delineating and modeling activity space using geotagged social media data. *Cartography and Geographic Information Science* 47(3):277–288
4. Guo, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., He, Q.: A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* (2020)
5. Da'u, A., Salim, N.: Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review* 53(4), 2709–2748 (2020)

6. Fan, H., Zhong, Y., Zeng, G., Ge, C.: Improving recommender system via knowledge graph based exploring user preference. *Applied Intelligence*, 1–13 (2022)
7. Yu T, Li J, Yu Q, Tian Y, Shun X, Xu L, Zhu L, Gao H (2017) Knowledge graph for tcm health preservation: Design, construction, and applications. *Artificial intelligence in medicine* 77:48–52
8. Zhang Y, Sheng M, Zhou R, Wang Y, Han G, Zhang H, Xing C, Dong J (2020) Hkgb: an inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated. *Information Processing & Management* 57(6):102324
9. Zeng X, Tu X, Liu Y, Fu X, Su Y (2022) Toward better drug discovery with knowledge graph. *Current opinion in structural biology* 72:114–126
10. Jia Y, Qi Y, Shang H, Jiang R, Li A (2018) A practical approach to constructing a knowledge graph for cybersecurity. *Engineering* 4(1):53–60
11. Berven A, Christensen OA, Moldeklev S, Opdahl AL, Villanger KJ (2020) A knowledge-graph platform for newsrooms. *Computers in Industry* 123:103321
12. Chen T, Zhang Y, Qian X, Li J (2022) A knowledge graph-based method for epidemic contact tracing in public transportation. *Transportation Research Part C: Emerging Technologies* 137:103587
13. Ahmed U, Srivastava G, Djenouri Y, Lin JC-W (2022) Knowledge graph based trajectory outlier detection in sustainable smart cities. *Sustainable Cities and Society* 78:103580
14. Tan J, Qiu Q, Guo W, Li T (2021) Research on the construction of a knowledge graph and knowledge reasoning model in the field of urban traffic. *Sustainability* 13(6):3191
15. Hsu, P.-Y., Chen, C.-T., Chou, C., Huang, S.-H.: Explainable mutual fund recommendation system developed based on knowledge graph embeddings. *Applied Intelligence*, 1–26 (2022)
16. Ji S, Pan S, Cambria E, Marttinen P, Philip SY (2021) A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems* 33(2):494–514
17. Dhelim S, Aung N, Ning H (2020) Mining user interest based on personality-aware hybrid filtering in social networks. *Knowledge-Based Systems* 206:106227
18. Li, S., Yang, B., Li, D.: Entity-driven user intent inference for knowledge graph-based recommendation. *Applied Intelligence*, 1–17 (2022)
19. Gao, M., Li, J.-Y., Chen, C.-H., Li, Y., Zhang, J., Zhan, Z.-H.: Enhanced multi-task learning and knowledge graph-based recommender system. *IEEE Transactions on Knowledge and Data Engineering* (2023)
20. Cai X, Xie L, Tian R, Cui Z (2022) Explicable recommendation based on knowledge graph. *Expert Systems with Applications* 200:117035
21. Zhang R, Hristovski D, Schutte D, Kastrin A, Fiszman M, Kilicoglu H (2021) Drug repurposing for covid-19 via knowledge graph completion. *Journal of biomedical informatics* 115:103696
22. Li L, Wang P, Yan J, Wang Y, Li S, Jiang J, Sun Z, Tang B, Chang T-H, Wang S et al.: Real-world data medical knowledge graph: construction and applications. *Artificial intelligence in medicine* 103:101817 (2020)
23. Xu J, Kim S, Song M, Jeong M, Kim D, Kang J, Rousseau JF, Li X, Xu W, Torvik VI, et al. Building a pubmed knowledge graph. *Scientific data* 7(1):205 (2020)
24. Yang, B., Liao, Y.-m.: Research on enterprise risk knowledge graph based on multi-source data fusion. *Neural Computing and Applications* 34(4), 2569–2582 (2022)
25. Du J, Wang S, Ye X, Sinton DS, Kemp K (2022) Gis-kg: building a large-scale hierarchical knowledge graph for geographic information science. *International Journal of Geographical Information Science* 36(5):873–897
26. Zhang Q, Wen Y, Zhou C, Long H, Han D, Zhang F, Xiao C (2019) Construction of knowledge graphs for maritime dangerous goods. *Sustainability* 11(10):2849
27. Mao S, Zhao Y, Chen J, Wang B, Tang Y (2020) Development of process safety knowledge graph: A case study on delayed coking process. *Computers & Chemical Engineering* 143:107094
28. Shan S, Cao B (2017) Follow a guide to solve urban problems: the creation and application of urban knowledge graph. *IET Software* 11(3):126–134
29. Wang, P., Liu, K., Jiang, L., Li, X., Fu, Y.: Incremental mobile user profiling: Reinforcement learning with spatial knowledge graph for modeling event streams. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 853–861 (2020)
30. Li G, Chen Y, Liao Q, He Z (2022) Potential destination discovery for low predictability individuals based on knowledge graph. *Transportation Research Part C: Emerging Technologies* 145:103928
31. Cui Z, Zhao P, Hu Z, Cai X, Zhang W, Chen J (2021) An improved matrix factorization based model for many-objective optimization recommendation. *Information Sciences* 579:1–14
32. Berahmand, K., Mohammadi, M., Saberi-Movahed, F., Li, Y., Xu, Y.: Graph regularized nonnegative matrix factorization for community detection in attributed networks. *IEEE Transactions on Network Science and Engineering* (2022)
33. Lin X, Boutros PC (2020) Optimization and expansion of non-negative matrix factorization. *BMC bioinformatics* 21(1):1–10
34. Smirnova A, Cudré-Mauroux P (2018) Relation extraction using distant supervision: A survey. *ACM Computing Surveys (CSUR)* 51(5):1–35
35. Nayak T, Majumder N, Goyal P, Poria S (2021) Deep neural approaches to relation triplets extraction: A comprehensive survey. *Cognitive Computation* 13(5):1215–1232
36. Wang, H., Qin, K., Zakari, R.Y., Lu, G., Yin, J.: Deep neural network-based relation extraction: an overview. *Neural Computing and Applications*, 1–21 (2022)
37. Wang X, El-Gohary N (2023) Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements. *Automation in Construction* 147:104696
38. Zhuang, C., Yuan, N.J., Song, R., Xie, X., Ma, Q.: Understanding people lifestyles: Construction of urban movement knowledge graph from gps trajectory. In: *Ijcai*, pp. 3616–3623 (2017)

Authors and Affiliations

Qi Zhang¹ · Zhenliang Ma¹ · Pengfei Zhang² · Erik Jenelius¹ · Xiaolei Ma³ · Yuanqiao Wen⁴

Qi Zhang
qzhan@kth.se

Pengfei Zhang
zhangpfscut@gmail.com

Erik Jenelius
erik.jenelius@abe.kth.se

Xiaolei Ma
xiaolei@buaa.edu.cn

Yuanqiao Wen
wenyqwhut@foxmail.com

¹ Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm 10044, Sweden

² Institute of Physics, Henan Academy of Sciences, Zhengzhou 450002, China

³ Beijing Key Laboratory for Cooperative Vehicle Infrastructure System and Safety Control, Beihang University, Beijing 100191, China

⁴ Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430070, China