Optimal Transport Enhanced Cross-City Site Recommendation

Xinhang Li
Tsinghua University
Beijing, China
xh-li20@mails.tsinghua.edu.cn

Xiangyu Zhao*
City University of Hong Kong
Hong Kong
xianzhao@cityu.edu.hk

Zihao Wang
HKUST
Hong Kong
zwanggc@cse.ust.hk

Yang Duan
UIUC
Urbana-Champaign, IL, USA
yangd4@illinois.edu

Yong Zhang*
Tsinghua University
Beijing, China
zhangyong05@tsinghua.edu.cn

Chunxiao Xing
Tsinghua University
Beijing, China
xingcx@tsinghua.edu.cn

ABSTRACT

Site recommendation, which aims at predicting the optimal location for brands to open new branches, has demonstrated an important role in assisting decision-making in modern business. In contrast to traditional recommender systems that can benefit from extensive information, site recommendation starkly suffers from extremely limited information and thus leads to unsatisfactory performance. Therefore, existing site recommendation methods primarily focus on several specific name brands and heavily rely on fine-grained human-crafted features to avoid the data sparsity problem. However, such solutions are not able to fulfill the demand for rapid development in modern business. Therefore, we aim to alleviate the data sparsity problem by effectively utilizing data across multiple cities and thereby propose a novel **O**ptimal **T**ransport enhanced **C**ross-city (**OTC**) framework for site recommendation. Specifically, OTC leverages optimal transport (OT) on the learned embeddings of brands and regions separately to project the brands and regions from the source city to the target city. Then, the projected embeddings of brands and regions are utilized to obtain the inference recommendation in the target city. By integrating the original recommendation and the inference recommendations from multiple cities, OTC is able to achieve enhanced recommendation results. The experimental results on the real-world OpenSiteRec dataset, encompassing thousands of brands and regions across four metropolises, demonstrate the effectiveness of our proposed OTC in further improving the performance of site recommendation models.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; **Learning to rank**; **Location based services**.

KEYWORDS

site recommendation, cross-domain recommendation, optimal transport

*Xiangyu Zhao and Yong Zhang are the corresponding authors.

SIGIR '24, July 14–18, 2024, Washington, DC, USA.

ACM ISBN 979-8-4007-0431-4/24/07.
https://doi.org/10.1145/3626772.3657757

ACM Reference Format:
Xinhang Li, Xiangyu Zhao, Zihao Wang, Yang Duan, Yong Zhang, and Chunxiao Xing. 2024. Optimal Transport Enhanced Cross-City Site Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3626772.3657757

1 INTRODUCTION

Site recommendation [7, 20, 31], which is a type of recommendation in the business area, has attracted increasing interest in recent years. Since an appropriate site will bring substantial profits while an inappropriate site may lead to business failure [24, 46, 47], accurately determining the most suitable location is crucial for business decision-making. However, information overload is more and more severe in our daily life and so as in the business area. Traditional recommendation aims to alleviate information overload and improve the user experience on the Web [1, 22, 25, 39], which have achieved promising results. Therefore, site recommendation is similarly proposed to predict the optimal location for commercial brands to open new branches in a data-driven manner to assist decision-making in modern business [41, 61].

Unlike traditional recommendation tasks that are oriented for millions of users and items, site recommendation targets commercial brands and city regions which are typically a few thousand. The extremely limited information may lead to unsatisfactory performance and present a serious challenge to site recommendation. Therefore, existing site recommendation methods mainly focus on several specific name brands and introduce fine-grained human-crafted features to avoid the data sparsity problem. For example, Geo-Spotting [20] explores the effectiveness of several geographical and mobility features on 3 popular fast food brands in New York City. DD3S [49] conducts site recommendation for 2 coffee shop brands and 2 chain hotel brands with various heuristic features such as traffic convenience and competition degree. Meanwhile, there are also some approaches that aim to leverage the data across different cities. CityTransfer [14] transfers knowledge to improve the performance of 3 chain hotel brands in a new city. WANT [33] predicts the potential consumption with massive user, geographical, commercial and time features for recommendation.

Although these methods have successfully tackled their own tasks, we argue that such solutions are unable to fulfill the demand for rapid development in modern business. First, the high

Table 1: Example of an instance in Tokyo of OpenSiteRec.

General	**ID** 259964052	**Name** Starbucks	**City** Tokyo
Commercial	**Broad Category** Food and Beverage	**Medium Category** Beverage Shop	**Narrow Category** Coffee and Tea Shop
	Brand Starbucks	**Competitive** Doutor, Tully's Coffee, ...	**Related** McDonald's, KFC, ...
Geographical	**Longitude** 139.7390476	**Latitude** 35.684236	**District** Chiyoda City
	Region Kōjimachi 3 Chōme	**NearBy** Kōjimachi 6 Chōme, Kioichō, ...	**Similar** Kōjimachi 4 Chōme, ...

dependence on the fine-grained human-crafted features will introduce human bias and thus the quality of the heuristic features will strongly affect the performance. The design of these features is not only time-consuming and labor-intensive but also requires rich domain knowledge, which is inconsistent with the purpose of the automatic data-driven manner. Next, the original data is difficult to obtain due to the problem of information integrity or the consideration of privacy protection. As a result, the information is typically very scarce and the heuristic features are not always available. Finally, the need for sufficient feature information forces the existing methods to take very few brands into account. Such a pathway of avoiding rather than solving the data sparsity problem strongly limits the applicability of site recommendation in practice. Therefore, a site recommendation model that is able to support all the brands without requirements of comprehensive information and heuristic features, is preferable and necessary for real-world application.

In order to achieve this goal, we aim to alleviate the data sparsity problem by leveraging the data from other cities and thereby propose a novel **O**ptimal **T**ransport enhanced **C**ross-city (**OTC**) framework for site recommendation. Specifically, we adopt the OpenSiteRec dataset which contains thousands of brands and regions in four metropolises. In this scenario, the recommendation is conducted by capturing and utilizing the correlations between different brands and regions, i.e., a collaborative filtering [8, 29, 40, 44] based method rather than a content-based method [45, 53, 59] like previous works. Here, we introduce optimal transport (OT) to site recommendation by projecting the brands and regions from the source city to the target city to transfer the inference recommendation as additional predictions. Then, OTC is able to achieve an enhancement over the backbone models by aggregating the direct recommendation results of the target city and inference recommendation results of multiple source cities. The convincing experimental results fully demonstrate the effectiveness of our proposed OTC in further improving the site recommendation performance. The contribution of this paper can be summarized as follows:

- We introduce optimal transport into site recommendation to effectively leverage the cross-city information in a light-weighted manner.
- We propose a novel **O**ptimal **T**ransport enhanced **C**ross-city (**OTC**) framework for site recommendation, which is able to further improve the performance of backbone models.
- We conduct extensive experiments on the real-world OpenSiteRec dataset and the results fully demonstrate the effectiveness and the efficiency of our proposed OTC method.
- The detailed analysis of recommendation results also indicates the potential of OTC in finding novel associations across different cities or valuable ideas for business success.

2 PRELIMINARY

In this section, we will first introduce the OpenSiteRec datasets used in this work. Then, we will provide the formal definitions of the problem that OTC addresses to elaborate on our work. Finally, we will deliver the basic knowledge of optimal transport theory to give a quick understanding of the techniques in OTC.

2.1 Data Description

To evaluate the effectiveness of OTC, we use the publicly available dataset OpenSiteRec [27]. It is a comprehensive open-source benchmarking dataset specifically designed for site recommendation research. Specifically, it collects plenty of real-world information on four major international cities: Chicago, New York City, Singapore, and Tokyo at December 1st, 2022. Compared to other datasets for site recommendation, OpenSiteRec provides a more extensive collection of information on a much larger number of brands. Moreover, the trustworthy commercial relations of these real-world entities are also included in OpenSiteRec and represented by a heterogeneous graph structure.

There are five types of entities in OpenSiteRec: Brand, Category, POI, Business Area, and Region. Brands denote institutions or commercial brands while categories in this dataset have three levels based on different criteria and refer to types of stores or functions of places. POIs (Point of Interests) represent the specific sites that provide different services to residents. Business areas are the areas planned and organized by governments to mainly cluster business companies and boost business activities while regions refer to geographical areas arranged by governments to help development and management. In particular, each POI belongs to one brand and is located in one region. Thus, the POIs serve as the bridge to connect commercial and geographical entities. Typically, an instance in OpenSiteRec is shown as Table 1.

As illustrated by the example, there are many different types of features available for each instance. However, the extreme differences across these features make the incorporation of them a really hard work and we do not want to introduce human bias as mentioned above. Therefore, we apply the vanilla settings as same as the benchmarking experimental settings in OpenSiteRec in this paper, which only utilize the Brand, POI and Region parts and formulate the site recommendation task as a top-n recommendation for brands.

Furthermore, due to the real-world essence, the data of OpenSiteRec is extremely sparse, highly diverse across different cities, very imbalanced on both brands and regions. Typically, the data sparsity is so high that each brand only has 5 sites on average. Meanwhile, the consumption habits and lifestyles of residents in different cities vary greatly, leading to significant differences in the distribution of POI categories among the cities. The distributions of brands and regions are also dramatically imbalanced that more than 80% of sites belong to less than 20% of brands and locate at less than 30% of regions. These phenomena present a serious challenge of alleviating the data sparsity problem to the site recommendation task, which is also the one we aim to tackle in this paper.

2.2 Problem Definition

In order to better illustrate the site recommendation task and the cross-city scenario, we deliver the formal problem definition here.

As mentioned above, there are three kinds of entities defined for site recommendation in OpenSiteRec, i.e., brands, POIs and regions. The concept of brands is adopted to unify various concepts and categories of sites, e.g., stores, shops, hotels, and public institutions. Similarly, the concept of regions is adopted to represent all the parts of a city according to the government planning. The concept of POIs is adopted to connect the brands and the regions where each POI could be seen as a brand-region pair to represent a site. Therefore, the site recommendation task could be formulated as:

Definition 2.1. Let $\mathcal{U} = \{u_1, u_2, ..., u_M\}$ denotes the set of brands and $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ denotes the set of regions of the given city. Then, the matrix of POIs for the city can be defined as $\mathbf{P} = \{0, 1\} \in \mathbb{R}^{M \times N}$. For each POI, if it belongs to the brand b_i and is located at the region v_j, this will contribute to $P_{ij} = 1$. If $\mathbf{P}$ has $P_{ij} = 0$, it means the brand b_i does not have a venue at the region v_j in the city. Given the above definitions, the site recommendation task aims at *making predictions of the region lists for the given brands*, i.e., predicting $\hat{\mathbf{P}}$.

Under the cross-city scenario, we aim to leverage the information from other cities to improve the performance of the target city. Suppose we have multiple source cities, represented by $\mathcal{S} = \{s_1, s_2, ..., s_K\}$ and the target city is denoted by symbol t. Then, the cross-city scenario is defined as follows:

Definition 2.2. Let $\mathcal{U}^s \in \mathbb{R}^{M_s}$ and $\mathcal{V}^s \in \mathbb{R}^{N_s}$ denotes the set of brands and regions in the source city s, $\mathcal{U}^t \in \mathbb{R}^{M_t}$ and $\mathcal{V}^t \in \mathbb{R}^{N_t}$ denotes the set of brands and regions in the target city t, respectively. The POI matrices of the source city s and the target city t are presented as $\mathbf{P}^s \in \mathbb{R}^{M_s \times N_s}$ and $\mathbf{P}^t \in \mathbb{R}^{M_t \times N_t}$. Therefore, the cross-city site recommendation aims at *predicting* $\hat{\mathbf{P}}^t$ *of the target city with additional information of every source city* $s \in \mathcal{S}$.

The formal problem definitions give a clearer view of cross-city site recommendation and the proposed method is strictly based on them in this paper.

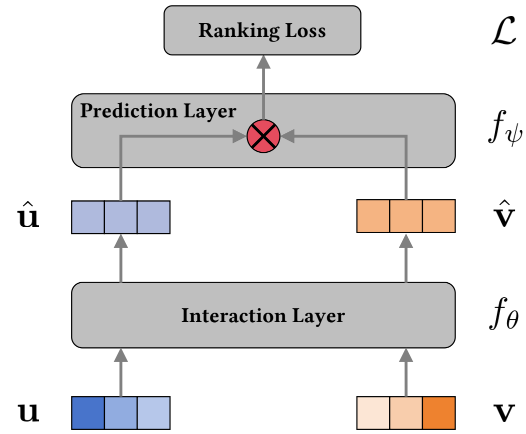

Figure 1: Backbone model structure of OTC. From bottom to top, there are input or embedding layer, interaction layer and prediction layer with ranking loss, respectively.

2.3 Optimal Transport

Optimal Transport (OT) aims at finding the optimal alignment between the source distribution and the target distribution with given conditions. From the view of transformation, OT problem can be seen as an assignment of weights or a projection, from the source distribution to the target distribution.

Suppose we have two discrete distributions $\mu = \sum_{i=1}^{M} u_i \delta_{x_i}$ and $\nu = \sum_{j=1}^{N} v_j \delta_{y_j}$. Specifically, δ is the delta function, $x_i, y_j \in \mathcal{X}$ are the feature vectors in space $\mathcal{X}$, and $\sum_{i=1}^{M} u_i = \sum_{j=1}^{N} v_j = 1$ are the two marginal distributions. Then, a typical OT problem can be formulated as obtaining the optimal alignment matrix T via minimizing the transport cost C as follows:

$$C(\mu, \nu) = \min_{\mathbf{T} \in \mathcal{P}} \sum_{i=1}^{M} \sum_{j=1}^{N} T_{ij} c(x_i, y_j), \tag{1}$$

where $\mathcal{P} = \{\mathbf{T} \in [0, 1]^{m \times n} : \sum_i T_{ij} = v_j, \sum_j T_{ij} = u_i\}$ denotes the set of all possible transport plans and $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ denotes the cost measure of two feature vectors. This formulation shows the basic form of OT and the minimal value of $C(\mu, v)$ is also recognized as Wasserstein distance between two distributions.

Although Wasserstein metric is powerful in addressing the alignment problem between two distributions, the restriction of the distributions being in the same space makes it unsuitable for the cross-domain scenario with data from different domains. Therefore, a quadratic formulation of OT, namely Gromov-Wasserstein (GW) metric, is often applied when the distributions of source domains $\mathcal{S}$ and target domains $\mathcal{T}$ are different. Different from Wasserstein metric, GW metric aims at aligning the intra-domain structures rather than the instances. Specifically, the optimization problem of GW distance is shown as follows:

$$\min_{\mathbf{T} \in \mathcal{P}} \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} d(c_s(x_{i_1}, x_{i_2}), c_t(y_{j_1}, y_{j_2}))^2 T_{i_1 j_1} T_{i_2 j_2}, \tag{2}$$

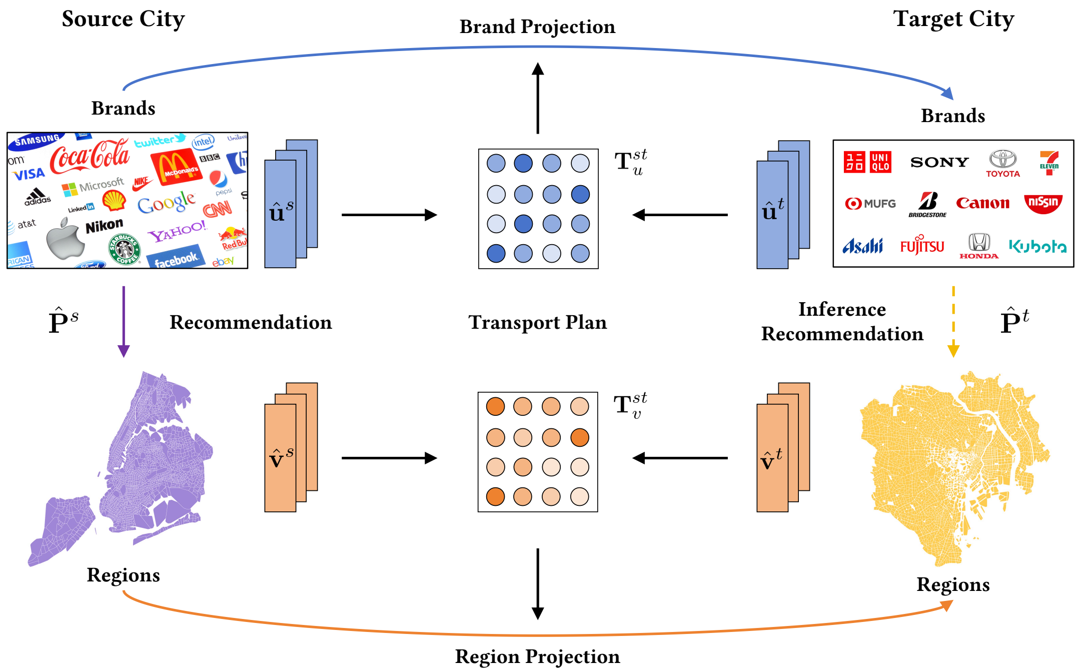

Figure 2: Overview of optimal transport enhanced cross-city model. The recommendations in the target city are inferred by transferring the recommendations in the source city via brand and region projection.

where c_s and c_t are the cost measure of source and target domains respectively, and d is a non-negative function to serve as the distance measure across domains. From another perspective, GW distance can be deemed as 'the distance between distances' and thus the optimization process of GW distance is well suitable for the cross-domain scenario.

3 METHODOLOGY

In this section, we will illustrate the proposed method OTC in detail. First, we will describe the backbone model structure for site recommendation in a single city. Then, we will introduce the novel optimal transport enhanced cross-city model for site recommendation to transfer knowledge from the source city to the target city. Finally, we will explain how to integrate knowledge from multiple cities for joint inference.

3.1 Backbone

As described above, the site recommendation task is essentially formulated as a top-n recommendation task. Therefore, regardless of the cross-city scenario, we first need a backbone model to achieve the site recommendation task in a single city. Typically, the structure of a top-n recommendation model can be split into three layers as shown in Figure 1: input/embedding layer, interaction layer and prediction layer. Note that the backbone model for OTC could be any specific model with such a structure and retrievable embedding representations for both sides. In this subsection, we merely aim to describe a general structure and training strategy rather than a specific model.

3.1.1 Embedding Layer. Since the main concern of this paper is to leverage cross-city information, we apply the vanilla settings with only IDs as the input features. Specifically, we project the IDs of the given brand $u \in \mathcal{U}$ and region $v \in \mathcal{V}$ into d-dimensional vectors by embedding matrices $\mathbf{U} \in \mathbb{R}^{M\times d}$ and $\mathbf{V} \in \mathbb{R}^{N\times d}$ to obtain ID embeddings $\mathbf{u}$ and $\mathbf{v}$.

3.1.2 Interaction Layer. The interaction layer aims at capturing the high-order correlations between embeddings via complex interaction operations for further enhancement. For example, LightGCN [15] aggregates the embeddings of neighbors via the graph message passing mechanism on user-item bipartite graph to enhance the representations of users and items. Such a process can be formulated as:

$$\hat{\mathbf{u}}, \hat{\mathbf{v}} = f_\theta(\mathbf{u}, \mathbf{v}) \tag{3}$$

where $\hat{\mathbf{u}}, \hat{\mathbf{v}} \in \mathbb{R}^d$ denote the final representations of the user and the item, θ represents the parameters of interaction layer. Also, this interaction layer can be as simple as identity mapping, which means $\hat{\mathbf{u}} = \mathbf{u}$ and $\hat{\mathbf{v}} = \mathbf{v}$, in some backbone models like matrix factorization (MF).

3.1.3 Prediction Layer. Once obtained the final representations of the brands $\hat{\mathbf{u}}$ and the regions $\hat{\mathbf{v}}$, a predictive function $f_\psi : \hat{\mathbf{u}}, \hat{\mathbf{v}} \rightarrow \mathbb{R}$ is then leveraged to get the interaction probability. ψ is the parameters of the prediction layer and this prediction layer could be any format such as inner product, cosine similarity and MLP. Here, we use the simple inner product between two representations to conduct prediction:

$$\hat{P}_{uv} = f_\psi(\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \hat{\mathbf{u}}^\top \hat{\mathbf{v}} \tag{4}$$

3.1.4 Single-City Training. Finally, for the model training, we employ the Bayesian Personalized Ranking (BPR) [38] loss as the objective for training:

$$\mathcal{L} = -\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{N}_u} \sum_{v' \notin \mathcal{N}_u} \ln \sigma(\hat{P}_{uv} - \hat{P}_{uv'}) + \lambda ||\Theta||^2 \tag{5}$$

where $\mathcal{N}_u = \{v \mid v \in \mathcal{V}, P_{uv} = 1\}$ is the set of regions that have at least one venue of brand u, σ is the sigmoid function, Θ denotes all the parameters in backbone model and λ is the coefficient of L2 regularization.

3.2 Optimal Transport Enhanced Cross-City Model

Although the distributions of brands over regions vary among different cities due to the lifestyles of people, e.g. the cafes concentrated at the city center of Singapore but dispersed throughout Chicago, there are still some commonalities between these distributions. Therefore, we aim to extract and transfer the knowledge in the distributions from a source city to a target city. Due to the limited intersection of brands and the completely disjointed regions between different cities, it is almost impossible to directly measure the correlations of brands or regions. Therefore, we propose to utilize optimal transport to effectively model the correlations for transferring the knowledge. Since the representations of brands and regions of different cities are in different spaces, we apply the Gromov-Wasserstein (GW) distance for measurement on the final embeddings.

Given the embedding representations of brands $\hat{\mathbf{u}}^s, \hat{\mathbf{u}}^t$ and regions $\hat{\mathbf{v}}^s, \hat{\mathbf{v}}^t$ in the source and target city respectively. We can obtain the transport plan $\mathbf{T}_u^{st}$ for brands and $\mathbf{T}_v^{st}$ for regions by optimizing Equation 2 where the cost measures $c_s(x, y) = c_t(x, y) = ||\mathbf{x} - \mathbf{y}||_2$ are L2 distance. Specifically, we assign the embeddings of source city $\hat{\mathbf{u}}^s, \hat{\mathbf{v}}^s$ and the target city $\hat{\mathbf{u}}^t, \hat{\mathbf{v}}^t$ to x and y in Equation 2 respectively. Then, we project the brands and the regions from the source city to the target city using the transport plan and obtain the projected brands $\hat{\mathbf{u}}^{st}$ and regions $\hat{\mathbf{v}}^{st}$ as follows:

$$\hat{\mathbf{u}}_j^{st} = \sum_{i=1}^{M_s} T_{u,ij}^{st} \cdot \hat{\mathbf{u}}_i^s \tag{6}$$

$$\hat{\mathbf{v}}_j^{st} = \sum_{i=1}^{N_s} T_{v,ij}^{st} \cdot \hat{\mathbf{v}}_i^s \tag{7}$$

where $\hat{\mathbf{u}}_i^s$ and $\hat{\mathbf{v}}_i^s$ are the embeddings of i-th brand and region in the source city s. $\hat{\mathbf{u}}_j^{st}$ and $\hat{\mathbf{v}}_j^{st}$ are the projected embeddings of j-th brand and region in the target city t with respect to the source city s. These projected embeddings in the target city are essentially a weighted sum of all the embeddings in the source city. Such a process can be seen as finding a 'counterpart' in the target city for the brand or the region of the source city.

Afterward, the recommendation results are transferred from the source city to the target city via the projected brands and regions. Thus, we have the inference recommendation $\hat{\mathbf{P}}^{st} = \{\hat{P}_{ij}^{st} | i \in M_t, j \in N_t\}$ formulated as:

$$\hat{P}_{ij}^{st} = \hat{\mathbf{u}}_i^{st\top} \hat{\mathbf{v}}_j^{st} = \sum_{k=1}^{M_s} \sum_{l=1}^{N_s} T_{u,ki}^{st} T_{v,lj}^{st} \cdot \hat{\mathbf{u}}_k^{s\top} \hat{\mathbf{v}}_l^s \tag{8}$$

Algorithm 1: Whole pipeline of OTC.

Input : Multi-City Site Recommendation Dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_K\}$

// Single-City Training

1 Train a single-city recommendation model $\mathcal{M}_k$ with dataset $\mathcal{D}_k$ for each city k via the objective in Equation 5.

2 Extract final brand embeddings $\hat{\mathbf{U}}^k = [\hat{\mathbf{u}}_1^k, \hat{\mathbf{u}}_2^k, \dots, \hat{\mathbf{u}}_{M_k}^k]$ and region embeddings $\hat{\mathbf{V}}^k = [\hat{\mathbf{v}}_1^k, \hat{\mathbf{v}}_2^k, \dots, \hat{\mathbf{v}}_{N_k}^k]$ for each city.

// Optimal Transport Plan Computation

3 Compute the OT plan of brands $\mathbf{T}_u^{st}$ and regions $\mathbf{T}_v^{st}$ for each pair of cities s, t via Equation 2.

// Cross-City Projection

4 Project brands from the source city to the target city with OT plan $\mathbf{T}_u^{st}$ via Equation 6 to obtain the projected brands $\hat{\mathbf{U}}^{st} = [\hat{\mathbf{u}}_1^{st}, \hat{\mathbf{u}}_2^{st}, \dots, \hat{\mathbf{u}}_{M_t}^{st}]$ in the target city.

5 Project regions from the source city to the target city with OT plan $\mathbf{T}_v^{st}$ via Equation 7 to obtain the projected regions $\hat{\mathbf{V}}^{st} = [\hat{\mathbf{v}}_1^{st}, \hat{\mathbf{v}}_2^{st}, \dots, \hat{\mathbf{v}}_{M_t}^{st}]$ in the target city.

// Cross-City Inference

6 Obtain the inference recommendation $\hat{\mathbf{P}}^{st}$ for each pair of cities s, t by embedding dot product between the projected brands $\hat{\mathbf{U}}^{st}$ and regions $\hat{\mathbf{V}}^{st}$ via Equation 8.

// Multi-City Fused Inference

7 Calculate the final recommendation $\bar{\mathbf{P}}^t$ for the target city by weighted sum of original recommendation $\hat{\mathbf{P}}^t$ and inference recommendation $\hat{\mathbf{P}}^{st}$ via Equation 10.

which is optimal transport-oriented and could be utilized to enhance the original recommendation $\hat{\mathbf{P}}^t$ in the target city.

The final prediction of the optimal transport enhanced cross-city model $\bar{\mathbf{P}}^t$ is obtained by weighted sum of $\hat{\mathbf{P}}^t$ and $\hat{\mathbf{P}}^{st}$:

$$\bar{P}_{ij}^t = \hat{P}_{ij}^t + \gamma^{st} \cdot \hat{P}_{ij}^{st} \tag{9}$$

where γ^{st} is the weight parameter to control the contribution of inference recommendation from the source city.

3.3 Multi-City Fused Inference

Since each source city is able to deliver beneficial knowledge to the target city, we further propose to integrate the knowledge from multiple source cities by fusing their inference recommendation. Given the set of source cities $\mathcal{S} = \{s_1, s_2, ..., s_K\}$, the overall recommendation results in the target city can be formulated as:

$$\bar{P}_{ij}^t = \hat{P}_{ij}^t + \sum_{k=1}^{K} \gamma^{s_k t} \cdot \hat{P}_{ij}^{s_k t} \tag{10}$$

where $\gamma^{s_k t}$ is the weight parameter for source city s_k and typically range in $(0, 5]$. Thus, the overall recommendation results in the target city are essentially a weighted sum of its original recommendation results $\hat{\mathbf{P}}^t$ and the inference recommendation results $\hat{\mathbf{P}}^{st}$ from all other cities. For the sake of better understanding of our OTC, we describe the whole pipeline of OTC in 1.

Table 2: Statistics of OpenSiteRec.

City	Brand	Region	Site
Chicago	969	801	8,044
New York City	2,702	2,325	14,189
Singapore	1,922	2,043	9,912
Tokyo	4,861	3,036	26,765

4 EXPERIMENTS

In this section, we will evaluate our proposed OTC on the dataset OpenSiteRec discussed above with a selected set of widely used metrics and baselines. Moreover, we will also present the significant experimental results, several analyses, and case studies to demonstrate the effectiveness and advantages of OTC.

4.1 Experimental Settings

4.1.1 Dataset. To evaluate the effectiveness of OTC, we utilize the aforementioned OpenSiteRec[1] dataset, which contains a comprehensive collection of data about four metropolises. Due to many city-specific characteristics and factors, such as urban planning, development, and culture, the distributions of data about these cities are different and imbalanced, especially the distributions of categories and brands in each city. Specifically, we split the data of each city into training, validation and test sets at the ratios of 70%, 10% and 20% respectively. The detailed statistics are shown in Table 2.

4.1.2 Evaluation Metrics. To alleviate the negative effects of the extremely imbalanced data of brands and regions in OpenSiteRec, we filter the dataset using 5-core setting on the brands to make sure that each of the remaining brands has at least 5 POIs associated with it. Additionally, 70%, 10%, and 20% of the POIs for each brand are randomly split into training, validation, and test sets respectively.

For evaluation, we choose two widely used metrics in top-n recommendations, **Recall@20** and **nDCG@20**. During evaluation, all the regions in the city that did not appear in the training set for the brand are considered as candidates, which is also known as the all-ranking protocol.

4.1.3 Baselines. We conduct experiments on the site recommendation task to compare our proposed method OTC with various representative recommendation models. Following the benchmarking experiments on OpenSiteRec, we choose the baseline models of traditional top-n recommendation tasks, such as machine learning models, collaborative filtering [40] models, and graph-based models.

For machine learning models, we choose:

- **LR** [17], i.e., logistic regression, is a popular model in classification.
- **GBDT** [11], i.e., gradient boosting decision tree, is an ensemble model with decision trees.
- **SVC** [4], i.e., support vector classifier, uses support vector machine (SVM) for classification tasks.

[1]https://OpenSiteRec.github.io/

- **RankNet** [5] is a famous neural network architecture in the field of Learning to Rank (LTR) in the recommendation area. In our experiment, a two-layer neural network is used as the backbone.

For collaborative filtering models, we choose:

- **MF-BPR** [38] is a variant of Matrix Factorization (MF) [23] with the Bayesian Personalized Ranking (BPR) loss optimizer.
- **NeuMF** [16] integrates neural network with Matrix Factorization (MF) for collaborative filtering with point-wise loss.

For graph-based models, we choose:

- **GC-MC** [42] is a common graph neural network architecture for recommendation tasks.
- **GraphRec** [9] applies graph neural network in social recommendation by aggregating the embeddings with social relationships. In our experiment, the social aggregation component is removed.
- **NGCF** [43], i.e., Neural Graph Collaborative Filtering, manipulates graph message passing on user-item interaction graph for recommendation tasks.
- **LightGCN** [15] simplifies the graph convolutional network by only using neighborhood aggregation for collaborative filtering.

Besides, we also take the cross-domain recommendation framework for comparison:

- **GWCDR** [26] is a cross-domain recommendation framework which minimizes the Gromov-Wasserstein distance of the interaction representations between the source city and the target city. Here we apply it with the same backbone model of our OTC and fuse its loss through multiple cities for fair comparison.

4.1.4 Implementation Details. The experiments are implemented on the server with an Intel Xeon Platinum 8255C CPU, a 315GB RAM and a NVIDIA Tesla V100 GPU. All the models are implemented using PyTorch 2.0.1. The model parameters are initialized with Xavier initialization and are optimized using Adam [21]. To obtain optimal performance, we perform a grid search on the combinations of hyper-parameters. The learning rate is chosen from {0.01, 0.005, 0.001, 0.0005, 0.0001} and the batch size is chosen from {16, 32, 64, 128, 256, 512, 1024, 2048, 4096}. We take the values per 0.5 within the range of $(0, 5]$ to define the best weight parameter for each city. For optimizing the optimal transport plan, we utilize the POT library[2] with square loss and 1e-09 relaxation parameter. The best epoch is selected for evaluation using an early stopping strategy on the tuning set.

4.2 Main Results

Table 3 shows all experimental results of the comparison of proposed OTC with the baseline methods mentioned above. From the experimental results, we have the following observations and conclusions:

- OTC-MF and OTC-LightGCN significantly beat all the baseline methods comparable to them, which demonstrates the effectiveness of our proposed OTC. Compared with GWCDR, which may even have negative effects on the performance of backbone models, our OTC can better capture the latent correlation across cities with limited data.

[2]https://pythonot.github.io/

Table 3: Experimental results on OpenSiteRec. The best results are highlighted in bold.

Method	Chicago		New York City		Singapore		Tokyo	
	Rec@20	nDCG@20	Rec@20	nDCG@20	Rec@20	nDCG@20	Rec@20	nDCG@20
LR	0.1203	0.0868	0.0886	0.0655	0.1784	0.1336	0.0795	0.0594
GBDT	0.1203	0.0868	0.0886	0.0655	0.1784	0.1336	0.0795	0.0594
SVC	0.1203	0.0868	0.0886	0.0655	0.1784	0.1336	0.0795	0.0594
RankNet	0.2269	0.1427	0.1224	0.0654	0.4297	0.2271	0.1213	0.0667
NeuMF	0.1942	0.1293	0.1200	0.0576	0.4289	0.2236	0.1225	0.0639
MF	0.2494	0.1465	0.1702	0.0917	0.4430	0.2351	0.1323	0.0781
GWCDR-MF	0.2518	0.1484	0.1688	0.0896	0.4397	0.2324	0.1329	0.0788
OTC-MF	**0.2669***	**0.1599***	**0.1766***	**0.0981***	**0.4574***	**0.2503***	**0.1372***	**0.0796***
%Improv. to MF	7.02%	9.15%	3.76%	6.98%	3.25%	6.45%	3.70%	1.92%
GC-MC	0.2332	0.1657	0.1514	0.0513	0.4685	0.2317	0.1558	0.0884
GraphRec	0.2365	0.1640	0.1538	0.0550	0.4697	0.2293	0.1594	0.0905
NGCF	0.2866	0.1838	0.1920	0.1102	0.4929	0.2674	0.1619	0.1012
LightGCN	0.2875	0.1902	0.2087	0.1088	0.5013	0.2745	0.1751	0.1068
GWCDR-LightGCN	0.2818	0.1845	0.2056	0.1077	0.4863	0.2677	0.1716	0.1052
OTC-LightGCN	**0.2977***	**0.2016***	**0.2108**	**0.1109***	**0.5078**	**0.2831***	**0.1780***	**0.1100***
%Improv. to LightGCN	3.55%	5.99%	1.01%	1.93%	1.30%	3.13%	1.66%	3.00%

"*" indicates the statistically significant improvements (i.e., two-sided t-test with $p < 0.01$) over the best baseline.

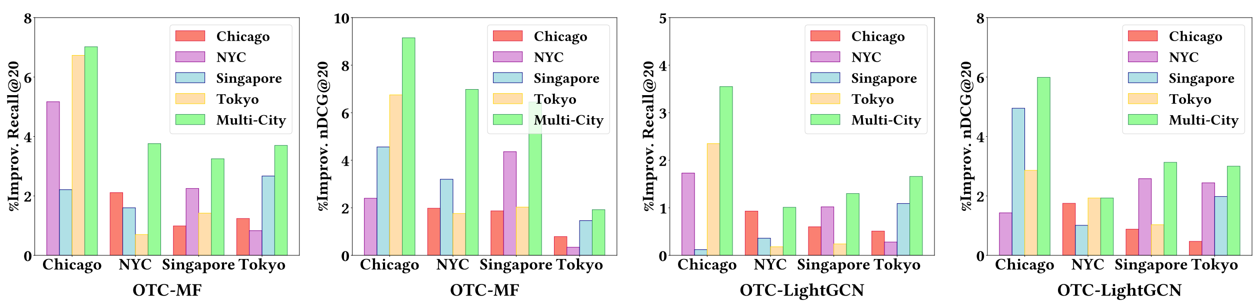

Figure 3: Ablation study with different source cities on the relative improvements.

- The results of LR, GBDT, and SVM are exactly the same. Since the data is extremely limited and the inputs of these traditional machine learning methods are high-dimensional one-hot vectors, they will easily overfit the training dataset and degenerate into the same linear classifier with the same results. Such a phenomenon also indicates the seriousness of data sparsity problem.
- Graph-based methods are usually better than machine learning methods and collaborative filtering methods. It is not surprising to get this observation of their strong ability to capture collaborative signals and alleviate the data sparsity problem.
- The performance gains are greater on Chicago and less on Tokyo. This phenomenon is owing to the different sparsity level of data, i.e., the data of Chicago is sparser while the data of Tokyo is denser. Since our proposed OTC aims at improving the recommendation performance by alleviating the data sparsity problem via cross-city information, the employment of OTC can bring greater improvements to sparser cities.
- The relative improvements by incorporating OTC is 5.3% for MF and 2.7% for LightGCN on average. This is because LightGCN is more robust than MF on sparse data by leveraging high-order interactions while our proposed OTC also aims to alleviate the data sparsity problem.
- The effects on nDCG metric are more significant than on recall metric. Such an observation shows that OTC is able to achieve more accurate recommendation and improve the quality of the whole ranking list.

4.3 Ablation Study

To better illustrate the effects of cross-city information, we further perform an ablation study on the relative improvements for evaluation. Specifically, we report the performance comparisons of OTC-MF and OTC-LightGCN with different source cities and with all three other cities on all these four cities. As illustrated in Figure 3, each source city will contribute to the recommendation performance on the target city at different extents. However, we could also find some interesting points:

Table 4: Efficiency analysis of OTC on the training time and the inference time with MF or LightGCN as the backbone models.

Method	Chicago		New York City		Singapore		Tokyo	
	Training	Inference	Training	Inference	Training	Inference	Training	Inference
MF	47s	0.3s	2min53s	2.4s	1min43s	1.5s	5min41s	5.9s
OTC-MF	+3.1s	+2.1s	+7.1s	+6s	+5.5s	+5.3s	+9.2s	+6.9s
LightGCN	10min32s	1.1s	24min7s	7.8s	13min58s	5.1s	35min37s	20.4s
OTC-LightGCN	+3.1s	+2.1s	+7.1s	+6s	+5.5s	+5.3s	+9.2s	+6.9s

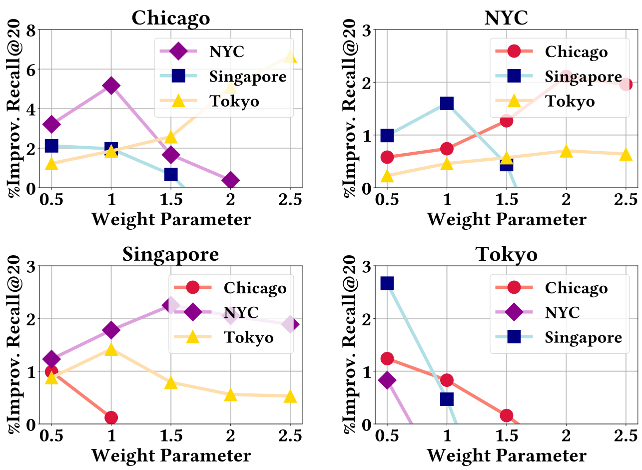

Figure 4: Parameter analysis of different weight parameter γ for OTC-MF on relative improvements of Recall@20.

- Tokyo can usually bring more improvements to Chicago, which implies the significant informativeness of its data. This may be because the data of Tokyo is much denser and thus the transferred knowledge is more robust.
- NYC can benefit more from the information of Chicago, which may owing to the similarities of urban planning in the same country.
- Singapore has the largest improvements with NYC as the source city. This indicates that Singapore may share some similarity with NYC in lifestyles to affect the brand distribution.
- Tokyo benefits the most from Singapore, which may caused by the large intersection of brands between Singapore and Tokyo.

4.4 Efficiency Analysis

Besides the effectiveness, our OTC is also efficient with the light-weighted and decoupling design of cross-city module. To better prove the high efficiency of OTC, we provide an efficiency analysis on a NVIDIA Tesla V100 GPU. As shown in Table 4, the additional time cost brought by OTC in both training and inference stages is marginal. This is because the computation of OT plan is very fast and the embeddings from source cities could be computed and stored previously. Meanwhile, the computation of OT plan is only related to the size of embeddings and is independent to the model structure. Thus, regardless the complexity of the model structure and the time of the training process, the additional time cost brought by OTC is the same for all backbone models.

4.5 Parameter Analysis

Although our proposed OTC is effective in improving the overall performance of site recommendation, it will be affected largely by the weight parameter γ. For better illustration, we plot the changing curves of relative improvements of different source cities with different weight parameters on OTC-MF. As shown in Figure 4, an appropriate value of the weight parameter is crucial to get desirable results. When the weight parameter is too large, which means the final predictions are more dominated by the cross-city information, OTC may even have a negative impact on the results as illustrated by the curves below the horizontal axis.

4.6 Case Study

Here we present a case study of three brands' transport plans from Tokyo to NYC to better support our statement of leveraging cross-city information to improve site recommendation performance in Table 5. Although the brand distribution varies from different cities for different lifestyles of people, it is still amazing to find that even the categories of the top assigned brands in the source city may be different from the given brand in the target city. For example, *Five Guys* is a fast food brand in NYC but the corresponding brands in Tokyo are mostly brands for alcoholic beverages. Similarly, *Dunkin' Donuts* is a dessert shop in NYC but the corresponding brands in Tokyo are all supermarkets and all have very high assignment weights. However, this phenomenon may suggest some ideas in business that the business strategy for some categories of brands is more beneficial to some other categories of brands in other cities. That is, those regions that are similar to the regions having more bars and supermarkets in Tokyo are more likely to have *Five Guys* and *Dunkin' Donuts* in NYC, respectively.

5 RELATED WORKS

In this section, we will introduce the related works of OTC on both task (site recommendation) and technique (optimal transport for recommendation) and discuss the differences.

5.1 Site Recommendation

Site recommendation aims at predicting the optimal location for brands and has significant practical value in urban computing [37, 56, 57, 60], and business strategy [3, 50, 54, 55]. Early attempts [18, 19] focus on investigating the important factors for site recommendation by data analysis, such as street centrality [35, 36]. With the

Table 5: Case study of three brands' transport plans from Tokyo to NYC. Specifically, *Five Guys* is a fast food brand, *Dunkin' Donuts* is a dessert shop brand and *Lululemon* is a sportswear brand. *HUB* and *Warawara* are alcoholic beverage shop brands. *Starbucks* is a cafe brand. *maruetsu petit*, *Inageya* and *Tokyu Store* are all supermarket brands. *Zara* and *H&M* are fast fashion brands. *Daigoku Drug* is a pharmacy and cosmetics store brand.

Target City (NYC) Brand	Source City (Tokyo) Brand			
Five Guys	HUB (0.28)	Warawara (0.17)	Starbucks (0.09)	...
Dunkin' Donuts	maruetsu petit (0.32)	Inageya (0.27)	Tokyu Store (0.21)	...
Lululemon	Zara (0.19)	H&M (0.12)	Daigoku Drug (0.09)	...

booming development of machine learning algorithms, many approaches have emerged to utilize various machine learning models to directly predict the potential locations []. Geo-Spotting [20] first utilizes machine learning models to explore the effectiveness of geographical and mobility features for site recommendation of 3 popular fast food brands in New York City. Similarly, ANNRR [6] and PAM [28] pay efforts to determine the locations for new bike sharing stations and ambulance stations using heterogeneous urban data. Some other methods employ more advanced models to better model the various features. For example, BL-G-CoSVD [52] incorporates geographical and commercial features into SVD model to recommend the shop type for a given region. DD3S [49] learns to rank the locations for 2 coffee shop and 2 chain hotel brands while DeepStore [32] utilizes a deep neural network to learn from a variety of features. O^2-SiteRec [51] considers the order number and delivery time from the courier capacity perspective for the site recommendation of Online-to-Offline (O2O) stores on delivery platforms. Meanwhile, there are also many methods focusing on leveraging the knowledge across different cities. Specifically, CityTransfer [14] applies both inter- and intra-city views of the association model to transfer knowledge and improve the site recommendation performance of 3 chain hotel brands in the target city. WANT [33] diminishes the distribution discrepancies to predict the consumption of store brands via adversarial learning.

Compared with them, our proposed OTC is able to achieve site recommendation without rich descriptive features. Therefore, our OTC can support thousands of brands on an extensive scale while the existing works can only handle very limited dozens of brands.

5.2 Optimal Transport for Recommendation

Optimal Transport (OT) has attracted increasing interest for its successes in other areas [2, 12, 13, 48, 58] and has recently been introduced into recommendation area. These methods mainly focus on aligning the distributions between different domains to conduct cross-domain recommendation. For example, WCF [34] predicts users' preference on cold-start items by minimizing the Wasserstein distance. DisAlign [30] employs Stein discrepancy on the item representations while GWCDR [26] applies Gromov-Wasserstein discrepancy on the interaction representations for distribution alignment to achieve cross-domain recommendation. MStein [10] measures mutual information between different item sequences using Wasserstein distance for sequential recommendation.

Unlike all these existing methods that leverage optimal transport to align the distributions for representation learning, our proposed OTC directly enhanced the recommendation results by transferring the brands and regions via the transport plan. In the situation of site recommendation that has extremely limited data, our solution to directly obtain recommendation results is much more robust than distribution alignment for the poor reliability of the distributions with insufficient data.

6 CONCLUSION

Site recommendation is a special type of recommendation task and plays an important role in assisting decision-making of modern business. The extremely limited data of real-world aspects has presented a serious data sparsity problem. Existing methods mainly avoid the data sparsity problem by only focusing on a few popular brands with descriptive information to manually design heuristic features. However, such methods are just escaping from the problem thereby making them unable to be applied in practice and create social good so far. Hence, we argue that an effective and efficient way to support site recommendation for all the brands is necessary. In this paper, we propose a novel optimal transport enhanced cross-city framework OTC, to leverage the cross-city information for further improvements. Specifically, OTC projects the brands and the regions from the source city into the target city via transport plans and then obtains the inference recommendation accordingly. The recommendation results in a given city are enhanced by incorporating the inference recommendations from the other cities. The experimental results on the real-world dataset strongly demonstrate the effectiveness of OTC in improving the performance of backbone models. Meanwhile, the detailed analyses of associations between different cities also conclude some valuable ideas for business success.

ACKNOWLEDGEMENTS

This research was partially supported by Research Impact Fund (No.R1015-23), APRC - CityU New Research Initiatives (No.9610565, Start-up Grant for New Faculty of CityU), CityU - HKIDS Early Career Research Grant (No.9360163), Hong Kong ITC Innovation and Technology Fund Midstream Research Programme for Universities Project (No.ITS/034/22MS), Hong Kong Environmental and Conservation Fund (No.88/2022), and SIRG - CityU Strategic Interdisciplinary Research Grant (No.7020046, No.7020074), Ant Group (CCF-Ant Research Fund, Ant Group Research Fund), Huawei (Huawei Innovation Research Program), Tencent (CCF-Tencent Open Fund, Tencent Rhino-Bird Focused Research Program), CCF-BaiChuan-Ebtech Foundation Model Fund, and Kuaishou.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. Knowl. Data Eng.* 17, 6 (2005), 734–749.

[2] David Alvarez-Melis and Tommi S. Jaakkola. 2018. Gromov-Wasserstein Alignment of Word Embedding Spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018.* Association for Computational Linguistics, 1881–1890.

[3] Oded Berman and Dmitry Krass. 2002. The generalized maximal covering location problem. *Comput. Oper. Res.* 29, 6 (2002), 563–581.

[4] Erin J. Bredensteiner and Kristin P. Bennett. 1999. Multicategory Classification by Support Vector Machines. *Comput. Optim. Appl.* 12, 1-3 (1999), 53–79.

[5] Christopher J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory N. Hullender. 2005. Learning to rank using gradient descent. In *Machine Learning, Proceedings of the Twenty-Second International Conference (ICML 2005), Bonn, Germany, August 7-11, 2005 (ACM International Conference Proceeding Series)*, Vol. 119. ACM, 89–96.

[6] Longbiao Chen, Daqing Zhang, Gang Pan, Xiaojuan Ma, Dingqi Yang, Kostadin Kushlev, Wangsheng Zhang, and Shijian Li. 2015. Bike sharing station placement leveraging heterogeneous urban open data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2015, Osaka, Japan, September 7-11, 2015.* ACM, 571–575.

[7] Richard L. Church and Alan T. Murray. 2008. Business Site Selection, Location Analysis and GIS.

[8] Mukund Deshpande and George Karypis. 2004. Item-based top-*N* recommendation algorithms. *ACM Trans. Inf. Syst.* 22, 1 (2004), 143–177.

[9] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Yihong Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019.* ACM, 417–426.

[10] Ziwei Fan, Zhiwei Liu, Hao Peng, and Philip S. Yu. 2023. Mutual Wasserstein Discrepancy Minimization for Sequential Recommendation. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023.* ACM, 1375–1385.

[11] Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29 (2001), 1189–1232.

[12] Ji Gao, Xiao Huang, and Jundong Li. 2021. Unsupervised Graph Alignment with Wasserstein Distance Discriminator. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021.* ACM, 426–435.

[13] Jingtong Gao, Xiangyu Zhao, Bo Chen, Fan Yan, Huifeng Guo, and Ruiming Tang. 2023. AutoTransfer: Instance Transfer for Cross-Domain Recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023.* ACM, 1478–1487.

[14] Bin Guo, Jing Li, Vincent W. Zheng, Zhu Wang, and Zhiwen Yu. 2017. CityTransfer: Transferring Inter- and Intra-City Knowledge for Chain Store Site Recommendation based on Multi-Source Urban Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4 (2017), 135:1–135:23.

[15] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020.* ACM, 639–648.

[16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017.* ACM, 173–182.

[17] David W. Hosmer and Stanley Lemeshow. 1991. Applied Logistic Regression.

[18] Pablo Jensen. 2006. Network-based predictions of retail store commercial categories and optimal locations. *Physical review. E, Statistical, nonlinear, and soft matter physics* 74 3 Pt 2 (2006), 035101.

[19] Pablo Jensen. 2009. Analyzing the Localization of Retail Stores with Complex Systems Tools. In *Advances in Intelligent Data Analysis VIII, 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31 - September 2, 2009. Proceedings (Lecture Notes in Computer Science)*, Vol. 5772. Springer, 10–20.

[20] Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, and Cecilia Mascolo. 2013. Geo-spotting: mining online location-based services for optimal retail store placement. In *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013.* ACM, 793–801.

[21] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.*

[22] Joseph A. Konstan, Bradley N. Miller, David A. Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. 1997. GroupLens: Applying Collaborative Filtering to Usenet News. *Commun. ACM* 40, 3 (1997), 77–87.

[23] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37.

[24] Prof Vikas Kumar and Kiran Karande. 2000. The Effect of Retail Store Environment on Retailer Performance. *Journal of Business Research* 49 (2000), 167–181.

[25] Xinhang Li, Zhaopeng Qiu, Jiacheng Jiang, Yong Zhang, Chunxiao Xing, and Xian Wu. 2024. Conditional Cross-Platform User Engagement Prediction. *ACM Trans. Inf. Syst.* 42, 1 (2024), 6:1–6:28.

[26] Xinhang Li, Zhaopeng Qiu, Xiangyu Zhao, Zihao Wang, Yong Zhang, Chunxiao Xing, and Xian Wu. 2022. Gromov-Wasserstein Guided Representation Learning for Cross-Domain Recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022.* ACM, 1199–1208.

[27] Xinhang Li, Xiangyu Zhao, Yejing Wang, Yu Liu, Yong Li, Cheng Long, Yong Zhang, and Chunxiao Xing. 2023. OpenSiteRec: An Open Dataset for Site Recommendation. *CoRR* abs/2307.00856 (2023).

[28] Yuhong Li, Yu Zheng, Shenggong Ji, Wenjun Wang, Leong Hou U, and Zhiguo Gong. 2015. Location selection for ambulance stations: a data-driven approach. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015.* ACM, 85:1–85:4.

[29] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Comput.* 7, 1 (2003), 76–80.

[30] Weiming Liu, Jiajie Su, Chaochao Chen, and Xiaolin Zheng. 2021. Leveraging Distribution Alignment via Stein Path for Cross-Domain Cold-Start Recommendation. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual.* 19223–19234.

[31] Yu Liu, Jingtao Ding, and Yong Li. 2022. Developing knowledge graph based system for urban computing. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs* (2022).

[32] Yan Liu, Bin Guo, Nuo Li, Jing Zhang, Jingmin Chen, Daqing Zhang, Yinxiao Liu, Zhiwen Yu, Sizhe Zhang, and Lina Yao. 2019. DeepStore: An Interaction-Aware Wide&Deep Model for Store Site Recommendation With Attentional Spatial Embeddings. *IEEE Internet Things J.* 6, 4 (2019), 7319–7333.

[33] Yan Liu, Bin Guo, Daqing Zhang, Djamal Zeghlache, Jingmin Chen, Ke Hu, Sizhe Zhang, Dan Zhou, and Zhiwen Yu. 2021. Knowledge Transfer with Weighted Adversarial Network for Cold-Start Store Site Recommendation. *ACM Trans. Knowl. Discov. Data* 15, 3 (2021), 47:1–47:27.

[34] Yitong Meng, Xiao Yan, Weiwen Liu, Huanhuan Wu, and James Cheng. 2020. Wasserstein Collaborative Filtering for Item Cold-start Recommendation. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020.* ACM, 318–322.

[35] Sergio Porta, Vito Latora, Fahui Wang, Salvador Rueda, Emanuele Strano, Salvatore Scellato, Allessio Cardillo, Eugenio Belli, Francisco Cardenas, Berta Cormenzana, and Laura Latora. 2012. Street Centrality and the Location of Economic Activities in Barcelona. *Urban Studies* 49 (2012), 1471 – 1488.

[36] Sergio Porta, Emanuele Strano, Valentino Iacoviello, Roberto Messora, Vito Latora, Alessio Cardillo, Fahui Wang, and Salvatore Scellato. 2009. Street Centrality and Densities of Retail and Services in Bologna, Italy. *Environment and Planning B: Planning and Design* 36 (2009), 450 – 465.

[37] M. Mazhar Rathore, Awais Ahmad, Anand Paul, and Seungmin Rho. 2016. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Comput. Networks* 101 (2016), 63–80.

[38] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009.* AUAI Press, 452–461.

[39] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *CSCW '94, Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, NC, USA, October 22-26, 1994.* ACM, 175–186.

[40] Badrul Munir Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the Tenth International World Wide Web Conference, WWW 10, Hong Kong, China, May 1-5, 2001.* ACM, 285–295.

[41] Meng Shao, Zhi Han, Jin Wei Sun, Chengsi Xiao, Shu lei Zhang, and Yuan xu Zhao. 2020. A review of multi-criteria decision making applications for renewable energy site selection. *Renewable Energy* 157 (2020), 377–403.

[42] Rianne van den Berg, Thomas N. Kipf, and Max Welling. 2017. Graph Convolutional Matrix Completion. *CoRR* abs/1706.02263 (2017).

[43] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019.* ACM, 165–174.

[44] Yejing Wang, Xiangyu Zhao, Tong Xu, and Xian Wu. 2022. AutoField: Automating Feature Selection in Deep Recommender Systems. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022.* ACM, 1977–1986.

[45] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: Neural News Recommendation with Personalized Attention. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*. ACM, 2576–2584.

[46] Yunna Wu, Chao Xie, Chuanbo Xu, and Fang Li. 2017. A Decision Framework for Electric Vehicle Charging Station Site Selection for Residential Communities under an Intuitionistic Fuzzy Environment: A Case of Beijing. *Energies* 10 (2017), 1270.

[47] Yunna Wu, Meng Yang, Haobo Zhang, Kaifeng Chen, and Yang Wang. 2016. Optimal Site Selection of Electric Vehicle Charging Stations Based on a Cloud Model and the PROMETHEE Method. *Energies* 9 (2016), 1–20.

[48] Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin. 2019. Gromov-Wasserstein Learning for Graph Matching and Node Embedding. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research)*, Vol. 97. PMLR, 6932–6941.

[49] Mengwen Xu, Tianyi Wang, Zhengwei Wu, Jingbo Zhou, Jian Li, and Haishan Wu. 2016. Demand driven store site selection via multiple spatial-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2016, Burlingame, California, USA, October 31 - November 3, 2016*. ACM, 40:1–40:10.

[50] Tong Xu, Hengshu Zhu, Xiangyu Zhao, Qi Liu, Hao Zhong, Enhong Chen, and Hui Xiong. 2016. Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. ACM, 1285–1294.

[51] Hua Yan, Shuai Wang, Yu Yang, Baoshen Guo, Tian He, and Desheng Zhang. 2022. O^{2}-SiteRec: Store Site Recommendation under the O2O Model via Multi-graph Attention Networks. In *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE, 525–538.

[52] Zhiwen Yu, Miao Tian, Zhu Wang, Bin Guo, and Tao Mei. 2016. Shop-Type Recommendation Leveraging the Data from Social Media and Location-Based Services. *ACM Trans. Knowl. Discov. Data* 11, 1 (2016), 1:1–1:21.

[53] Zheng Yuan, Fajie Yuan, Yu Song, Youhua Li, Junchen Fu, Fei Yang, Yunzhu Pan, and Yongxin Ni. 2023. Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*. ACM, 2639–2649.

[54] Zijian Zhang, Ze Huang, Zhiwei Hu, Xiangyu Zhao, Wanyu Wang, Zitao Liu, Junbo Zhang, S. Joe Qin, and Hongwei Zhao. 2023. MLPST: MLP is All You Need for Spatio-Temporal Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*. ACM, 3381–3390.

[55] Zijian Zhang, Xiangyu Zhao, Hao Miao, Chunxu Zhang, Hongwei Zhao, and Junbo Zhang. 2023. AutoSTL: Automated Spatio-Temporal Multi-Task Learning. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*. AAAI Press, 4902–4910.

[56] Xiangyu Zhao, Wenqi Fan, Hui Liu, and Jiliang Tang. 2022. Multi-Type Urban Crime Prediction. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 4388–4396.

[57] Xiangyu Zhao and Jiliang Tang. 2017. Modeling Temporal-Spatial Correlations for Crime Prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*. ACM, 497–506.

[58] Xu Zhao, Zihao Wang, Yong Zhang, and Hao Wu. 2020. A Relaxed Matching Procedure for Unsupervised BLI. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*. Association for Computational Linguistics, 3036–3041.

[59] Zhi Zheng, Zhaopeng Qiu, Tong Xu, Xian Wu, Xiangyu Zhao, Enhong Chen, and Hui Xiong. 2022. CBR: Context Bias aware Recommendation for Debiasing User Modeling and Click Prediction10033. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*. ACM, 2268–2276.

[60] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. 2023. DiffTraj: Generating GPS Trajectory with Diffusion Probabilistic Model. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

[61] Tezcan Kaşmer Şahin, Saffet Ocak, and Mehmet Top. 2019. Analytic hierarchy process for hospital site selection. *Health Policy and Technology* (2019).