# CityTransfer: Transferring Inter- and Intra-City Knowledge for Chain Store Site Recommendation based on Multi-Source Urban Data

BIN GUO, Northwestern Polytechnical University
JING LI, Northwestern Polytechnical University
VINCENT W. ZHENG, UIUC, Advanced Digital Sciences Center
ZHU WANG, Northwestern Polytechnical University
ZHIWEN YU, Northwestern Polytechnical University

Chain businesses have been dominating the market in many parts of the world. It is important to identify the optimal locations for a new chain store. Recently, numerous studies have been done on chain store location recommendation. These studies typically learn a model based on the features of existing chain stores in the city and then predict what other sites are suitable for running a new one. However, these models do not work when a chain enterprise wants to open business in a new city where there is not enough data about this chain store. To solve the cold-start problem, we propose *CityTransfer*, which transfers chain store knowledge from semantically-relevant domains (e.g., other cities with rich knowledge, similar chain enterprises in the target city) for chain store placement recommendation in a new city. In particular, *CityTransfer* is a two-fold knowledge transfer framework based on collaborative filtering, which consists of the transfer rating prediction model, the inter-city knowledge association method and the intra-city semantic extraction method. Experiments using data of chain hotels from four different cities crawled from Ctrip (a popular travel reservation website in China) and the urban characters extracted from several other data sources validate the effectiveness of our approach on store site recommendation.

CCS Concepts: • **Information System Applications: Miscellaneous**

## KEYWORDS

Chain Store Site Recommendation, Urban Computing, Knowledge Transfer, Collaborative Filtering, Recommendation

**ACM Reference format:**

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, Article 135, (December 2017) 22 pages[1].
DOI: 10.1145/3161411

This work is supported by the National Basic Research Program of China (grant no.2015CB352400), and the National Natural Science Foundation of China (grant nos. 61772428, 61332005, 61373119, and 61402369).

Author's addresses: B. Guo (e-mail: guob@nwpu.edu.cn), No. 127, Youyi-West Rd., Xi'an 710072, China; J. Li, No. 127, Youyi-West Rd., Xi'an 710072, China; V.W. Zheng, Advanced Digital Sciences Center (ADSC), Singapore; Z. Wang, No. 127, Youyi-West Rd., Xi'an 710072, China; Z. Yu, No. 127, Youyi-West Rd., Xi'an 710072, China.

## 1 INTRODUCTION

Chain businesses are dominating the market in modern cities, which usually locate in different areas of the city and can enhance the provision of high-quality and beneficial services for the citizens. One of the critical issues in chain business is to select locations for new outlets/stores of a chain enterprise, which may not only affect the chain store's profit, but also impact the future development of the chain enterprise [5]. Thereby, an effective chain store recommendation system becomes necessary to chain enterprise managers.

Traditionally, to make valuable suggestions, operators need to carry out questionnaire surveys to understand people's needs and make detailed investigation to all the candidate places to learn their characteristics, e.g., traffic convenience, human flow, etc. Obviously, these traditional methods are time-consuming and cannot adapt to the rapid development of the city automatically.

Recently, the proliferation of big data in cities has fostered unprecedented opportunities to deliver intelligent urban services by leveraging data mining and machine learning techniques [14, 42]. Researchers thus resort to the data-driven approach to address the store site recommendation problem in urban environments. For example, quite a number of studies [8,18,19,26,27,39] explore the data from location-based social networks (LBSNs) to predict business locations [20,28]. Most of them typically learn a regression model based on the features of existing chain stores in the same city. However, when a chain enterprise plans to expand its market in a new city, it usually lacks enough data which can be directly referenced, known as the cold-start problem [6,24]. Existing works fail to address this problem. In this work, we aim to address this practical and interesting issue to learn a chain store recommendation system for the given chain enterprise in the target city by using a data-driven approach.

To the best of our knowledge, the above mentioned issue has not been explored in existing studies. Taking insight of this problem, we find that though we do not have knowledge of the target chain enterprise in the target city, we do have information about its running in other cities. This endows us knowledge about the location bias of the chain enterprise. Furthermore, in the target city, there often exist similar chain enterprises, which can provide knowledge about the local features of local enterprises. In both cases, we can have rich 'foreign' knowledge which can be transferred to enable chain store recommendation. We use the following scenario to illustrate this.

Supposing that Home Inn (a popular economy hotel in China) would like to expand its market in a 'new' city Xi'an (a second-tier city in China). There exists no Home Inn in Xi'an, but there are many Home Inn sites in Beijing (a first-tier city in China). Furthermore, there are many stores of other similar economy hotel enterprises, such as 7 Days Inn and Hanting Inn in Xi'an. An interesting question arises: can we use the knowledge learned from Home Inn in Beijing and other economy hotel enterprises in Xi'an to identify the optimal sites for Home Inn startups in Xi'an?

When transferring knowledge from other cities, there exists a problem that the training and testing data may have different feature and rating distribution [37], which cannot work well using traditional machine learning methods. However, transfer learning (TL) [30] has been proved to be an effective approach for knowledge transferring from different domains. For example, in the field of ubiquitous computing, it has been successfully used for human activity recognition (HAR) [38], air quality prediction [37], human mobility [11], and so on. However, existing studies of transfer learning only address one-fold knowledge transfer. They do not support the transferring of two-fold knowledge (inter-city and intra-city) as investigated in the above scenario. From another perspective, the problem to be solved in the above scenario is a typical recommendation problem. Among the numerous recommendation methods, collaborating filtering (CF) [6,13,22,35] has shown high performance in many recommendation systems. While traditional CF cannot be directly used to solve the above two-fold knowledge transfer problem, it has been proved easily-extensible to address complicated recommendation problems [17,22].

In this paper, we combine collaborating filtering and knowledge transfer to address the two-fold knowledge transfer problem. This is not trivial and we are faced with at least the following challenges.

1) How to transfer target chain enterprise knowledge and target city knowledge to deal with the cold-start problem in chain store site recommendation?

2) Data from different cities normally have different distributions in feature and rating spaces. Therefore, we cannot directly compare places located in two different cities, which is the second challenge.

3) Local features extracted from multi-source datasets might be redundant and noisy, which can affect the performance on knowledge transfer and chain store site recommendation.

To address these challenges, we propose *CityTransfer*, a two-fold knowledge transfer framework based on collaborating filtering. It supports chain store site recommendation in cold-start situations by leveraging both inter-city and intra-city knowledge transfer. In particular, our contributions are as follows.

1) We present a new and practical 'cold-start' problem in optimal site recommendation. To address this problem, we propose a two-fold knowledge transfer framework, which significantly extends the SVD (Singular Value Decomposition)-based CF model to make it possible to transfer the enterprise bias knowledge from other cities, and also learn local characterizing features from similar enterprises in the same city.

2) To bridge the discrepancy on feature/rating distribution between the source city and the target city, we divide each city into same-sized location grids (e.g., 500m*500m) and build correspondence between the grids by measuring the Pearson correlation coefficient over their features [1,3]. It is then used in an optimization function to generate a shared semantic space for both cities and make the location grids comparable between them.

3) To ensure the features of each location grid preserve high-quality semantics, we use AutoEncoder [4,10,29] to construct original features of each location grid, which makes the new features become more robust and informative and further help improve the quality of chain store recommendation.

We choose chain hotel enterprises as a case study to evaluate the performance of our framework. Multi-source data of four different cities in China are collected for experiments, including chain hotel enterprise data from Ctrip [2], check-in data from Sina Weibo[3], and Point of Interests (POIs) data from Gaode-Map[4]. Experimental results indicate that our approach outperforms baselines from many aspects. We also obtain some findings that can be used to steer better knowledge transfer among cities.

## 2 RELATED WORK

Our work is closely related to the following three areas: recommendation systems, optimal site recommendation, and knowledge transfer in ubiquitous computing, as summarized below.

### 2.1 Recommendation Systems

Recommendation systems have become increasingly popular in recent years, which have been utilized in a variety of application areas. Traditional recommendation systems could be distinguished into three different approaches [9]: rule-based filtering, content-based filtering, and collaborative filtering.

*Rule-based filtering* [15] creates a user-specific utility function and then applies it to the candidate items. It is easy to fail due to the burden on user configuration. *Content-based filtering* is based on item contents and user preferences [31]. It tries to recommend items that are similar to those that a user preferred in the past. *Collaborative filtering* (CF) [34] uses the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users. A key advantage of the CF approach is that it does not rely on machine analyzable content. It has become one of the most successful and widely-used approaches to build recommender systems.

---

[2] http://hotels.ctrip.com/
[3] http://open.weibo.com/wiki/
[4] http://lbs.amap.com/

There are generally two ways to build CF systems: *memory-based methods* and *model-based methods* [34]. Early CF systems, such as GroupLens [21], use the user-rating data to calculate the similarity between users or items and leverage them to make prediction/recommendation. However, there are several limitations, such as the fact that the similarity values are based on common items and therefore are unreliable when data are sparse. To overcome the shortcoming of memory-based CF algorithms, model-based CF approaches are investigated. It uses the rating data to estimate or learn a model to make predictions. Well-known model-based CF techniques include Bayesian belief nets (BNs) CF models [34], clustering CF models [36], and Singular Value Decomposition (SVD) CF models [22]. However, such pure collaborative filtering cannot help in *cold-start* settings [33], where new users or new items exist.

To address the cold-start problem in CF, additional information about users (user attributes, e.g. gender, age, geographical location, occupation) and items (item attributes, e.g. genres, product categories, keywords) should be used. Several *hybrid recommender systems* have been developed to tackle the cold-start problem by combining two or more recommendation techniques. For example, Schein *et al.* [33] propose the aspect model latent variable method for cold-start recommendation, which combines both collaborative and content information in model fitting. Based on all features in user and content profiles, Chu and Park [9] develop predictive bilinear regression models to provide accurate personalized recommendations of new items. Gantner *et al.* [12] describe a method that maps entity (e.g. new user or item) attributes to the latent features of a matrix factorization model. With such mappings, the factors of a matrix factorization model trained by standard techniques can be applied to the new-user and the new-item problem.

Our work is inspired by the CF-based recommendation method. In particular, we view the optimal site recommendation as the multi-user, multi-item recommendation problem, where chain enterprises are viewed as 'users' and location grids are viewed as 'items'. In the target city, we face the 'cold-start' problem for recommending sites for a new enterprise. Different from existing solutions for cold-start problems in CF, which combine both collaborative and content information for model fitting, we devote to transferring the knowledge of the target enterprise in a source city (where there are ratings for the target enterprise). However, due to the discrepancy between the source and target cities, the knowledge cannot be directly used in the traditional CF model. In our work, we significantly extend the SVD-based CF model to make it possible to transfer knowledge of the same enterprise from other cities, and also learn local characterizing features from similar enterprises in the same city. To achieve this, we build the transfer rating prediction model and use the inter-city knowledge association method to make the location grids in different cities comparable, i.e., representing them in the shared semantic space which can then be used in the prediction model for the target city.

## 2.2 Optimal Site Recommendation

Numerous studies [8,18,19,26,27,39] have been done on the recommendation of optimal sites for retail stores or public facilities. Most of them devote to building feature-based learning models, which first extract features from different sources, and then learn a regression model to recommend locations for the chain store in the same city. For example, Jensen [18] studies the diversity of retail store location patterns and investigates various features that may affect chain store performance. Geo-Spotting [19] investigates the predictive power of geographical features and mobility features on the popularity of retail stores in the city. Chen *et al.* [8] formulate the bike station placement issue as a bike trip demand prediction problem and propose a semi-supervised feature selection method to extract customized features from the highly variant, heterogeneous urban open data to predict bike trip demand. ChainRec [26] proposes a supervised regression method for chain store placement recommendation considering its scale, where three types of features are used, including geographic features, commercial features, and scale-relevant features. Yu *et al.* [39] treat the shop type recommendation problem as a traditional recommendation task, and propose the bias-learning matrix factorization method for shop popularity prediction.

Our work differs from these existing studies in the following two aspects. First, the problem definition is different. The above studies give recommendations by exploring sufficient labelled training data in the same city, while we aim to address the site recommendation problem for an enterprise with no labelled data in the target city, i.e., the cold-start problem. Second, while there are no-labelled data in the target city, we want to leverage two-fold knowledge in the prediction model: *local features* that can be learned from similar enterprises in the same city, and *enterprise bias* that can be learned from other cities with rich data. To achieve this, we propose the two-fold knowledge transfer model founded on the matrix factorization-based CF model, and introduce the inter-city knowledge aggregation and intra-city semantic extraction methods to enable high-quality knowledge transfer.

### 2.3 Knowledge Transfer in Ubiquitous Computing

A major assumption in many data mining and machine learning algorithms is that the training and future data must be in the same feature space and have the same distribution. However, in many real-world applications, this assumption does not hold. For example, we sometimes have a classification task in one domain of interest, but we only have sufficient training data in another domain, where the latter data may be in a different feature space or follow a different data distribution. In such cases, knowledge transfer becomes important that can greatly improve the learning performance by avoiding expensive data labeling efforts. Transfer learning has emerged as a promising learning framework to address this problem. It tries to transfer the knowledge from some *source* tasks to the *target* task when the latter does not have enough training data [30].

There have been recently a few studies that leverage transfer learning to address the no-training-data issue in the ubiquitous computing field [11,37,38,41]. Earlier studies focus on indoor localization and activity recognition. For example, Zheng *et al.* [41] study the problem of transfer learning using a semi-supervised HMM for adaptive localization in an indoor environment. Wei *et al.* [38] propose the co-regularized heterogeneous transfer learning model to transfer the knowledge from social media platforms to sensors in human activity recognition. Quite recently, there have been studies about transfer learning in urban computing. For example, Fan *et al.* [11] present CityCoupling, which establishes an inter-city spatial mapping in one city as input and reproduces human mobility in another city. Wei *et al.* [37] propose the flexible transfer learning method for air quality prediction, which aims to transfer knowledge from a city with sufficient multimodal data and labels, to the cities with insufficient data and scarce labels. All of the above studies leverage knowledge transfer to solve the label scarcity problem in different fields. However, they are mainly limited to one-fold knowledge transfer, and they rarely explore transfer learning in the recommendation field. In our work, we want to transfer both local characterizing knowledge from similar enterprises in the target city and enterprise bias knowledge from the same enterprise in other cities, which is a new problem and is not addressed by traditional transfer learning methods. Alternatively, we are inspired by the SVD-based CF method in recommendation systems, where we view the enterprises in the same city as users and the location grids as items. The framework enables us to transfer local knowledge from similar enterprises in the same city. Furthermore, to address the cold-start problem in our framework, we leverage both inter-city knowledge association and intra-city semantic extraction methods to build a shared semantic space for inter-city knowledge transfer (considering that different cities have different feature and label distributions), i.e., transferring knowledge of the target enterprise from other cities with rich data.

## 3 CITYTRANSFER: AN OVERVIEW

### 3.1 Problem Formulation

We first give formal definitions of the main concepts as well as the problem of our work. A summary of the definition of notations used in our work is given in Table 1.

***Definition 1.*** A *city* is represented as a set of two-dimensional location grids, each of which has a size of $\kappa$ m× $\kappa$ m (e.g., $\kappa$ = 500m). We consider two different cities in knowledge transfer. A *source city s* (like Beijing in the above scenario) has $m_1$ grids, denoted as $G^{(s)} = \{g_1^{(s)},\ldots , g_{m1}^{(s)}\}$. For each grid $g_j^{(s)}$, we extract a feature vector $f_j^{(s)} \in R^d$. Similarly, a *target city t* (like Xi'an in above scenario) has $m_2$ grids, denoted as $G^{(t)} = \{g_1^{(t)}; \ldots, g_{m2}^{(t)}\}$. We extract a feature vector $f_l^{(t)} \in R^d$ for each grid $g_l^{(t)}$.

***Definition 2.*** A *chain enterprise* is an enterprise that manages a number of hotel outlets located in different places.

We consider two types of chain enterprises. A *source chain enterprise* (like Hanting Inn and 7 Days Inn in the above scenario) has stores in both the source city $s$ and the target city $t$. A *target chain enterprise* (like Home Inn) only has stores in the source city $s$. Suppose that we have $n$ chain enterprises, denoted as $H = \{h_1,\ldots, h_n\}$. Without loss of generality, we consider $h_n$ as the target chain enterprise and the remaining $\{h_1,\ldots , h_{n-1}\}$ as the source chain enterprise. For a chain enterprise $h_i$ that has a store in a location grid $g_j^{(s)}$ in the source city $s$, we calculate its rating $r_{ij}^{(s)} \in [0, 5]$, indicating the popularity of this store. As a result, we can denote the rating of all the chain enterprises $H$ in the source city as $R^{(s)} = \{r_{ij}^{(s)}\}$. Similarly, for a chain enterprise $h_i$ which has a hotel in a location grid $g_l^{(t)}$ in the target city $t$, we also calculate a rating $r_{il}^{(t)} \in [0, 5]$ and the ratings of all the source chain enterprise $H / h_n$ in the target city can be denoted as $R^{(t)} = \{r_{il}^{(t)}\}$.

***The CityTransfer problem***: Based on the above definitions, we aim to use $R^{(s)}$ and $R^{(t)}$, together with the location grid features $f_j^{(s)}$ and $f_l^{(t)}$, to predict the ratings for the target chain enterprise $h_n$ over the location grids in the target city $t$.

Table 1. A summary of the definition of notations

| Notation | Description |
|---|---|
| $s, t$ | Source city, target city |
| $G^{(s)} = \{g_1^{(s)},\ldots , g_{m1}^{(s)}\}$ $G^{(t)} = \{g_1^{(t)},\ldots ,g_{m2}^{(t)}\}$ | The set of grids in source or target city. |
| $f_j^{(s)}, f_l^{(t)}$ | The set of original features in source or target city. |
| $H = \{h_1,\ldots, h_n\}$ | The set of all chain enterprises. $h_n$ is the target enterprise, $H/h_n$ is source enterprises. |
| $\boldsymbol{R}^{(s)} = \{r_{ij}^{(s)}\}$ | The set of ratings (ratings matrix) in source or target city. |
| $\boldsymbol{R}^{(t)} = \{ r_{il}^{(t)}\}$ | $r_{ij}$ is the ratings for $h_i$ in $g_j$, $r_{ij=}$ [0,5] |
| $\boldsymbol{u} = \{u_i\}$ | The matrix of chain enterprise, $u_i$ is the vector of $\boldsymbol{u}$ about $h_i$ |
| $\boldsymbol{v}^{(s)} = \{v_j\}$, $\boldsymbol{v}^{(t)} = \{v_l\}$ | The matrix of location grids in source or target city, $v_i$ is the vector of $\boldsymbol{v}$ in $g_j$, it is also the vector of new feature constructed by AutoEncoder. |
| $b_i$ | The bias of chain enterprise $h_i$, |
| $e_j^{(s)},e_l^{(t)}$ | The bias of location grid $g_j$ in source or target city. |
| $\boldsymbol{W}^{(s)},\boldsymbol{W}^{(t)}, y_\tau^{(s)}, y_\tau^{(t)}$ | The parameters of AutoEncoder[25]. |
| $\Delta = \{(g_j^{(s)}, g_l^{(t)})\}$ | The set of all the location grid pairs between two cities. |

## 3.2 Data Collection and Analysis

We choose chain hotel enterprise for a case study and collect data about two major stakeholders in our work: 1) *Chain hotel enterprise data*, including the profile information (e.g., name, location, and category) and all the reviews contributed by the consumers for each chain hotel. It should be noted that the number of reviews for a chain store can to some extent reflect its popularity, and thus we use it as the ground truth to evaluate the prediction result. 2) *Urban characterizing dat*a, including POIs, check-ins, and house price. It is used to characterize the features of different places of a city and further understand the impact of enterprise bias and local characters on chain enterprise site selection.

A summary of the multi-source data collected for four major cities (Beijing, Shanghai, Xi'an, and Nanjing) in China is given in Table 2. (1) The chain hotel enterprise data is crawled from Ctrip (a popular travel reservation website in China), where three popular *economic* chain hotels, namely Hanting Inn, 7 Days Inn, and Home Inn, are chosen for the study. We view them as similar chain enterprises in this work. (2) The POIs have 12 different categories (e.g., education, scenic spots, sports, commercial spots, financial services, and transport facilities), and they are crawled from GaoDe Map (one of the most popular digital map service providers in China). (3) The check-in data is collected from the geo-tagged posts in Sina Weibo (the most famous microblogging service in China) of the four cities, dating from Aug. 1 to Nov. 30 in 2015. (4) The house price for each city is crawled from SouFang[5] (a popular website for real estate in China).

For each location grid, we treat it as an *instance* and extract features from multiple data sources. We measure the business value of a location grid to a chain enterprise by counting the average number of reviews for each chain store of the enterprise within the location grid. To ensure prediction quality, the business value is finally normalized to a *rating score* within the range of [0, 5]. As discussed earlier, the rating score is used as the ground truth for the recommendation model.

Table 2. A summary of the multi-source data collected for four major cities in China.

| Sources | Cities | | | |
|---|---|---|---|---|
| | Beijing | Shanghai | Xi'an | Nanjing |
| Hanting Inn Hotels | 123 | 147 | 69 | 73 |
| 7 Days Inn Hotels | 160 | 46 | 59 | 46 |
| Home Inn Hotels | 179 | 156 | 189 | 57 |
| Hanting Inn Reviews | 18195 | 22875 | 8415 | 10020 |
| 7 Days Inn Reviews | 31215 | 8610 | 8985 | 9915 |
| Home Inn Reviews | 35310 | 45146 | 14956 | 16095 |
| POIs | 348863 | 444703 | 143397 | 166648 |
| Check-ins | 21222070 | 16928489 | 3011698 | 3410412 |
| House price | 55030 | 50224 | 7758 | 24781 |

Table 3. The impact of enterprise bias (for Hanting) on site selection in different cities (PCC).

| Beijing | | Shanghai | | Xi'an | | Nanjing | |
|---|---|---|---|---|---|---|---|
| Scenic spots | 0.69 | Scenic spots | 0.33 | Government | 0.42 | Education | 0.59 |
| Medical | 0.62 | Transport | 0.24 | Life | 0.35 | Scenic spots | 0.52 |
| Government | 0.61 | Financial | 0.23 | Medical | 0.26 | Commercial | 0.51 |
| Financial | 0.61 | Medical | 0.23 | Financial | 0.25 | Government | 0.50 |
| Sport | 0.59 | Commercial | 0.22 | Commercial | 0.25 | Public facility | 0.46 |
| Commercial | 0.58 | Sport | 0.20 | Education | 0.24 | Financial | 0.46 |

[5] http://www.fang.com/

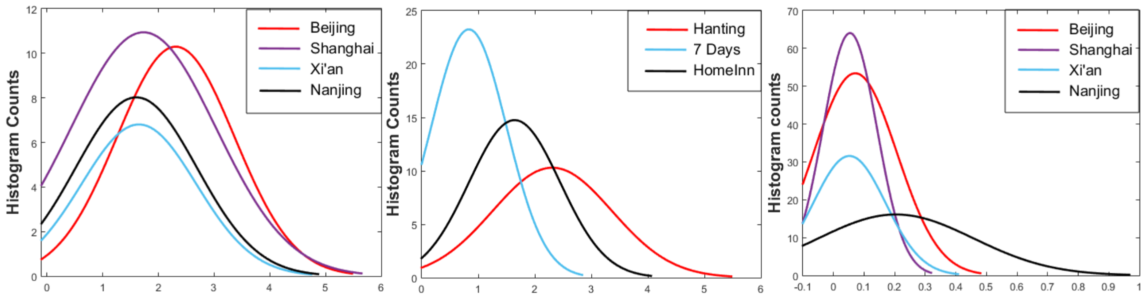
*3.2.1 The impact of enterprise bias on site selection in different cities.* To understand the impact of chain enterprise bias on site selection, we measure the Pearson Correlation Coefficient (PCC) between the rating score of each location grid and the number of POIs (of different categories) in the grid. The top six correlated POI categories for the Hanting Inn enterprise in four cities are given in Table 3. We can find that for most of these cities, *scenic spots*, *financial services*, *government agencies*, and *commercial institutions* are of high correlation on site selection. It indicates that there is enterprise bias on site selection in different cities and we can transfer the enterprise bias knowledge from the source city to the target city. However, we also find that there are certain differences in terms of the members of highly-correlated POI categories and the correlation strength among them. For example, the POI category 'education' is of high relevance to Xi'an and Nanjing while not for the other two cities. Therefore, only transferring the knowledge from a source city is not sufficient for accurate rating prediction.

*3.2.2 The impact of local features on site selection for similar enterprises.* To understand the impact of local features on site selection, we propose the *co-occurrence ratio* (COR) between the *target* enterprise and the *source* enterprise. For example, if we assume Hanting Inn and 7 Days Inn as the target enterprise and the source enterprise, respectively, we calculate the COR for Hanting to 7 Days, i.e., *COR(Hanting, 7 Days)*, as the ratio of the number of stores of Hanting that co-locates with 7 Days stores in any location grids to all Hanting stores in the city. The COR results for different target-source enterprise pairs in Beijing and Shanghai are given in Table 4. We can find that the COR is higher than 0.5 for most target-source enterprise pairs, which indicates that there is significant correlation on site selection for similar enterprises in the same city. This motivates us to transfer the local knowledge from similar enterprises in the target city. However, we also find that for some enterprise pairs, the COR is low and thus only local knowledge is not enough for accurate site recommendation.

Table 4. The impact of local features on site selection for similar enterprises (COR).

| | Beijing | | | Shanghai | | |
|---|---|---|---|---|---|---|
| | Hanting | 7 Days | Home Inn | Hanting | 7 Days | Home Inn |
| Hanting | - | 0.86 | 0.83 | - | 0.3 | 0.62 |
| 7 Days | 0.58 | - | 0.65 | 0.75 | - | 0.79 |
| Home Inn | 0.70 | 0.82 | - | 0.72 | 0.37 | - |

*3.2.3 The difference of feature/rating distribution in different cities.* Data from different cities may have different distributions in feature and rating spaces. To understand this in our problem, we create the fitting curve of rating scores (as discussed earlier in this section) and features under different settings. As shown in Fig. 1, the distribution of ratings of Hanting in four different cities is different (Fig. 1a), and the distribution of ratings for the three enterprises in the same city (i.e., Beijing) is quite different (Fig. 1b). We can also find in Fig. 1(c) that the distribution of house price (i.e., a kind of feature) is also quite different in different cities. Therefore, we need to build a shared semantic space to make the features/ratings comparable between different cities and enable inter-city knowledge transfer.




(a) Grid Rating (b) enterprise Rating (c) House Pring

Fig 1. The difference of feature/rating distribution in different cities.

## 3.3 The CityTransfer Framework

The framework of *CityTransfer* is given in Fig. 2, which consists of three modules: the data preparation module, the CityTransfer learning model, and the recommendation module.

***Data preparation***. We collect multi-source data for both the source city and the target city, and then divide each city into a set of two-dimensional grids. For each grid, we extract useful features, including geographic features and commercial features. More details about the features will be given in Section 4.

***The CityTransfer learning model***. It consists of three major components:

1) *Transfer Rating Prediction Model*. It predicts the rating of location grids to the target chain enterprise by extending the SVD-based CF algorithm [13,22], by which we can learn local characterizing knowledge from similar enterprises in the target city.

2) *Inter-City Knowledge Association*. To enable enterprise bias knowledge transfer, we compute the Pearson correlation coefficient among the location grids in the source and target city, and use an optimization function to build a shared semantic space to address the incomparability issue between the location grids of different cities.

3) *Intra-City Semantic Extraction*. For each city, the local grid features extracted from multi-source datasets are redundant and noisy. This component extracts high quality semantics by using AutoEncoder [4,10,29] to reconstruct original features and improve the knowledge transfer performance.

Finally, we combine the above three aspects as the *Joint Optimization Algorithm* to predict the rating for a target enterprise in the target city.

***Recommendation***. We rank ratings over all location grids for the target enterprise in a target city and recommend top *N* places accordingly.

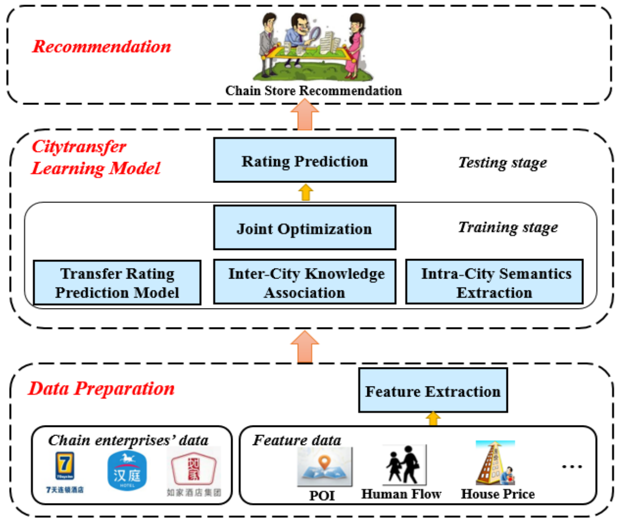



Fig. 2. The CityTransfer Framework.

## 4 FEATURE EXTRACTION

As discussed earlier, similar to previous urban computing studies [37, 40], we divide the city (both source and target) into a set of location grids, and extract a set of features to characterize each grid. Specifically, the extracted features are classified into two categories: *geographic features* (e.g., diversity, human flow, traffic convenience, POI set) and *commercial features* (density, competitiveness, complementarity, and house price) [19, 26].

### 4.1 Geographic features

The performance of chain stores is closely related to the spatial characteristics of the place where the store resides. We use the following geographic features to characterize it.

***Diversity***. It reflects the grid's spatial heterogeneity. We refer to $N_c$ as the number of POIs of category $c$ ($c \in \Gamma$, $\Gamma$ is the set of all the POI categories mentioned in Section 3) in the grid $g_j$. Diversity is defined as Eq. (1) based on the information entropy theory.

$$f_j^{div} = -\sum_{c\in\Gamma} \frac{N_c(j)}{N(j)} \times \log \frac{N_c(j)}{N(j)} \tag{1}$$

***Human flow***. It reflects the grid's area popularity. We measure it based on the total number of check-ins in grid $g_j$, as define in Eq. (2). Here, $CH$ refers to all check-in places, and place $p$ is a check-in point that falls in location grid $g_j$.

$$f_j^{HF} = -\sum_{p\in CH} num(p)(p \in g_j) \tag{2}$$

***Traffic convenience***. It reflects the location grid's traffic convenience. We use public transportation $trans=\{t_1, t_2, t_3\}$ (buses, subways, and parking lots) to characterize it. Specifically, $num(t_i)$ is the number of stations for $t_i$ ($t_i \in$ trans) in grid $g_j$, which is defined as Eq. (3).

$$f_j^{traffic} = -\sum_{t_i \in trans} num(t_i) \tag{3}$$

***POI set***. It combines all the categories of POIs in location grid $g_j$, which is defined as Eq. (4). Here, $f^c$ ($c \in \Gamma$) is the number of POIs of category $c$ in the grid.

$$f_j^{POI} = \left\{ f_j^{c_1}, f_j^{c_2}, \ldots, f_j^{c_n} \right\} \tag{4}$$

### 4.2 Commercial features

To characterize the commercial features of a location grid, we extract the following commercial features.

***Density.*** It is an indicator of the popularity of a target chain enterprise, calculated as the number of stores of this enterprise in a grid, as formulated in Eq. (5).

$$f_j^{den} = N_{tar} \tag{5}$$

***Competitiveness.*** We assume that competition mostly comes from the nearby stores with the same POI category, and we define the competitiveness feature as Eq. (6). Here, $N_t(r)$ represents the number of chain stores of the same category in grid $g_j$, and $N_{c/tar}$ is the number of chain stores of the same category except the target enterprise's stores.

$$f_j^{com} = -\frac{N_{c/tar}}{N_c} \tag{6}$$

***Complementarity.*** We firstly measure the complementary coefficient between different POI category pairs as Eq. (7). After that we compute the aggregated complementarity based on the POI category distribution of location grid $g_j$ as Eq. (8).

$$\rho_{t \to t'} = \frac{N_{set}(t,t')}{N_\Gamma (N_\Gamma - 1)/2} \tag{7}$$

$$f_j^{cp} = \sum_{t \in \Gamma} \log(\rho_{t \to t'}) \times (N_{jt'} - \overline{N_{t'}}) \tag{8}$$

where $N_{set}(t,t')$ is the number of co-occurrences of category pair $(t, t')$ in the same grid; $N_\Gamma (N_\Gamma - 1)/2$ refers to the total number of different category-pairs. $N_{jt'}$ is the number of $t'$ that appears in grid $g_j$, and $\overline{N_{t'}}$ is the average number of POIs of category $t'$ that appear in all location grids.

***House price.*** For each location grid, we use the average price of all collected housing points within this grid as its house price feature.

$$f^{HP} = \overline{hp_i} \tag{9}$$

## 5 CITYTRANSFER: DETAILED DESIGN

To address the cold-start problem in chain store site recommendation, *CityTransfer* significantly extends the SVD-based CF model to enable two-fold knowledge transfer. It consists of three major components: *Transfer Rating Prediction Model*, *Inter-City Knowledge Association* and *Intra-City Semantic Extraction*. We first describe each of the three components and then present the proposed joint optimization algorithm.

### 5.1 Transfer Rating Prediction Model.

To leverage chain enterprise features and location grid features for chain location recommendation, we introduce chain enterprise matrix $\boldsymbol{u}$ and location grids matrix $\boldsymbol{v}$ and use them to predict the rating of each candidate location. Specifically, $u_i \in R^k$ is a vector of $\boldsymbol{u}$, which represents the feature of one chain enterprise, $v_j \in R^k$ is a vector of $\boldsymbol{v}$, which represents the feature of a grid. Founded on the SVD-based CF model [21], we establish connections between $\boldsymbol{u}$ and $\boldsymbol{v}$ by analyzing past behaviors of source and target enterprise in each city. That is, we first divide the rating matrix ($\boldsymbol{R}^{(s)}$ or $\boldsymbol{R}^{(t)}$) into $\boldsymbol{u}$ and $\boldsymbol{v}$ for each city, and then use the product of $u_i$ and $v_j$ to predict the rating of $g_j$. As the rating scale might be different for different chain enterprises and different location grids, we introduce the bias for a chain enterprise as $b_i$ and the bias for location grids as $e_j$, and combine them with $u_i$ and $v_j$ to improve the quality of rating prediction.

For the chain enterprise, the feature $u_i$ and the bias $b_i$ are invariant across different cities, as a given chain enterprise often has a fixed market positioning and location preference. Thereby, we can transfer chain enterprise knowledge from source city to the target city. In contrast, for the location grid, the feature $v_j$ and the bias $e_j$ of different cities are not the same. When transferring knowledge from the source to the target city, there exists the problem whether $v_j^{(s)}$ *and* $v_j^{(t)}$ are comparable, and we will address this issue based on *Inter-City Knowledge Association* (presented in Section 5.2). Therefore, we predict the rating for $h_i \in H$ in the source city using Eq. (10).

$$\hat{r}_{ij}^{(s)} = b_i + e_j^{(s)} + u_i^T v_j^{(s)} \tag{10}$$

Similarly, we predict the rating for a source chain enterprise $hi$, $\forall i = 1, \ldots, n-1$, in the target city $t$ with Eq. (11).

$$\hat{r}_{ij}^{(t)} = b_i + e_j^{(t)} + u_i^T v_j^{(t)} \tag{11}$$

Finally, to minimize the prediction error, the optimization objective is defined as Eq. (12).

$$O_1 = \sum_{r_{ij}^{(t)} \in R^{(t)}} (\hat{r}_{ij}^{(t)} - r_{ij}^{(t)})^2 + \lambda_1 \sum_{r_{ij}^{(s)} \in R^{(s)}} (\hat{r}_{ij}^{(s)} - r_{ij}^{(s)})^2 \tag{12}$$

where $\lambda_1 > 0$ is a trade-off parameter.

## 5.2 Inter-City Knowledge Association.

As presented in Section 3.2, there exists discrepancy between the source city and the target city, i.e., the distribution diversity of features and ratings [37]. Therefore, we cannot directly compare two location grids in different cities. To make the location grids to be comparable across different cities, we propose to actively build correspondence between two cities and represent them in a shared semantic space. After the conversion, $v^{(s)}$ and $v^{(t)}$ become comparable, and we can transfer knowledge from the source to the target city. Specifically, to build the correspondence between $g_j^{(s)}$ and $g_l^{(t)}$, we consider measuring the *Pearson correlation coefficient* [1,3] between $\boldsymbol{f}_j^{(s)}$ and $\boldsymbol{f}_l^{(t)}$ as defined in Eq. (13).

$$\rho(f_j^{(s)}, f_l^{(t)}) = \frac{\sum_{i=1}^{k} (f_{j,i}^{(s)} - f_j^{(s)})(f_{l,i}^{(t)} - f_l^{(t)})}{\sqrt{\sum_{i=1}^{k} (f_{j,i}^{(s)} - \overline{f}_j^{(s)})^2 \sum_{i=1}^{k} (f_{l,i}^{(t)} - \overline{f}_l^{(t)})^2}} \tag{13}$$

where $\overline{f}_j^{(s)} = \frac{1}{k}\sum_{i=1}^{k} f_{j,i}^{(s)}$ and $\overline{f}_j^{(t)} = \frac{1}{k}\sum_{i=1}^{k} f_{l,i}^{(t)}$. Based on the correlation, to build the grid-grid correspondence, for each location grid $g_j^{(s)}$ in the source city $s$, we pick out top $\gamma$ location grids $g_l^{(t)}$ in the target city $t$. Similarly, for each location grid $g_l^{(t)}$ in the target city $t$, we pick out top $\gamma$ location grids $g_j^{(s)}$ in the source city. In total, we denote the set of all the location grid pairs between two cities as $\Delta = \{(g_j^{(s)}, g_l^{(t)})\}$.

To make each pair of location grids from two cities be similar in the shared semantic space, we need to optimize the function described in Eq. (14).

$$O_2 = \sum_{(g_j^{(s)}, g_j^{(t)}) \in \Delta} \rho(f_j^{(s)}, f_l^{(t)})(v_j^{(s)} - v_j^{(t)})^2 \tag{14}$$

## 5.3 Intra-City Semantic Extraction

As explored in other urban computing studies [8, 37], local grid features extracted from multi-source datasets are redundant and noisy. For example, Figure 3 gives the Pearson Correlation Coefficient (PCC) among different feature pairs in the Beijing city, and we find that there is certain correlation among various

POI categories, human flow, traffic convenience features, and so on. The intra-city semantic extraction component is to reduce the impact of redundancy and noisy and improve the prediction performance.

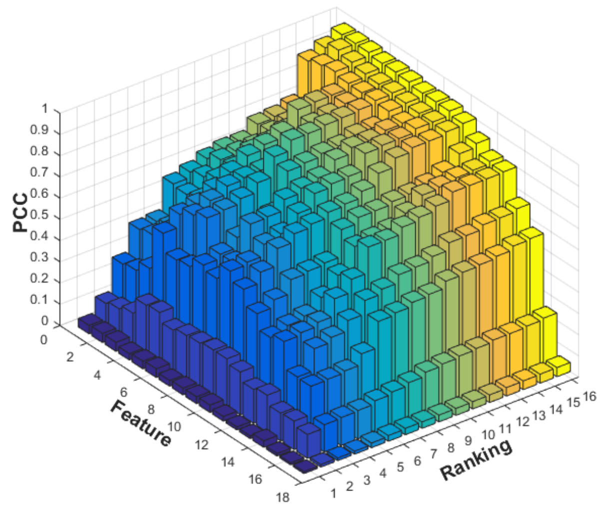



Fig. 3. The Pearson correlation coefficient (PCC) among diverse features.

To ensure each location grid $v_j$ is robust and informative, and preserve more semantics, we reconstruct the location grid's original features $f_j$ to its corresponding $v_j$. Specifically, we use AutoEncoder [4,10,29] to construct $v_j$ from $f_j$. In general, there can be multiple layers in the network architecture of AutoEncoder. For simplicity, in this paper we use a three-layer AutoEncoder as an example to illustrate how we construct $v_j$ from its corresponding $f_j$. In particular, we give the formal definition as Eq. (15) and Eq. (16), respectively.

$$v_j^{(s)} = \sigma(W^{(s)} f_j^{(s)} + y_1^{(s)}) \tag{15}$$

$$v_j^{(t)} = \sigma(W^{(t)} f_j^{(t)} + y_1^{(t)}) \tag{16}$$

where $\boldsymbol{W}^{(s)} \in \mathrm{R}^{d_* k}$, $\boldsymbol{W}(t) \in \mathrm{R}^{d_* k,}$ $\boldsymbol{t_1}^{(s)} \in \mathrm{R}^k$ and $\boldsymbol{t_1}^{(t)} \in \mathrm{R}^k$ are parameters; $\sigma(\mathbf{z})$ is a neural activation function, which returns a vector of sigmoid function value. We reconstruct $f_j$ from $v_j$ by a nonlinear function using the tied parameters $W^{(s)}$ and $W^{(t)}$, as defined with Eq. (17) and Eq. (18), respectively.

$$\hat{f}_j^{(s)} = \sigma(W^{(s)T} v_j^{(s)} + y_2^{(s)}) \tag{17}$$

$$\hat{f}_j^{(t)} = \sigma(W^{(t)T} v_j^{(t)} + y_2^{(t)}) \tag{18}$$

where $\mathbf{y}_2^{(s)} \in \mathrm{R}^d$ and $\mathbf{y}_2^{(t)} \in \mathrm{R}^d$ are also parameters. Finally, to minimize the reconstruction error, we request $f_j$ to be close to $\hat{f}_j$ which is defined as Eq. (19).

$$O_3 = \sum_{j=1}^{m_1} \left\| \hat{f}_j^{(s)} - f_j^{(s)} \right\|^2 + \sum_{j=1}^{m_2} \left\| \hat{f}_j^{(t)} - f_j^{(t)} \right\|^2 \tag{19}$$

## 5.4 The Optimization Algorithm

Based on the above components, we first present the joint optimization function. We denote $\Theta = \{b_i, e_j^{(s)}, e_l^{(t)}, u_i, \boldsymbol{W}^{(s)}, \boldsymbol{W}^{(t)}, y_\tau^{(s)}, y_\tau^{(t)} | \ i=1,..,n;\ j=1,\ldots m_1;\ l=1,\ldots m_2;\ \tau=1,2\}$. To avid overfitting, we introduce a regularization term as Eq. (20).

$$R = \sum_{i=1}^{n} b_i^2 + \sum_{i=1}^{m_1} e_i^{(s)2} + \sum_{i=1}^{n} e_i^{(t)2} + \sum_{i=1}^{n} \left\| u_i \right\|^2 + \left\| W^{(s)} \right\|^2 + \left\| W^{(t)} \right\|^2 + \sum_{\tau=1}^{2} \left\| y_\tau^{(s)} \right\|^2 + \sum_{\tau=1}^{2} \left\| y_\tau^{(t)} \right\|^2 \tag{20}$$

In all, we combine $O_1$, $O_2$, $O_3$ and $R$ as the final objective function. We optimize $\Theta$ by minimizing $\zeta$ in Eq. (21).

$$\zeta = \frac{1}{2} O_1 + \frac{\lambda_2}{2} O_2 + \frac{\lambda_3}{2} O_3 + \frac{\lambda_4}{2} R \tag{21}$$

where $\lambda_2 > 0$, $\lambda_3 > 0$ and $\lambda_4 > 0$ are trade-off parameters.

We implemented our proposed methods based on Theano[6], and choose stochastic gradient descent (SGD) [40] as the optimization methods. Once $\Theta$ is learned to optimization, we can use Eq. (10) to predict the rating in each location grid of the target city $t$ for the target hotel chain $h_n$.

To conclude this section, we give the entire pseudo-code of the *CityTransfer* algorithm in Algorithm 1, which includes a training stage and a testing stage. The inputs of our model include feature vector $f_j$ and rating matrix R from both source and target cities. The training stage (line1~8) is responsible for building grid-grid correspondence, initializing $\Theta$ and optimizing $\Theta$ by minimizing in Eq. (20). In the testing stage (line 9~13), the rating for the target chain enterprise $h_n$ in target city $t$ is predicted with the overall optimization function.

---

ALGORITHM 1: The *CityTransfer* Algorithm

Input: Source city: $G^{(s)}$, each of which has a feature vector $f_j^{(s)}$ $(j=1,\ldots,m_1)$,
ratings in source and target chain enterprises H ,which is presented as $R^{(s)} = \{r_{ij}\}$
Target city: $G^{(t)}$, each of which has a feature vector $f_l^{(t)}$ $(l=1,\ldots,m_2)$,
ratings in source chain enterprises $H/h_n$, which is presented as $R^{(t)} = \{r_{il}\}$
Output: Rating predictions for $h_n$ over $G^{(t)}$ in target city.

1. % Training stage
2. % build actively grid-grid correspondence;
3. **for all** l =1,...,$m_2$ do
4. Compute the correlation between $f_j^{(s)}$ and any $f_j^{(t)}$ by Eq. 13;
5. based on the correlation scores, choose the top $\gamma$ grids $g_j^{(s)}$ for the grid $g_l^{(t)}$ and add $(g_j^{(s)}, g_l^{(t)})$ into $\Delta$;
6. **end for**
7. Initialize $\Theta$;
8. Use Theano to optimize $\Theta$ by minimizing $\zeta$ in Eq. 21;
9. % Testing stage
10. **for all** $\boldsymbol{l}$=1,…,$\boldsymbol{m_2}$ do
11. Construct the location grid feature $v_j^{(t)}$ by Eq.16;
12. Predict the rating $r_{nl}^{(t)} \leftarrow b_n + e_j^{(t)} + u_n^T v_j^{(t)}$ by Eq. 11;
13. **end for**

## 6 EVALUATION AND DISCUSSION

We first give a summary of the experiment purposes to evaluate the performance of *CityTransfer*, based on which a set of baselines are defined. We then present the experiment settings as well as the significant experimental results. Finally, we

[6] http://deeplearning.net/software/theano/

make a discussion of the deep insights and limitations of our work. The experiments are based on the datasets described in Section 3.

## 6.1 Experiment Purposes and Baselines.

The successful of CityTransfer relates to several different aspects, and we present the following purposes for experiment design.

- We firstly compare the performance of *CityTransfer* with empirical methods, where certain important factors may be considered in site selection, such as human flow, traffic convenience, house price, etc. This helps us understand the usefulness of our approach regarding frequently-used empirical approaches.
- The unique feature of *CityTransfer* largely depends on its two-fold knowledge transfer framework. We wonder whether the transferring of both inter-city and intra-city knowledge performs better than single-fold knowledge transfer, i.e., only transferring enterprise bias knowledge from a source city or local characterizing knowledge from similar enterprises in the target city.
- The *CityTransfer* framework consists of three major components, and a joint optimization function is used to combine them. The question is *whether each component plays important roles in the whole framework*. In other words, we should investigate the effects of the inter-city knowledge association and the intra-city semantic extraction modules.
- For a target city, there could be several candidate source cities to be chosen for knowledge transfer. The following questions are thus interesting to be investigated: *Are there significant performance differences when choosing different source cities? Is there prior knowledge that can be leveraged to steer source city selection?*
- For a site recommendation problem, there might be several candidates to be chosen in practice. The geo-distribution of them can be different in the city, e.g., in a *centralized* manner or in a *scattered* manner. We thus want to validate the performance of our framework under diverse real-world settings.

Based on the above experiment purposes, we define and adopt the following six baselines for comparison.

**1) Empirical selection (ES)**, which recommends optimal locations based on certain significant features in the target city, such as the human flow, traffic convenience, and house price. In particular, we use the ranking of location grids based on the selected factor/feature (e.g., human flow) as the ranking result for site recommendation.

**2) Inter-city-transfer (INTER_T)**, which only transfers knowledge about the target chain enterprise from the source city to the target city. It learns a regression model for the target chain enterprise in the source city, and then uses this model in the target city to score the candidate places for target chain enterprise recommendation.

**3) Intra-city-transfer (INTRA_T)**, which only transfers knowledge about similar chain enterprises in the target city. It learns a regression model about similar enterprises in the target city, and then uses it to do location recommendation for the target enterprise.

**4) Matrix Factorization (MF)**, which only uses the SVD-based CF method to predict ratings [2,23,35]. It omits the incomparable (i.e., distribution diversity of labels and features) nature between cities and directly 'employ' the labelled-instances from a source city to address the cold-start issue. Therefore, we only use $O_1$ as the objective function to be optimized.

**5) MF& Semantic Extraction (MF_SE)**, which uses AutoEncoder to preserve more semantics, and adds it to the matrix factorization method for rating prediction. In other words, we use $O_1 + O_2$ as the optimization objective function.

**6) MF& Knowledge Association (MF_KA)**, which firstly builds the correspondence between the location grids of two cities by measuring the Pearson correlation coefficient, and then introduces it to the matrix factorization method for rating prediction. In this way, the location grid features are comparable across different cities, and we can transfer knowledge from the source city. In a word, we use $O_1 + O_3$ as the optimization objective function.

## 6.2 Experiment Settings

*6.2.1 Parameter setting and training.* The *CityTransfer* models are learnt by optimizing the loss function of Eq. (20). We use a pre-learning scheme to determine the initial values of parameter $W$ and $b$, and other model parameters are learned using SGD and AutoEncoder. Based on the empirical knowledge learned from previous studies [8,37,42] and our experiments, we set the location grid size to 500m*500m in the experiments, which is not too small that has insufficient features nor too big that results in less labelled instances. Based on a large number of experiments, we find that the best performance can be achieved when the number of hidden layers (or latent features) in AutoEncoder is 8-10 in most cases, and we set it to '9' in the following experiments. The learning rate (LR) decay method [4, 10] is used in SGD for LR adjusting. For trade-off parameters, we try different pairs of $\lambda_i (i = 1,2,3,4)$, and employ the cross-validation method [1] to identify the best pair. Finally, we find that $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\lambda_3 = 0.5$, $\lambda_4 = 0.025$ perform the best mostly and use them in the experiments. The features have different range of value and, to improve learning performance, we normalize the feature values within the range of [0, 1]. For each experiment, we repeat it 1,000 times and use the average value as the final result.

*6.2.2 Evaluation metrics.* As rating prediction is a typical regression problem, we employ the often-used metric RMSE (Root Mean Square Error) [16] for result comparison. Furthermore, as a ranking-based site recommendation problem, we assume that there could be several candidate places to be chosen from, and the output is to make a ranking among them for recommendation. In this way, we also adopt the metric that is generally used in ranking algorithms, namely NDCG (Normalized Discounted Normalized Gain)[18], defined by Eqs. (22) and (23).

$$NDCG@K = \frac{DCG}{IDCG} \tag{22}$$

$$DCG = rel_i + \sum_{i=2}^{n} \frac{rel_i}{\log_2 i} \tag{22}$$

As explored in other urban computing studies [8, 37], local grid features extracted from multi-source datasets are redundant and noisy. For example, Figure 3 gives the Pearson Correlation Coefficient (PCC) among different feature pairs in the Beijing city, and we find that there is certain correlation among various POI categories, human flow, traffic convenience features, and so on. The intra-city semantic extraction component is to reduce the impact of redundancy and noisy and improve the prediction performance.

Here, $K$ represents the number of candidate places to be ranked, and it is set to '10' in the experiments. We use $R$ to represent the predicted ranking list, and $R'$ to denote the real ranking list. $DCG$ is the discounted cumulative gain of $R$, and $IDCG$ is the $DCG$ value of $R'$. $rel_i$ is the rating score of $i$-th object in the ranking list.

## 6.3 Experimental Results

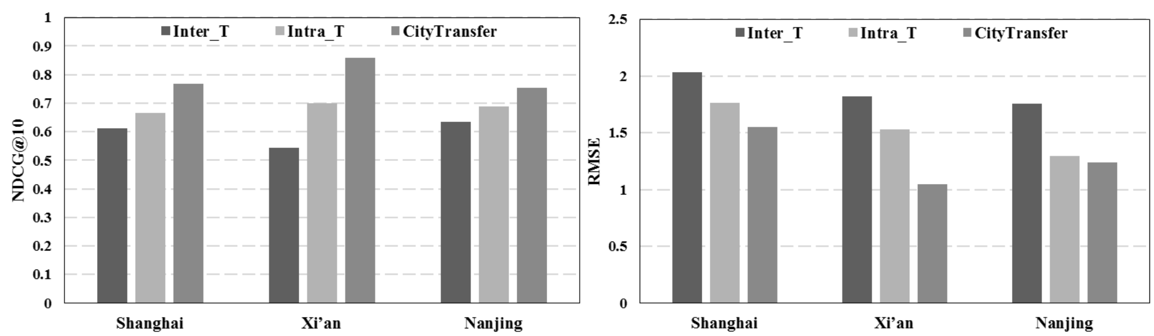
Having depicted the experiment settings and baselines, we present the experimental results regarding the four experiment purposes given in Section 6.1.

*6.3.1 Comparing with empirical selection methods.* We choose Hanting Inn as the target company, and use the ES method (presented in Section 6.1) that leverages local features for site recommendation as the baseline. The performance under six different features for the four different cities is given in Table 5. We can find that the average NDCG value for each city is ranging from 0.2 to 0.5, which is not high. The features perform differently for different cities, for example, human flow performs poor in Shanghai but well in Nanjing. Over the six features, *scenic spot*, *commercial spot*, and *traffic convenience* perform better than other features. However, as shown in Fig. 4, *CityTransfer* largely outperforms ES on Hanting recommendation. For example, for Xi'an, the NDCG value increases from an average of 0.28 for ES to 0.85 by using *CityTransfer*. The results indicate that by leveraging two-fold knowledge transfer, our approach is much better on site recommendation than empirical selection methods.

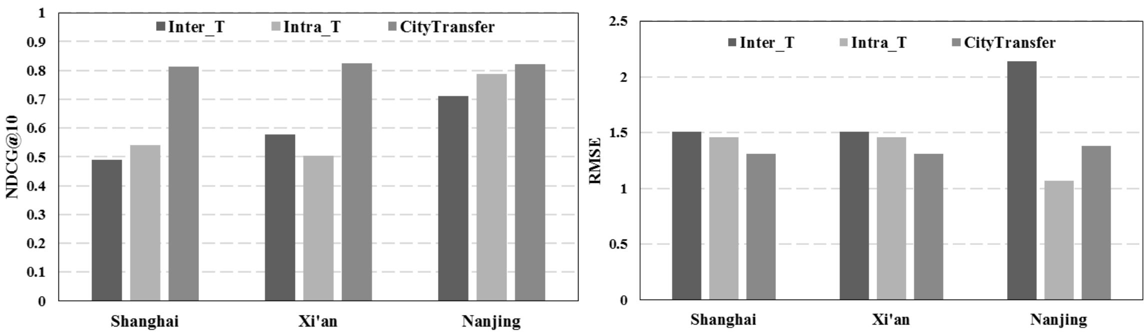
Table 5. The performance of empirical selection methods (NDCG@10).

| Feature | Shanghai | Beijing | Xi'an | Nanjing |
|---|---|---|---|---|
| Scenic spot | 0.61 | 0.42 | 0.26 | 0.48 |
| Commercial spot | 0.55 | 0.42 | 0.28 | 0.46 |
| Traffic convenience | 0.52 | 0.41 | 0.36 | 0.43 |
| POI diversity | 0.60 | 0.37 | 0.26 | 0.41 |
| Human flow | 0.27 | 0.30 | 0.30 | 0.52 |
| House price | 0.48 | 0.45 | 0.23 | 0.48 |
| Average | 0.50 | 0.39 | 0.28 | 0.46 |

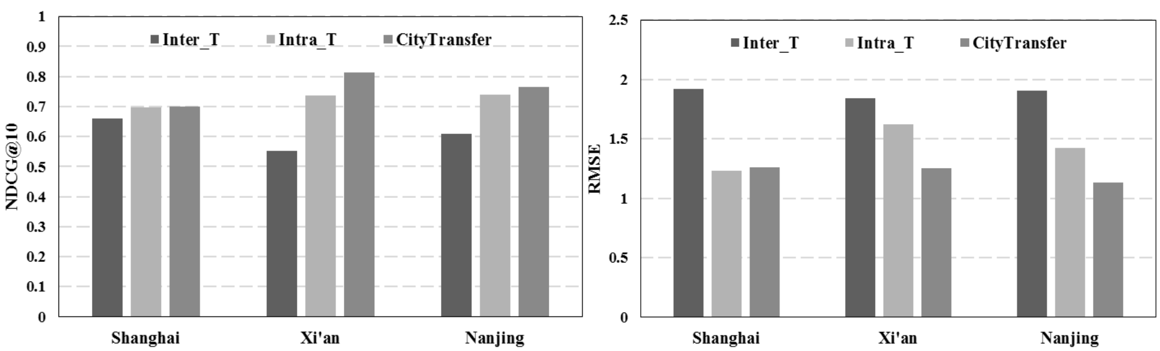
*6.3.2 The usefulness of two-fold knowledge transfer.* To understand the usefulness of two-fold knowledge transfer, we compare *CityTransfer* with two baselines: Inter-city-transfer (INTER_T) and Intra-city-transfer (INTRA_T). We assume Beijing as the source city, and Shanghai, Xi'an, and Nanjing as three target cities. We can find in Fig. 4 that *CityTransfer* outperforms the two baselines in all target cities for three different enterprises (around 0.7-0.85 for NDCG@10, and the relatively lowest RMSE value). It indicates that transferring both enterprise bias knowledge from source cities and local characterizing knowledge from the target city is effective on chain enterprise site recommendation. In other words, the two-fold knowledge facts are complementary and they together improve the recommendation performance.




(a) Hanting




(b) 7 Days




(c) Home Inn

Fig. 4. The performance comparison of one-fold and two-fold knowledge transfer.

We can also find that Intra-city-transfer (INTRA_T) performs better than Inter-city-transfer (INTER_T) in most cases. The reason could be that INTRA_T transfers more local characterizing knowledge which may play an important role in site recommendation. However, when we recommend for 7 Days Inn in Xi'an, INTER_T shows better performance in comparison with INTRA_T. Overall, both INTER_T and INTRA_T have their own advantages, which are critical to be integrated to improve the recommendation performance.

*6.3.3 The effectiveness of inter-city knowledge association and intra-city semantic extraction.* To address the discrepancy of feature and rating distribution, we propose the inter-city knowledge association method. To deal with the redundancy and noisy issues caused by extracting features from multi-source urban data, we develop the intra-city semantic extraction method. In order to evaluate the effectiveness of these proposed methods, we compared *CityTransfer* with three baselines: MF, MF_SE and MF_KA. In the experiment, we choose Beijing as the source city and Shanghai as the target city.

Table 6. The effectiveness of inter-city knowledge association and intra-city semantic extraction.

| | Hanting Inn | | 7 days Inn | | Home Inn | |
|---|---|---|---|---|---|---|
| | NDCG | RMSE | NDCG | RMSE | NDCG | RMSE |
| MF | 0.663 | 1.469 | 0.652 | 1.592 | 0.628 | 1.435 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MF_SE | 0.683 | 1.413 | 0.788 | 1.098 | 0.736 | 1.346 |
| MF_KA | 0.741 | 1.689 | 0.623 | 1.716 | 0.782 | 1.735 |
| CityTransfer | 0.769 | 1.548 | 0.812 | 1.205 | 0.701 | 1.261 |

The results shown in Table 6 indicate that compared to MF, there is around 15-20% performance improvement on NDCG@10 by using *CityTransfer*. The RMSE value is to some extent reduced by using *CityTransfer*. The reason is mainly that MF omits the incomparable nature between cities and the learning process is based on original, redundant features. We also find that by additionally considering inter-city knowledge association (MF_KA) or intra-city semantic extraction (MF_SE) over MF can significantly improve the NDCG@10 performance. The results also indicate that there could be slight difference on performance improvement for MF_KA and MF_SE in different tasks, but a combination of them (*CityTransfer*) usually results in better results. It should be noted that NDCG and RMSE results sometimes conflict. There are two possible reasons. First, the two evaluation metrics cover different aspects: RMSE estimates the error between the predicted value and the true value (of one site), and NDCG calculates the difference between two ranked lists (of a list of candidate sites). Therefore, the changing of them may not coincide under different experiment settings. Second, the ratings of some candidate sites may be similar; so even though the RMSE value is small, the predicted ranking order of the sites can be of significant difference. This may also lead to conflicts between NDCG and RMSE results. Finally, as the site recommendation problem in our work is based on candidate site ranking, NDCG value can better present the performance of different approaches.

*6.3.4 The performance of knowledge transfer from different source cities.* To evaluate the performance of different source cities, we treat Beijing or Shanghai as the source city, and Xi'an or Nanjing as the target city, respectively. From Table 7 and 8, we have the following observations. Overall, as the source city, Beijing performs better than Shanghai on knowledge transfer, the reason of which could be that Beijing has more chain hotels for the three chain enterprises. When we treat Xi'an as the target city, we find that the performance of transferring from Beijing is a bit better than Shanghai, the reason might be that both Beijing and Xi'an are located in north China and have similar geographical features. When the target city is Nanjing, transferring from Shanghai performs better. The reason is similar, as Nanjing is geographically close to Shanghai.

Table 7. The performance of transferring from different source cities to Xi'an (NDCG@10)

| Source City | Hanting Inn | 7 Days Inn | Home Inn |
|---|---|---|---|
| Beijing | 0.859 | 0.826 | 0.814 |
| Shanghai | 0.729 | 0.798 | 0.729 |

Table 8. The performance of transferring from different source cities to Nanjing (NDCG@10)

| Source City | Hanting Inn | 7 Days Inn | Home Inn |
|---|---|---|---|
| Beijing | 0.754 | 0.821 | 0.764 |
| Shanghai | 0.801 | 0.848 | 0.765 |

To sum up, the knowledge transfer among different city pairs generally performs well, with the NDCG@10 value bigger than 0.7. However, carefully choosing source cities can to some extent improve the performance. There are several common factors that could be useful for source city selection: First, the source city should have more instances of the target enterprise. Second, the source city should better be geographically close to the target city, which usually represents higher city similarity. Data-driven urban profiling has also become a hot research area in recent years [25,32]. We can also leverage their approaches to intelligently measure the similarity of different city pairs for source city selection.

*6.3.5 The impact of different distribution of candidate grid sets.* In the above experiments, we randomly sample location grids (the number is '10') as the candidate set for ranking. However, in practice, the candidate grids to be recommended (for opening a new store) may follow different distribution patterns. In other words, the places available (to be rent or sold) for opening new stores in a city can have different geo-distributions. Here we want to study the recommendation performance under different distribution of candidates. In particular, we study three different patterns based on the location grid's geographical distribution: *centralized* (e.g., all candidates come from the central business district area), *scattered* (e.g., the candidates are within different areas of the city), and *random*. We choose Beijing as the source city and Shanghai as the target city in the experiments. Because 7 Days Inn in Shanghai has a small number of stores, we only consider Hanting Inn and Home Inn in the following experiment.

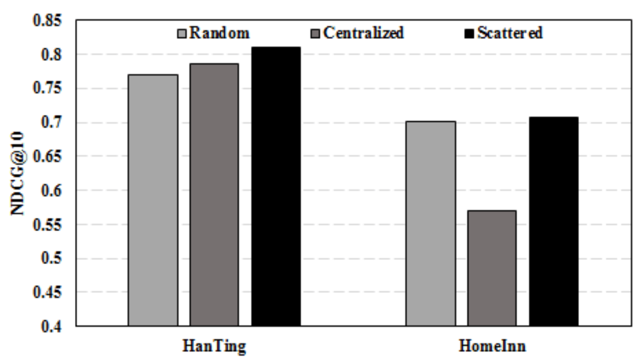



Fig. 5. The performance of different candidate choosing.

We can observe from Fig. 5 that our work generally performs well for different distribution of candidate grid sets. It can also be observed that centralized candidate set performs worse than the other two settings. The reason might be that the feature variance among different grids under the centralized distribution (e.g., all candidates locates in geo-closed areas) could be quite small, the prediction value is thus similar and the ranking result could be error-prone. For the scattered setting, the candidates are far from each other and may have large distance over their feature vector, and thus results in better ranking performance.

## 6.4 Discussions

Having presented the methods and evaluation results, we give a deep discussion of our work and present the limitations to be improved in the future work.

***Learning with multi-source urban data***. AutoEncoder is used for significant feature extraction and noise elimination. It should perform better when working with richer data. Beyond the urban data sources used in the present paper, we intend to investigate the usage of more data sources (e.g., taxi trajectory data, geo-tagged microblogging posts) about urban dynamics. We believe that the performance of our approach can be further improved with the usage of more data sources.

***The data insufficiency issue***. The current work assumes that different cities have equivalent available data sources for ranking prediction. However, in some cases we may have enough data in a big city (e.g., Beijing) but this type of data is insufficient or even missing in another city (e.g., Xi'an). We are planning to consider the data insufficiency or data missing issue in the future work.

***System implementation and intelligent transferring.*** The experiments indicate that for different cities, the performance of knowledge transfer differs, mainly because of distinct city characteristics and the richness of data. Therefore, we will study in depth the effective patterns and develop system support for intelligent knowledge transferring. Several general guidelines and approaches have been discussed in Section 6.3.4.

***Ground truth and experiments for other chain enterprises.*** In our work, the business performance of a location grid for a chain enterprise (i.e., the ground truth for the learning model) is reflected by the average number of reviews for the chain stores of the enterprise in this grid. The framework could be more accurate if we can obtain the operational data from chain enterprises. Furthermore, we use chain hotels as a case study, and we plan to collect data about chain restaurants (e.g., KFC, Starbucks) which have wider distribution and bigger number of stores to validate our work in the future.

***The extension and usage of the CityTransfer model to other applications.*** *CityTransfer* is a two-fold knowledge transfer model. Although our approach is proposed for optimal site selection, it is applicable to deal with different problems of chain business, such as pricing, product positioning, marketing strategy, targeted advertising, activity organization, etc. For other problems in urban computing, there might be more aspects of complementary knowledge that can be transferred and integrated. We thus intend to extend our model on integrated knowledge transferring and apply it to different application areas. For instance, for gesture recognition to a new user, we can transfer knowledge from both the target user's activity learning model as well as other users' gesture recognition model.

## 7. CONCLUSION

In this paper, we have proposed a two-fold knowledge transfer model called *CityTransfer* for chain store recommendation using multi-source urban data. It aims to solve the cold-start problem when a chain enterprise wants to expand its market in a new city. *CityTransfer* introduces knowledge transfer to collaborative filtering and significantly extend the SVD-based CF model, which make it possible to transfer the enterprise bias knowledge from source cities and learn local characterizing features from similar enterprises in the target city. Experimental results show that our method outperforms the baselines, from different aspects, on recommending optimal chain store sites. As for future works, we will on one hand to extend our model to deal with the data insufficiency issues, and on the other hand, investigate the usage of our method in other business scenarios or application areas and develop more general integrated-knowledge-transfer models.

## ACKNOWLEDGMENTS

This work was partially supported by the National Basic Research Program of China (No.2015CB352400), the National Natural Science Foundation of China (No. 61772428, 61332005, 61373119, 61402369).

## REFERENCES

[1] Y. Anzai, "Pattern recognition and machine learning," Elsevier Press, 2012.
[2] L. Baltrunas, B. Ludwig, and F. Ricci, "Matrix factorization techniques for context aware recommendation," in *Proc. of the ffth ACM conference on Recommender systems*, 2011, pp. 301-304.
[3] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Proc. of Noise reduction in speech processing, Springer*, 2009, pp. 1–4.
[4] Y. Bengio et al., "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009.
[5] B. Berman, J. R. Evans, and J. R. Lowry, "Retail Management: A Strategic Approach," *Prentice-Hall Press, London*, 1995.

[6] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem," *Knowledge-Based Systems*, vol.26, no. 2, pp. 225-238, 2012.
[7] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proce. of COMPSTAT'10,* Springer, pp. 177-186, 2010.
[8] L. Chen, D. Zhang, G. Pan, X. Ma, D. Yang, K. Kushlev, W. Zhang, and S. Li, "Bike sharing station placement leveraging heterogeneous urban open data," in *Proc. of ACM UbiComp'15,* pp. 571-575. 2015.
[9] W. Chu and S. T. Park, "Personalized recommendation on dynamic content using predictive bilinear models," in *Proc. of WWW'09*, ACM, 2009.
[10] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse Autoencoder-based feature transfer learning for speech emotion recognition," in *Proc. of IEEE Affective Computing and Intelligent Interaction (ACII'13),* 2013, pp.511–516.
[11] Z. Fan, X. Song, R. Shibasaki, T. Li, and H. Kaneda, "CityCoupling: bridging intercity human mobility," in *Proc. of ACM UbiComp'16*, 2016, pp.718–728.
[12] Z. Gantner, et al., "Learning attribute-to-feature mappings for cold-start recommendations," in Proc. of IEEE ICDM'10, 2010, pp. 176-185.
[13] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.
[14] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Computing Surveys (CSUR)*, vol. 48, no. 1, pp. 1-31, 2015.
[15] R. Guttman, A. Moukas, and P. Maes, "Agent-mediated electronic commerce: A survey," *Knowledge Engineering Review*, vol. 13, no. 3, pp. 147-159, 1998.
[16] J. Han, J. Pei, and M. Kamber, "Data mining: concepts and techniques," Elsevier Press, 2011.
[17] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via cross-domain triadic factorization," in *Proc. of ACM WWW'13*, 2013, pp. 595–606.
[18] P. Jensen, "Network-based predictions of retail store commercial categories and optimal locations," *Physical Review E*, vol. 74, no. 3, pp. 1-3, 2006.
[19] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-spotting: mining online location-based services for optimal retail store placement," in *Proc. of ACM SIGKDD'13*, 2013, pp.793–801.
[20] H. A. Karimi, Advanced location-based technologies and services, CRC Press, 2013.
[21] J. A. Konstan, et al., "GroupLens: applying collaborative filtering to Usenet news," *Communications of the ACM*, vol. 40 no. 3, pp. 77-87, 1997.
[22] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative fltering model," in *Proc. of ACM SIGKDD'08*, 2008, pp. 426–434.
[23] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30-37, 2009.
[24] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong, "Addressing cold-start problem in recommendation systems," in *Proc. of the 2nd international conference on Ubiquitous information management and communication*, 2008, pp. 208-211.
[25] R. Lee, S. Wakamiya, and K. Sumiya, "Urban area characterization based on crowd behavioral lifelogs over Twitter," *Personal and ubiquitous computing*, vol. 17, no. 4, pp. 605-620, 2013.
[26] J. Li, B. Guo, Z. Wang, M. Li, and Z. Yu, "Where to place the next outlet? harnessing cross-space urban data for multi-scale chain store recommendation," in *Proc. of ACM UbiComp'16: Adjunct*, 2016, pp.149-152.
[27] Y. Li, et al., "Location selection for ambulance stations: a data-driven approach," in *Proc. of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, pp. 1-4.
[28] E. Malmi, T. M. T. Do, and D. Gatica-Perez, "Checking in or checked in: comparing large-scale manual and automatic location disclosure patterns," in *Proc. of the 11th ACM International Conference on Mobile and Ubiquitous Multimedia*, 2012, pp. 1-10.
[29] A.Ng, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, 2011, pp. 1-19.
[30] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, 2010, pp. 1345–1359.
[31] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," *The adaptive web*, Springer, Berlin, Heidelberg, 2007, pp. 325-341.
[32] D. Quercia, L. M. Aiello, and R. Schifanella, "The Emotional and Chromatic Layers of Urban Smells," in *Proc. of ICWSM'16*, 2016, pp. 309-318.
[33] A. I. Schein, et al., "Methods and metrics for cold-start recommendations," in Proc. ACM SIGIR'02, 2002, pp. 253-260.
[34] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, no. 4, 2009, pp. 1-19.
[35] G. H. Toroslu, "*A singular value decomposition approach for recommendation systems*," The graduate school of natural and applied sciences of middle east technical university, Ph.D. Dissertation, 2010.
[36] L. H. Ungar and D. P. Foster, "Clustering methods for collaborative filtering," in *Proc. of the Workshop on Recommendation Systems*, AAAI Press, 1998.
[37] Y. Wei, Y. Zheng, and Q. Yang, "Transfer knowledge between cities," in *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1905-1914.
[38] Y. Wei, Y. Zhu, C. W. Leung, Y. Song, and Q. Yang, "Instilling social to physical: Co-regularized heterogeneous transfer learning," in *Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*, 2016, pp. 1338-1344.

[39] Z. Yu, M. Tian, Z. Wang, B. Guo, and T. Mei, "Shop-Type Recommendation Leveraging the Data from Social Media and Location-Based Services," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 1, 2016, pp. 1-21.
[40] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp.186–194.
[41] V. W. Zheng, E. W. Xiang, Q. Yang, and D. Shen, "Transferring Localization Models over Time," in *Proc. of AAAI*'08, 2008, pp. 1421-1426.
[42] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, 2014, pp. 1-55.

Received May 2017; revised July 2017; accepted October 2017.