Expert Systems With Applications 170 (2021) 114537

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multi-criteria tensor model for tourism recommender systems

Minsung Hong [a], Jason J. Jung [b,*]

[a] Big Data Research Group, Western Norway Research Institute, Box 163, NO-6851 Sogndal, Norway
[b] Department of Computer Engineering, Chung-Ang University, 84 Heukseok-ro, Heukseok-dong, Dongjak-gu, Seoul, South Korea

ARTICLE INFO

Keywords:
Tourism recommender system
Multi-criteria recommendation
Cultural difference
User preference modeling
Multi-dimensional model
Tensor factorization

ABSTRACT

Many tourism recommender systems have been studied to offer users the items meeting their interests. However, it is a non-trivial task to reflect the multi-criteria ratings and the cultural differences, which significantly influence users' reviews of tourism facilities, into recommendation services. This paper proposes two "single tensor" models, consisting of users (or countries), items, multi-criteria ratings, and cultural groups, in order to consider *simultaneously* an inherent structure and interrelations of these factors into recommendation processes. With one Tripadvisor dataset, including 13 K users from 120 countries, experiments demonstrated that, in terms of MAE, the two proposed models for user and country give an improvement of 21.31% and 7.11% than other collaborative filtering and multi-criteria recommendation techniques. Besides, there were the positive influences of multiple-criteria ratings and cultural group factors on recommendation performances. The comparative analysis of several variants of the proposed models showed that considering Western and Eastern cultures is appropriate for improving predictive performances and their stability.

1. Introduction

As the valuable information available on the Internet and the number of its users have increased hugely in the last decade, the amount of information provided to any query on the Web using a search engine or other application is often overwhelming. In turn, users need a lot of energy and time to find information useful for them. Intelligent systems using personalized information have been studied as a way to cope with this overload and provide an intellectually manageable number of possible recommendations (Walek & Fojtik, 2020). In the tourism industry, such recommender systems automatically extract tourists' preferences through analysis of their explicit or implicit feedback and match the features of tourism items with their needs (Cai, Lee, & Lee, 2018; Esmaeili, Mardani, Golpayegani, & Madar, 2020).

Collaborative filtering (CF) is one of the most well-known and frequently used methods to recommend items in various fields. Traditional CFs are typically based on a single type of rating score. Whereas, in the case of restaurant or hotel recommendation, ratings of multiple aspects (e.g., overall, staff, service, or atmosphere) can often be collected and reflect various characteristics of the restaurant or hotel (Turk & Bilge, 2019). Indeed, in online review platforms such as Tripadvisor, Hotel.com, and Booking.com, restaurants or accommodations are often evaluated for multiple aspects, unlike movies and books. Such multiple rating data is a source of rich information to provide personalized restaurant or hotel recommendations (Fu, Liu, Ge, Yao, & Xiong, 2014). However, it is a non-trivial task to reflect the multiple ratings into recommendation services due to the unique features of multi-aspect user reviews. Moreover, the task becomes more complicated when the features have inter-relation with other factors such as spatial and temporal context (Salehan, Zhang, & Aghakhani, 2017; Viktoratos, Tsadiras, & Bassiliades, 2018; Zhang, Salehan, Leung, Cabral, & Aghakhani, 2018; Wang & Yi, 2019). In order to use multiple factors in a recommendation, multi-criteria recommender systems have been studied. However, most of the existing research (Adomavicius, Manouselis, & Kwon, 2011; Zheng, 2017) considered multi-criteria independently or sequentially. Whereas, we simultaneously reflect the multi-criteria ratings in rating prediction by using a "single tensor" that keeps an inherent structure of and interrelations between multi-criteria directly.

On the other hand, according to Lee (2016), cultural difference is often considered as a barrier to technology transfer. Moreover, Jung, Lee, Chung, and tom Dieck (2018) pointed out that information systems in the tourism industry are mostly affected by cultural factor. Therefore, researchers (Chen & Pu, 2008; Tang, Winoto, & Ye, 2011; Berkovsky, Taib, Hijikata, Braslavski, & Knijnenburg, 2018) have examined cultural influences on recommender systems. Hofstede (1980) defined "culture" as "the collective programming of the mind which distinguishes the members of

* Corresponding author.
E-mail addresses: msh@vestforsk.no (M. Hong), j2jung@gmail.com (J.J. Jung).

https://doi.org/10.1016/j.eswa.2020.114537
Received 25 April 2020; Received in revised form 23 November 2020; Accepted 25 December 2020
Available online 30 December 2020
0957-4174/

one human group from another." Moreover, he and his colleague distinguished countries by five cultural dimensions: masculinity/femininity; power distance; time orientation; uncertainty avoidance; and individualism/collectivism (Hofstede & Bond, 1988). Among these five dimensions, recommender systems relate to uncertainty avoidance and collectivism (Hong, An, Akerkar, Camacho, & Jung, 2019). The uncertainty avoidance links with the purpose of recommender systems to reduce overloading information and alleviate uncertainty on decision making (Choi, Lee, Sajjad, & Lee, 2014). Collectivism associates with the recommendation functionality based on a collaborative filtering algorithm that utilizes the preferences of users similar to an active user. However, few studies have focused on the cultural influence in the item recommendation despite its substantial impact. Even existing studies (Choi et al., 2014; Chen & Pu, 2014; Chu & Huang, 2017) mainly analyze cultural differences in recommendation results through user surveys, rather than applying the cultural factors to user preference modeling and rating prediction. In contrast, the proposed model considers cultural differences directly and uses them in rating prediction.

In summary, there have been few studies that apply cultural differences to user preference modeling and evaluate its impacts on recommendation performance. To the best of our knowledge, this study is the first research work to consolidate multi-criteria ratings and a cultural factor into a single model for tourism recommendation. Moreover, the proposed model enables us to preserve the inherent structure of and interrelations between various factors (i.e., the multiple ratings and cultural factor). By using tensor factorization, the proposed model is approximated to predict user ratings for multi-criteria. Lastly, cultural differences are analyzed via the results of experiments designed to consider the cultural factor. In this regard, our primary contributions are as follows.

- Contrary to other related work, we consolidate user preferences along with multiple rating and cultural factor simultaneously.
- The proposed model outperforms other well-known techniques in terms of the recommendation prediction and its stability.
- This model can be easily applied to other domains such as hotel and point-of-Interest recommendations if the multiple ratings for a recommended item are collected.
- Like other related work, experimental results show that classifying cultures into Western and Eastern groups is an effective manner, especially to improve recommendation performance in this study.

The rest of this paper is organized as follows: Section 2 reviews relevant studies associated with cultural differences and multi-criteria in recommender systems. In Section 3, the models for consolidating multiple user preferences and cultural factor are defined, and tensor factorization-based prediction is described. Section 4 presents experimental methodology such as a used dataset and compared techniques. In Section 5, we evaluate several proposed models in various views and then discuss their performances by comparing them with other techniques. Lastly, Section 6 concludes with open challenges.

Table 1
Summary of other related and this studies.

Reference	Evaluation factor	Evaluation measure	Data collection
(Chen & Pu, 2008)	Rec. interfaces	TLX	Online survey
(Chen & Pu, 2014)	Rule-based interfaces	TLX	Online survey
(Tang et al., 2011)	Rec. prediction	Spearman correlation	Online survey
(Choi et al., 2014)	Rec. quality & trust	ANOVA & Duncan's test	Interview
(Berkovsky et al., 2018)	Rec. interfaces	Chi-square test	Online survey
(Chu & Huang, 2017)	Item multi-criteria	Rating score analysis	Web crawler
(Hong et al., 2019)	Rec. prediction	RMSE, cosine similarity	Web crawler
This study	Rec. prediction	RMSE, MAE	Web crawler

Rec. indicates recommendation.

2. Related work

Cultural difference has been conceived as one of the essential factors in the tourism research field. According to Jung et al. (2018), tourists' cultural background relates to the experience they want. In addition, tourism supported by information technology is becoming more international along with the increasing number of travelers from various countries (Li, 2014). Likewise, tourists from different cultures have a variety of rating behaviors (Chu & Huang, 2017). Many studies have shown cultural differences' influences in recommender systems (Tang et al., 2011; Chen & Pu, 2014; Choi et al., 2014; Berkovsky et al., 2018). Chen and Pu (2008) proposed an organization-based recommendation interface following five design principles that had been elicited through an evaluation of 13 interface prototypes (Pu & Chen, 2006). They analyzed cultural differences by comparing the interface with typical list views for recommended items. Analyzing a survey with 120 volunteers from five countries (China, Switzerland, France, Italy, and Germany) showed that when using the item view interface people from various cultures behave differently. Besides, the interface was extended based on an association rule mining to reflect the current interests and potential needs of users (Chen & Pu, 2014). They measured a task load index (TLX) factor by using an eye-tracker, and their experiments verified that the user preferences of participants from Eastern and Western cultures differ. Tang et al. (2011) proposed a collaborative filtering-based multi-domain recommendation for movies, TV series, books, music, and games. Experimental results from 333 students from China and Hong Kong showed that cultural differences affect users' preferences in the cross-domain recommendation. Choi et al. (2014) analyzed the impact of cultural collectivism and uncertainty avoidance in a theoretical mobile recommender system. They evaluated various factors (e.g., users' previous App purchases, and use patterns) with 461 participants from three countries (Korea, China, and the UK). Their experiments pointed out that cultures influence the two cultural dimensions in the system differently. Berkovsky et al. (2018) investigated user perceptions of the presentation, explanation, and priority of recommended items according to cultural differences. Their experiment with 102 participants from four countries (France, Japan, Russia, and the USA) showed that cultural differences impact user preferences on the recommendation presentation and explanation. Chu and Huang (2017) extended heterogeneous hotel data collected from Tripadvisor to large-scale hotel information. They predicted users' ratings of hotels based on an extended Matrix Factorization (MF) combined with other factors, such as date, price, nationality, comment, and the visual concept. Their experiment demonstrated the influence of culture on the different rating behaviors of various hotel features (e.g., business service, check-in, cleanliness, overall, location, rooms, service, and sleep quality). Our previous work considered around 80 countries in order to analyze the cultural influence and compute cultural differences between the countries in the recommendation services (Hong et al., 2019). Experimental results showed the cultural differences affecting users' rating behaviors and the effectiveness of recommendations based on cultural groups. To avoid misleading the valuable contributions of the mentioned researchers, note that we only reviewed the results related to cultural differences and factors since others are out of scope in this paper.

On the one hand, traditional CF techniques use a single rating as element of the two-dimensional item-user matrix. Such techniques focus on one type (i.e., overall) of ratings. Although the single rating-based approach shows a smooth and satisfying performance, its accuracy has been perceived to be relatively lower than multi-criteria

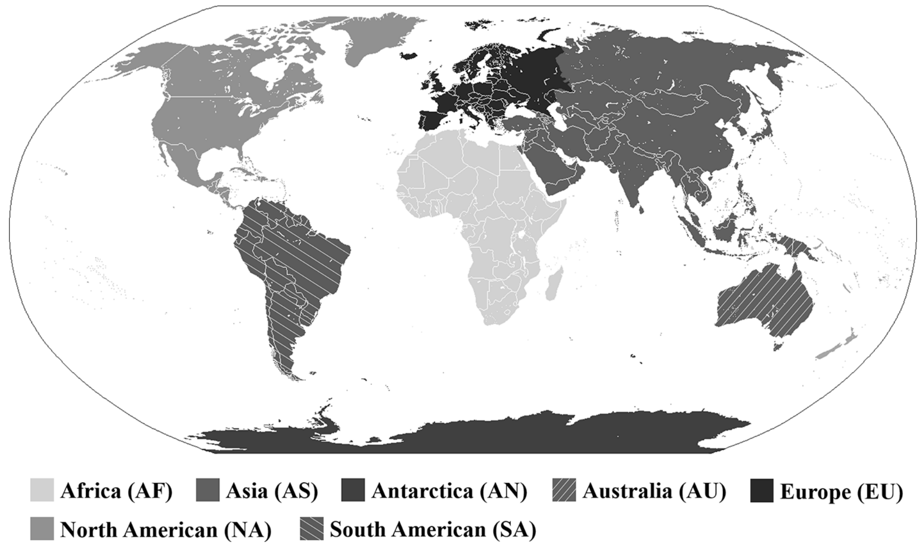

Fig. 1. Geographic continent category (McArthur et al., 2018).

recommendation techniques (Lakiotaki, Matsatsinis, & Tsoukiàs, 2011; Farokhi, Vahid, Nilashi, & Ibrahim, 2016; Zhang et al., 2018). Thus, several approaches for multi-criteria recommender systems have been studied and can be classified into memory-based and model-based methods, like CF techniques. In the memory-based method, user similarities are mainly computed in two ways. One computes similarities for each criterion and combines them into a single similarity based on aggregation methods (e.g., average and weighted sum) (Adomavicius & Kwon, 2007). The other approach directly calculates distances by using multi-dimensional distance metrics (e.g., Euclidean and Manhattan). The model-based approach models user preferences to predict missing ratings. There are two representative methods: aggregation function and multi-linear singular value decomposition (MSVD). The aggregation function method is based on the idea that an overall rating is the combination of other multi-criteria ratings. This method has three steps: estimating ratings by criteria, designing an aggregation function, and predicting overall ratings based on the function. The MSVD approach builds a predictive model for multi-dimensional ratings and is based on the assumption that criteria depend on each other (Hassan & Hamada, 2017).

Table 1 summarizes the aforementioned studies on the impact of cultural differences in recommender systems. Regarding the evaluation factor affected by the cultures, half of these studies focus on recommendation interfaces and quality, rather than recommendation prediction. Furthermore, the majority of related studies evaluated cultural impacts through statistical analysis of questionnaire data. Whereas our study *directly applies the cultural differences into the recommendation modeling of real-world data* and evaluates prediction results according to cultural groups. In addition, most of the above-mentioned research analyzed cultural differences between a few countries (less than 10), except for our previous research. In this study, we consider data from 120 countries and focus on the prediction of users' preferences for 2,000 restaurants. Besides, a single type of tensor is proposed to keep an inherent structure and model interrelations between multi-criteria in recommendation processes. The proposed model consists of not only the multiple ratings but also a cultural factor.

Regarding recommendation algorithms, the proposed method basically belongs to the MSVD. We compare it with one traditional and two

Table 2
Continent models.

Category	Name	Continent						
7 Continents[1]	C7	AF	AS	EU	NA	SA	AN	AU
6 Continents[2]	C6	AF	AS	EU	Americas		AN	AU
5 Continents[3]	C5	AF	Eurasia		Americas		AN	AU
4 Continents (McColl, 2014)	C4	Afro-Eurasia			AC		AN	AU
2 Continents[4,5]	C2	Eastern		Western				

[1] https://www.nationalgeographic.org/encyclopedia/Continent/
[2] https://en.wikipedia.org/wiki/Continent
[3] https://en.wikipedia.org/wiki/Continent
[4] https://en.wikipedia.org/wiki/Eastern_world
[5] https://en.wikipedia.org/wiki/Western_world

recent multi-criteria recommendation techniques and five traditional CFs. The traditional technique is the aggregation function approach proposed by Adomavicius et al. (2011). It uses the singular value decomposition (SVD) to predict ratings of extra criteria individually and aggregates them into overall ratings. Two recent methods are the aggregation function and the MSVD proposed by Zheng (2017). One is based on the assumption that multi-criteria ratings are independent of each other, the other assumes that multi-criteria have sequential dependencies. Thus, it sequentially predicts the ratings of multi-criteria based on the other criteria. These methods are described in Section 4.2. Note that the proposed model includes a cultural factor and multi-criteria ratings, and considers their interdependencies, while the other baselines use only overall or multi-criteria ratings.

3. Multi-criteria tensor model

This section introduces consolidation models to reflect multi-criteria and cultural factors in the context of restaurant recommendation. A tensor factorization predicting missing ratings in the models is also explained.

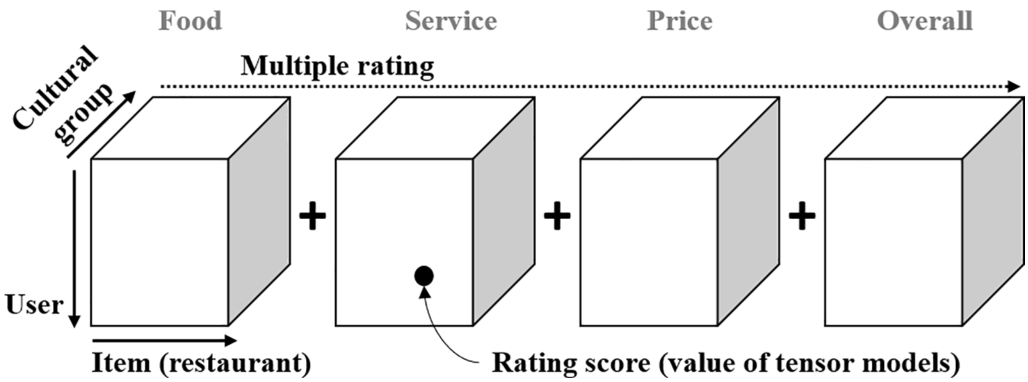

Fig. 2. Proposed tensor model.

3.1. Tensor modeling

User experiences in restaurants can be classified into tangible food and intangible service components (Bojanic & Rosen, 1994). Restaurant service providers interact with their customers through not only price, quality, and variety of food but also other factors such as the service quality and speed, as well as the dining atmosphere (Salehan et al., 2017; Zhang et al., 2018). An unpleasant experience from lousy service may affect customers' perceptional, emotional, and physiological responses negatively. In turn, the experience would lead to adverse reviews of the restaurant and its food (Yüksel & Yüksel, 2003; López & Farzan, 2015). In summary, the reviews given by users are complex and related to multiple aspects of this process. Additionally, external factors (e.g. demographics and weather) can significantly influence the judgment (Bakhshi, Kanuparthy, & Gilbert, 2014). In this regard, we take into account, not only food, service, and price (i.e. internal factors) but also cultural groups of users (i.e. an external factor).

To classify user nationality into cultural groups, geographic continents from definitions of National Geographic[1] and Oxford (McArthur, Lam-McArthur, & Fontaine, 2018) are considered as shown in Fig. 1. As a result, various continent models listed in Table 2 are applied to construct our user preference models.

Traditional CF techniques are generally based on overall ratings modeled by a user-item matrix. Therefore, they cannot consider internal and external factors as aforementioned and may fail to represent the latent generative structure of user reviews of restaurants comprehensively (Fu et al., 2014). A tensor has been used to model a multi-dimensional structure of data in recommender systems as it can preserve the interdependency between multiple factors such as users, items, contexts, and so on (Frolov & Oseledets, 2017; Hong & Jung, 2018; Hong, Akerkar, & Jung, 2019). Thus, we use the 4-order tensor to model the multi-criteria and cultural factors of restaurants' users. In other words, the proposed model enables us to consider the latent interrelations between the multiple rating and cultural group factors simultaneously, as shown in Fig. 2.

The illustrated tensor model consists of four dimensions as follows:

- the "user" indicates users or user groups (i.e. countries) requesting a recommendation in our study.
- The "item" dimension denotes restaurants.
- The "multiple rating" marked by dots contains four multi-criteria ("food," "service," "price," and "overall").
- The "cultural group" is one of the continent models listed in Table 2.
- Lastly, the element value of the tensor model is a rating score given by users of restaurants in terms of a criterion. This integer value is ranged from 1 to 5, where 1 and 5 denote the most negative and positive reviews, respectively.

3.2. Tensor factorization

This section describes tensor factorization to predict unobserved users' preferences for restaurants. To simplify the explanation, we use the 3-order tensor model in the following equations and an algorithm. Given I users (or countries), J items (i.e., restaurants), and K cultural groups, the proposed model $\mathscr{S}$ is defined by

$$\mathscr{S} = \left\{S_{ijk}\right\} \in \mathbb{R}^{I\times J\times K}, \tag{1}$$

where the value S_{ijk} indicates a rating of the i^{th} user from the k^{th} culture group for the j^{th} restaurant.

Like a conventional tensor factorization, we aim to minimize loss between observed and approximate tensors and regularization risks. Given original and approximate ratings (i.e., S_{ijk} and $\widehat{S}_{ijk}$) of users to restaurants, the objective is to minimize a loss function $L(S_{ijk}, \widehat{S}_{ijk})$. For better generalization performance, a regularization term $\Omega(S_{ijk})$ is also added. Thus, our objective function $F()$ is $L(S_{ijk}, \widehat{S}_{ijk}) + \Omega(S_{ijk})$ (Karatzoglou, Amatriain, Baltrunas, & Oliver, 2010). Least squares loss function and Frobenius norm are used as standard choices for the $L()$ and Ω. A tensor-matrix multiplication operator is expressed by $\times_U$, where subscript denotes the tensor direction on which the matrix is multiplied. In addition, the i^{th} row's entries of the matrix U are denoted by U_{i*}. Therefore, the objective function of tensor factorization is defined as follows:

$$F(\mathscr{S}, \mathscr{C}, U, R, G) = 1/2\|\mathscr{C}\times_U U\times_R R\times_G G - S\|_F^2 + 1/2\left[\lambda_U\|U\|_F^2 + \lambda_R\|R\|_F^2 + \lambda_G\|G\|_F^2\right], \tag{2}$$

where $\mathscr{C}$ is a core tensor; $\|\cdot\|_F^2$ indicates the Frobenius norm; $U \in \mathbb{R}^{I\times d_U}$, $R \in \mathbb{R}^{I\times d_R}$, and $G \in \mathbb{R}^{I\times d_G}$ are latent factors of users, restaurants, and cultural groups; d_U, d_R, and d_G are the numbers of latent features; λ_U, λ_R, and λ_G as the regularization parameters are equally set.

Due to the absence of a closed-form solution for the minimization of Eq. (2), the objective function is minimized by Stochastic Gradient Descent (SGD) for each factor with fixing the others. Algorithm 1 shows the procedures of tensor factorization by using Higher Order Singular Value Decomposition (HOSVD) (Karatzoglou et al., 2010), where the gradients of our objective function are calculated as follows:

$$\begin{aligned} \eta\partial_{U_{i*}}F^l &= \left(\widehat{\mathscr{S}}_{ijk} - \mathscr{S}_{ijk}\right) \times \mathscr{C}\times_R R_{j*}\times_G G_{k*}, \\ \eta\partial_{R_{j*}}F^l &= \left(\widehat{\mathscr{S}}_{ijk} - \mathscr{S}_{ijk}\right) \times \mathscr{C}\times_U U_{i*}\times_G G_{k*}, \end{aligned} \tag{3}$$

$$\eta\partial_{G_{k*}}F^l = \left(\widehat{\mathscr{S}}_{ijk} - \mathscr{S}_{ijk}\right) \times \mathscr{C}\times_U U_{i*}\times_R R_{j*}.$$

[1] National Geographic https://www.nationalgeographic.org/encyclopedia/Continent/

Algorithm 1: Tensor factorization

Data: observed tensor $\mathcal{S}$, learning rate t_0, regularization parameters $\lambda = \lambda_U = \lambda_R = \lambda_G$
Result: approximate tensor $\hat{\mathcal{S}}$

Initialize $\mathcal{S}$, $\mathcal{C}$, U, R, G with zero ;
Set $l = 0$ and $t = t_0$;
while *not converged* **do**
  $\eta = 1/\sqrt{t}$ and $t = t + 1$;
  for *each* $\mathcal{S}_{ijk} \neq 0$ **do**
    $U_{i*} \leftarrow U_{i*} - \eta \partial_{U_{i*}} F^l$;
    $R_{j*} \leftarrow R_{j*} - \eta \partial_{R_{j*}} F^l$;
    $G_{k*} \leftarrow G_{k*} - \eta \partial_{G_{k*}} F^{l++}$;
    Compute the l^{th} iteration of the objective function F^l by Eq. (2) ;
  end
end
Return $\hat{\mathcal{S}} = \mathcal{C} \times_U U \times_R R \times_G G$;

This algorithm linearly scales to the number of rating scores R and the dimensionalities I,J, and K of the user, restaurant, and cultural group factors. Therefore, the time complexity of the proposed algorithm is $\mathcal{O}(RIJKM)$ actually, where M is the number of the 'multiple rating' item (i.e., 4).

4. Experimental methodology

4.1. Dataset

This section compares our real-world dataset with the other datasets of the studies reviewed in Section 2. Table 3 shows the statistical information of all the mentioned datasets. The "Ques." means that the dataset is gathered by a questionnaire. Half of the relevant studies (excluding our previous work) haven't used rating scores generated by users to analyze cultural differences in recommender systems. Although two other works used rating scores, they only considered a small number of countries and overall ratings. Our previous study used user ratings from more than 80 countries, but a cultural factor was not directly applied to user preference model. Furthermore, the study considered a single type of ratings (i.e., overall rating) into restaurant recommendations.

The dataset used here was collected from Tripadvisor. According to Alexa traffic statistics[2], Tripadvisor shows continuously increasing traffic and is ranked in the top 300 websites having highest traffic volume in the world. The dataset contains 36,795 reviews of 13,620 users of 2,000 restaurants and includes the approximate nationalities of the users. Since we consider four multi-criteria ratings, the proposed models have higher sparsities than the other compared methods based on a user-item matrix. For example, if the tensor model consists of the users, restaurants, multiple ratings, and two cultural groups, its sparsity is very high - 99.976% (density is 0.024%). According to Singh (2018), this sparsity is natural in a real-world situation but hampers predictive performances.

4.2. Comparison methods

This section explains the other techniques compared in our experiments and various proposed models according to considered factors. The comparison techniques are as follows.

- K-Nearest Neighbors-based CF (KNN): is one of the basic CF algorithms. The prediction of $\hat{r}_{ui}$ is defined by $\hat{r}_{ui} = \sum_{v\in\sum_i^k(u)} sim(u,v)\cdot r_{vi} / \sum_{v\in\sum_i^k(u)} sim(u,v)$, where k is the number of neighbors and $sim()$ denotes a similarity function.
- Co-clustering-based CF (COC) (George & Merugu, 2005): assigns users and items into some clusters, respectively. The missing rating scores of an active user are then predicted based on ratings of the others belonging to the same cluster with the user. The approximate rating $\hat{r}_{ui}$ is calculated as follows: $\hat{r}_{ui} = \overline{C_{ui}} + (\eta_u - \overline{C_u}) + (\eta_i - \overline{C_i})$, where the $\overline{C_{ui}}, \overline{C_u}$ and $\overline{C_i}$ are the average ratings of co-cluster C_{ui}, u's cluster and i's cluster, respectively.
- Singular Value Decomposition-based MF (SVD): has been popularized by Simon Funk during the Netflix Prize[3]. The prediction $\hat{r}_{ui}$ is set as: $\hat{r}_{ui} = \eta + b_u + b_i + q_i^T p_u$. If user u or item i is unknown, then the biases b_u or b_i and the factors p_u or q_i are assumed to be zero.
- SVD++-based MF (SVD++) (Koren, 2008): is an extension of the SVD considering implicit ratings. The prediction $\hat{r}_{ui}$ is calculated by $\hat{r}_{ui} = \eta + b_u + b_i + q_i^T * (p_u + |I_u|^{-1/2} \sum_{j\in I_u} y_j)$, where the y_j term indicates a new set of item factors capturing implicit ratings. Also, an implicit rating means that a user u rated an item j, regardless of the rating value.
- Non-negative Matrix Factorization (NMF) (Luo, Zhou, Xia, & Zhu, 2014): is similar to SVD. The predicted rating $\hat{r}_{ui}$ is set as: $\hat{r}_{ui} = q_i^T p_u$, where user and item factors keep positive. For optimization, a (regularized) stochastic gradient descent is used.
- Aggregation-based Multi-Criteria Recommendation (AMCR) (Adomavicius et al., 2011): is a model-based approach built on the SVD technique. It consists of three steps. The first step predicts missing ratings for each criterion through MF based on the SVD. And then, the relationships (i.e., coefficients) between overall ratings and the rating scores for the other criteria are estimated. Lastly, the other criteria' predicted ratings are aggregated into approximate overall ratings by using the coefficients. Note that we used linear, Ridge, and Lasso regressions to obtain the best coefficient.
- Criteria-Independent Contextual (CIC) and Criteria-Chains Contextual (CCC) methods (Zheng, 2017) are based on context-aware matrix factorization (CAMF-C) proposed by Baltrunas, Ludwig, and Ricci (2011). The CIC predicts ratings for multi-criteria by the SVD separately and uses the CAMF-C to predict overall ratings, while the

[2] Alexa, accessed on 01/22/2020, http://www.alexa.com/siteinfo/tripadvisor.com

[3] https://sifter.org/simon/journal/20061211.html

Table 3
Summary of other dataset.

Reference	# of items	# of users	# of countries
(Chen & Pu, 2008; Chen & Pu, 2014)	Ques.	120	5
(Tang et al., 2011)	108	333	2
(Choi et al., 2014)	Ques.	461	3
(Berkovsky et al., 2018)	Ques.	112	4
(Chu & Huang, 2017)	12,773	1.3 M	8
(Hong et al., 2019)	50	15,424	81

CCC estimates multi-criteria ratings sequentially. Note that we used Pearson correlation coefficients between multi-criteria ratings to decide a sequence of respective rating predictions for the CCC.

These techniques were implemented by using Python libraries and CARSKit developed by Zheng, Mobasher, and Burke (2015). All experiments including the grid-search of optimal parameters were conducted in the same computation environment.

The proposed tensor models are divided into two types according to the "user" factor (i.e., users and countries). Besides, the models excluding the "cultural group" factor are also considered. Such variants help us to evaluate the influences of multiple ratings or/and cultural groups as well as their combinations in a restaurant recommendation service. Therefore, the variants of our models are named by MCTu/c–C#R. We basically call our models MCT (Multi-Criteria Tensor) and use suffix "u" and "c" for the user and country models. The "#" denotes the number of cultural groups, as listed in Table 2. For example, if a tensor model consists of users, restaurants, multiple ratings, and four cultural groups, we call the model MCTu-C4R. User- and country-based models excluding the cultural group factor are presented as MCTu-C0R and MCTc–C0R. The other techniques are expressed by adding small letters "u" and "c", respectively for user- and country-based models. For instance, the SVD++ for users is indicated by SVDu++.

4.3. Experimental measure

To evaluate the predictive performances of the MCT and other techniques, we use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) defined as follows:

$$RMSE = \sqrt{\sum_{i}^{N} (\widehat{a}_i - a_i)^2 / N}, and \tag{4}$$

$$MAE = \sum_{i}^{N} |\widehat{a}_i - a_i| / N, \tag{5}$$

where $\widehat{a}_i$ and a_i are predicted and observed ratings. The N indicates the number of the observed rating scores in a test set. We use the k-fold cross-validation scheme to avoid overfitting results to a test set. The k was set as 5 following Kuhn and Johnson (2013) explanation that "typically, $k = 5$ or $k = 10$ have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor very high variance".

The details of datasets used for the proposed methods (i.e., MCTu and MCTc) and comparison techniques (Comp.) are listed in Table 4.

Table 4
Statistical information of used dataset for cross-validation.

	User model		County model	
	Comp.	MCTu-C#R	Comp.	MCTc-C#R
Train set	10,544	29,436	2,683	7,363
Test set	2,635	7,359	670	1,840
Total	13,179	36,795	3,353	9,203
# of rating types	1	4	1	4
# of user/country	13,620		120	
# of item (restaurant)	2,000		2,000	

According to the numbers of users, cultural groups, and ratings considered in the models above, training and testing sets differ. Even though the datasets of comparison methods are smaller than those of the MCTs, it is fair to consider their sparsities and the datasets generated from the same dataset.

5. Evaluation and discussion

5.1. Influences of multiple rating and cultural group factors

This section assesses how multi-criteria ratings and cultural groups affect the predictive performance of restaurant recommendation. Also, we evaluate the proposed methods based on user and country models to analyze the influences of cultural differences on the recommendation performance.

5.1.1. Multi-criteria ratings

In this section, we analyze how the auxiliary ratings of extra criteria (i.e. "food," "price," and "service") impact the restaurant recommendation. These criteria are expressed by the suffixes "f," "p," and "s". For example, if a model includes "food" and "service" ratings, we call it MCT-Rfs. We do not consider the cultural factor here (i.e., # is 0). The learning rate t_0 and regularization parameters λ of the MCTs are equally set as 0.001 and 0.01. Fig. 3 shows the RMSEs and MAEs of the variant models according to the multi-criteria combinations. As a result, the MCTu-C0Rf and MCTu-C0Rp models have better performances than MCTu-C0Rs, and the MCTu-C#fp is best among the models consisting of two extra criteria. These results are natural when we see correlation coefficients between rating elements, as listed in Table 5. As presented by the bold font in the table, the "food" and "price" ratings have high correlations (around 0.8) not only with each other but also with the "overall" score. The "service" has relatively low correlations with the factors. These results show that the criteria of "food" and "price" influence predictive performances more positively than the "service" criterion. Consequently, the MCTu-C0R, which contains all the ratings, has the best performances (i.e. all the extra criteria have correlations higher than 0.65 with the overall rating). These experimental results imply that considering correlations between multi-criteria ratings before applying them to recommender systems is simple but useful.

5.1.2. Cultural groups

This section compares several MCT models according to the numbers of cultural groups in order to analyze the cultural factor's influence. The multi-criteria ratings are not considered here. We leverage the MCTu-C0, which includes only user and item factors, as a baseline. Table 6 lists the RMSEs and MAEs of the models according to the cultural groups' numbers. In general, the RMSE has a lower bound equal to the MAE, and its upper bound tends to be increased more than MAE as the test sample size rises. Thereby, the comparison of RMSEs for different sizes of test sets can cause misinterpretation of the experimental results. However, the problem does not occur in our experiment, since we use the same number of observed overall ratings in test sets, as shown in Table 4. Additionally, RMSE tends to increase by high variance associated with the frequency distribution of error magnitudes (Willmott & Matsuura, 2006). Note that the distribution means error magnitudes for each predicted data element and is not the standard deviation (SD) shown in Table 6. In other words, the more deviations between RMSE and MAE (i.e. "Diff." in the table), the bigger substantial error variance. It means that prediction performance is unstable. In terms of predictive stability, the MCTu-C4 has the most stable performance. Moreover, the MCTu-C5 shows strength against overfitting sub-trainings, since it has the smallest Diff.

Consequently, the MCTu-C2 shows the best performances in terms of RMSE and MAE. Compared with the MCTu-C0, the variants of the MCTu models considering the cultural factor show on average around 12.7% (44.5%) and 13.94% (48.4%) improvements in terms of RMSE (SD_{RMSE})

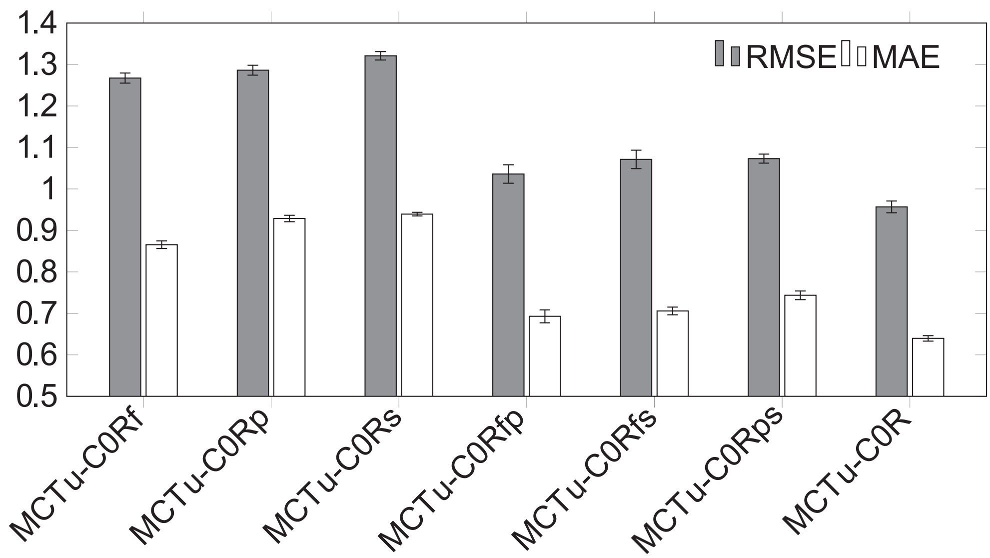

Fig. 3. Influence comparison of multi-criteria ratings.

Table 5
Correlation between multi-criteria ratings.

	overall	food	price	service
overall	1.0000			
food	**0.8699**	1.0000		
price	**0.8329**	**0.8090**	1.0000	
service	0.7872	0.6779	0.6875	1.0000

and MAE (SD_{MAE}). These results show that the "cultural group" factor positively affects the restaurant recommendation's predictive performances.

5.1.3. *Combination of multiple ratings and cultural groups*

This section analyzes how the combination of "multiple rating" and "cultural group" factors influences recommendation performances. Two cultural groups are considered, since MCTu-C2 showed the best predictive performances in the previous section. Table 7 shows the recommendation performances of the MCTu-C0R# and MCTu-C2R# according to the considered multi-criteria. The C0R and C2R models consist of all multi-criteria ratings. We calculated the RMSEs and MAEs with the same number of overall ratings. For example, the MCTu-C0Rf and MCTu-C2Rf were evaluated based on 4,182 samples. Therefore, the comparison between these two types of MCT models is fair. The MCTu-C2# has higher model sparsities than the MCTu-C0# because of the various structure of tensor models. Therefore, we need to note that it might decrease the predictive performances of the MCTu-C2#. According to the combination of multi-criteria, the MCTu-C2R# models have similar trends to the MCTu-C0R#. In other words, a model comprising multi-criteria ratings, which have high correlations, shows better predictive performances. As a result, the MCTu-C2R outperforms the other models in terms of MAE, despite its higher sparsity. Furthermore, it shows better performances than others in the SD_{MAE}, unlike the MCTu-C0R.

Table 6
Influence comparison for cultural groups.

MCTu-	C0	C2	C4	C5	C6	C7
RMSE	2.0772	**1.7969**	1.7986	1.8126	1.8239	1.8337
SD_{RMSE}	0.0252	0.0134	0.0118	**0.0105**	0.0228	0.0117
MAE	1.7041	**1.4492**	1.4571	1.4660	1.4754	1.4847
SD_{MAE}	0.0267	0.0174	0.0135	**0.0072**	0.0207	0.0101
Diff.	0.3731	0.3477	**0.3416**	0.3466	0.3485	0.3490

Table 8 shows performance differences between the models with and without the cultural factor. The values in this table are computed by $v_{C0R} - v_{C2R}$, where the v_{C0R} and v_{C2R} denote corresponding result values of MCTu-C0R# and MCTu-C2R#. Therefore, positive values mean that there are performance improvements caused by combining multi-criteria ratings with cultural groups. Consequently, the MCT-C2R shows the most outstanding improvements (around 0.0154 and 0.0039) in RMSE and its standard deviation. Besides, combining the "food" and "price" criteria with the cultural factor (i.e. MCTu-C2Rfp) improves the performance stabilities (i.e. the SD_{MAE}) more than the other MCTu-C2R#. Interestingly, the MCTu-C2Rs has the biggest improvement in terms of MAE, although it consists of the "service" criterion that has relatively low correlations with overall ratings. This result implies that the cultural factor has interrelations with rating types and may overcome lower correlations between the rating types. Such complex results demonstrate that it is a difficult task to construct a single model with various factors (i.e. multi-ratings and cultural groups), as mentioned in Section 1. In addition, the C2Rf, C2Rp, and C2Rs models show improvements in the RMSEs and MAEs when combined with the cultural factor. It demonstrates that people from different cultures might have various preferences when it comes to the multi-criteria of restaurants and could differently assess a restaurant according to their preferences for those multi-criteria. Thus, it is significant to consider cultural differences in tourism recommendations.

In summary, the combination of these two factors improves in RMSE (SD_{RMSE}) and MAE (SD_{MAE}) by on average 1.61% (27.7%) and 2.5% (28.67%). Compared with the MCTu-C2, the MCTu-C2R improves significantly (56.73% (63.51%)) in MAE and its standard deviation. Since the numbers of test samples for these two models differ, only the MAEs are considered here. These results imply that a combination of cultural factor and multiple ratings influences the restaurant recommendation positively.

5.1.4. *Cultural difference*

This section evaluates cultural differences by comparing models that contain the "cultural group" factor. Country-based models are also considered. We use the models, including all multi-criteria ratings, since they showed the best performances in previous sections. Table 9 lists the RMSE and MAE results of the MCT models according to the cultural groups listed in Table 2. Note that the numbers of test samples are equal to 7,359 and 1,840 for the MCTu-C#R and MCTc–C#R, as shown in Table 4.

The bold font expresses the best results. When an outcome is between

Table 7
Influence comparison of combined multi-criteria ratings and cultural groups.

MCTu-	C0Rf	C0Rp	C0Rs	C0Rfp	C0Rfs	C0Rps	C0R
RMSE	1.2673	1.2862	1.3210	1.0360	1.0712	1.0731	**0.9568**
SD_{RMSE}	0.0122	0.0119	**0.0101**	0.0223	0.0222	0.0109	0.0142
MAE	0.8656	0.9287	0.9393	0.6929	0.7059	0.7436	**0.6397**
SD_{MAE}	0.0093	0.0077	**0.0043**	0.0156	0.0093	0.0105	0.0066
Diff.	0.4017	0.3574	0.3818	0.3431	0.3653	0.3295	**0.3172**
Density	0.0384	0.0387	0.0389	0.0352	0.0354	0.0356	0.0338
MCTu-	C2Rf	C2Rp	C2Rs	C2Rfp	C2Rfs	C2Rps	C2R
RMSE	1.2634	1.2843	1.3184	1.0218	1.0576	1.0680	**0.9414**
SD_{RMSE}	0.0227	0.0137	0.0107	0.0190	0.0209	0.0104	**0.0103**
MAE	0.8467	0.9110	0.9202	0.6756	0.6891	0.7300	**0.6270**
SD_{MAE}	0.0150	0.0109	0.0090	0.0111	0.0099	0.0086	**0.0063**
Diff.	0.4167	0.3734	0.3983	0.3462	0.3685	0.3380	**0.3144**
Density	0.0192	0.0193	0.0194	0.0176	0.0177	0.0178	0.0169
Training set	16,731	16,855	16,936	23,043	23,124	23,248	29,436
Test set	4,182	4,213	4,234	5,760	5,781	5,812	7,359

0.003 and 0.004 higher than the best ones, it is marked with underlining. In terms of user models, most of the MCTu-C6R results are similar to those of the MCTu-C2R. The MCTu-C4R and MCTu-C6R have similar performances in the SD_{RMSE}. The SD_{MAE} of MCTu-C6R and MCTu-C7R are similar to the MCTu-C2R's one (i.e., the two models have similar resistance to the MCTu-C2R against overfitting to training sets). Regarding the Diff., the stabilities of the MCTu-C4R and MCTu-C6R outperform the MCTu-C2R. Considering the country models' RMSEs, the MCTc–C7R has similar performances to the MCTc–C2R. Also, the MAEs of MCTc–C5R, MCTc–C6R, and MCTc–C7R are similar to the MCTc–C2R's one. In terms of stability, even the MCTc–C7R has a better performance than the MCTc–C2R. Consequently, the MCTu-C2R and MCTc–C2R have lower predictive errors and show acceptable stabilities as their deviations between RMSE and MAE are similar to or lower than of the other models.

To determine a proper candidate for the "cultural group" factor in this study, we analyze differences between the MCT-C0R and MCT-C#R models. On the right side of the table, we list differences between the results of the MCTc–C0R and the MCTc–C#R. The MCTu-2C and MCTc-2C are the most proper candidates as they have the most positive influences on recommendation predictive performances. Interestingly, the MCTu-C6R and MCTc–C7R show similar improvements in terms of RMSE or MAE. However, their SD_{RMSE} and SD_{MAE} are similar to or lower than in the MCTu-C0R and MCTc–C0R. Therefore, we selected the MCTu-C2R and MCTc–C2R to be compared with other techniques in the next section. These results imply that considering two cultural groups might be an appropriate way for modeling multi-criteria ratings of restaurants' users.

5.2. Comparative analysis with other techniques

This section evaluates the proposed methods by use of comparative analysis with other techniques. It also discusses experimental results. As shown in Table 4, as the structures of used models for user preferences differ, application of the same settings to all algorithms is not appropriate. Table 10 lists optimal settings for each technique found through a grid search. Note that we used 5-fold cross-validations to find optimal settings. The MCT-C0R and MCT-C2R are represented in two kinds of model structures: the user/country-restaurant-multiple rating and user/country-restaurant-multiple rating-cultural group, respectively. Since their optimal factor sets of users and countries in the models are two or three, we could preliminarily infer that users and countries could be divided into two or three groups, as selected above. Note that the CIC's and CCC's optimal settings are for the CAMF-C, and we sequentially predicted ratings for each criterion that has high correlations for the CCC (i.e. food→price→service→overall).

Fig. 4 shows the RMSE and MAE results of each method based on users. The MCTu-C2R is superior to the other techniques (traditional CFs), as its MAE is much lower (on average 0.224) compared to the other methods. Moreover, the MAE standard deviations for this method are much smaller than for others (except for SVDu++). It means that the proposed method both performs and avoids overfitting better than the traditional CF techniques. Even the MCTu-C0R, which includes only multi-criteria ratings, is superior to them. These results imply that multiple ratings have positive influence on restaurant recommendation.

In terms of multi-criteria recommendation, the AMCRu shows high RMSE and MAE. The reason for it might the lacking consideration of individuals' correlations (i.e. personal rating behaviors) for multi-criteria, as the AMCRu obtains the coefficient from entire multi-criteria ratings. The CICu and CCCu show better RMSEs and MAEs than the CF techniques based on a user-item matrix, although their deviations are a little bigger than for the CF methods. These results mean that the multi-criteria recommendation techniques have better predictive performances but relatively lower stabilities than the traditional CFs.

Consequently, our proposed methods have not only better predictive but also more stable performances than multi-criteria recommendations. Interestingly, the RMSEs and MAEs in the recent multi-criteria methods (i.e. the CICu and CCCu) are similar to the ones in traditional MF techniques (i.e. the SVDu and SVDu++). In comparison, the proposed methods outperform the SVDu and SVDu++. These results emphasize the importance of considering interdependencies between multi-criteria

Table 8
Performance improvement of MCTu-R# through cultural group factor.

MCTu-	C2Rf	C2Rp	C2Rs	C2Rfp	C2Rfs	C2Rps	C2R
RMSE	0.0039	0.0018	0.0026	0.0142	0.0136	0.0051	**0.0154**
SD_{RMSE}	−0.0104	−0.0018	−0.0007	0.0034	0.0013	0.0005	**0.0039**
MAE	0.0189	0.0178	**0.0191**	0.0173	0.0169	0.0136	0.0126
SD_{MAE}	−0.0056	−0.0033	−0.0047	**0.0045**	−0.0006	0.0019	0.0002
RMSE-MAE	−0.0150	−0.0159	−0.0165	−0.0031	−0.0032	−0.0085	**0.0028**

Table 9
RSME and MAE of the proposed methods.

MCT	Experiment result for tensor models				Diff.	Difference with MCT-C0R			
	RMSE	SD_{RMSE}	MAE	SD_{MAE}		RMSE	SD_{RMSE}	MAE	SD_{MAE}
u-C2R	**0.9414**	**0.0107**	**0.6270**	**0.0063**	0.3144	**0.0154**	**0.0035**	**0.0127**	**0.0003**
u-C4R	0.9466	0.0129	0.6346	0.0097	**0.3120**	0.0102	0.0013	0.0051	−0.0031
u-C5R	0.9521	0.0166	0.6338	0.0101	0.3183	0.0047	−0.0024	0.0059	−0.0035
u-C6R	0.9417	0.0123	0.6297	0.0071	**0.3120**	0.0151	0.0019	0.0100	−0.0005
u-C7R	0.9459	0.0136	0.6307	0.0071	0.3152	0.0109	0.0006	0.0090	−0.0005
u-C0R	0.9568	0.0142	0.6397	0.0066	0.3171				
c–C2R	**0.9778**	**0.0177**	**0.7147**	**0.0142**	0.2631	**0.0201**	**0.0103**	**0.0083**	**0.0138**
c–C4R	0.9860	0.0307	0.7189	0.0181	0.2671	0.0119	−0.0027	0.0041	0.0099
c–C5R	0.9901	0.0276	0.7162	0.0163	0.2739	0.0078	0.0004	0.0068	0.0117
c–C6R	0.9946	0.0362	0.7175	0.0186	0.2771	0.0033	−0.0082	0.0055	0.0094
c–C7R	0.9787	0.0365	0.7181	0.0313	**0.2606**	0.0192	−0.0085	0.0049	−0.0033
c–C0R	0.9979	0.0280	0.7230	0.0280	0.2749				

simultaneously in order to improve the predictive performances (see the MCTu-C0R). As a result, when MAE and RMSE are concerned, the MCTu-C0R and MCTu-C2R improve by on average **20.05**% (16.32%) and **21.31**% (17.86%) when compared to the other techniques. It implies that the combination of the cultural group and multi-criteria factors affects restaurant recommendation positively.

Fig. 5 shows the RMSEs and MAEs when the country model-based methods are in use. The SVDc and CCCc show better RMSE performances than others, including the proposed method. Although the MAE performances of MCTc–C0R and MCTc–C2R are still better than those of the other techniques. They show similar or worse RMSEs than the SVD-based MFs, the CICc, and CCCc. Besides, the proposed methods show different results compared to the other techniques. RMSEs and MAEs in the MCTc–C0R and MCTc–C2R are worse than in the user-based models (i.e. the MCTu-C0R and the MCTu-C2R). In contrast, other country model-based techniques show better performances than using user-based models. Such results may occur due to structural damage of interrelations with the other factor (i.e., cultural groups) that are caused when we transform the user-based models into the country-based ones. In particular, since the tensor model tends to keep the inherent structure of multi-dimensional data, the defect may impact the recommendation performance. These results teach us that we need to be careful when modifying or reconstructing data models to create a tensor model that keeps the inherent structure and latent interrelations between multiple factors.

Table 10
Optimization setting of compared methods for best prediction results.

Method	Set specification for optimization
KNNu	# of neighbors: 10, similarity measure: Mean Squared Difference (MSD)
COCu	# of user and item clusters: 2 and 3, Iter. # of ALS procedure: 20
SVDu	lrS: 0.0036, regS: 0.024, # of factors: 100, Iter. # of SGD procedure: 20
SVDu++	lrS: 0.0034, regS: 0.036, # of factors: 20, Iter. # of SGD procedure: 20
NMFu	lrS: 0.005, regS: 0.006, # of factors: 15, Iter. # of SGD procedure: 50
AMCRu	lrS: 0.003, regS: 0.025, # of factors: 100, Iter. # of SGD procedure: 20
CICu	lrS: 0.001, regS: 0.001, # of factors: 20, Iter. # of SGD procedure: 50
CCCu	lrS: 0.0015, regS: 0.0015, # of factors: 30, Iter. # of SGD procedure: 50
MCTu-C0R	lrS: 0.001, regS: 0.01, # of factors: 3–2-3, Iter. # of SGD procedure: 9
MCTu-C2R	lrS: 0.001, regS: 0.01, # of factors: 3–2-3–2, Iter. # of SGD procedure: 8
KNNc	# of neighbors: 15, similarity measure: cosine similarity
COCc	# of user and item clusters: 2 and 3, Iter. # of ALS procedure: 20
SVDc	lrS: 0.0019, regS: 0.05, # of factors: 100, Iter. # of SGD procedure: 20
SVDc++	lrS: 0.0014, regS: 0.025, # of factors: 20, Iter. # of SGD procedure: 20
NMFc	lrS: 0.005, regS: 0.006, # of factors: 15, Iter. # of SGD procedure: 50
AMCRc	lrS: 0.002, regS: 0.045, # of factors: 100, Iter. # of SGD procedure: 20
CICc	lrS: 0.0015, regS: 0.0015, # of factors: 15, Iter. # of SGD procedure: 40
CCCc	lrS: 0.0015, regS: 0.0015, # of factors: 20, Iter. # of SGD procedure: 30
MCTc–C0R	lrS: 0.001, regS: 0.001, # of factors: 3–2-7, Iter. # of SGD procedure: 7
MCTc–C2R	lrS: 0.001, regS: 0.001, # of factors: 3–2-5–2, Iter. # of SGD procedure: 7

lrS and regS denote learning rates and regularization parameters.

Furthermore, we analyze performance differences between methods based on user and country models, as listed in Table 11. The values in this table are calculated by subtracting country-based models' errors from those of user-based ones. Positive values mean that a corresponding method based on a country model has better performances than the one based on a user model.

The table also shows that RMSE and MAE in the country model-based compared methods improve by, on average, 8.53% and 8.20% compared to than ones based user models. It might be due to high densities, as shown in Table 12. Indeed, the difference in the density between the user and the country model-based comparison methods (Comp.) is equal to 1.35%. In terms of RMSE and MAE, the performance decreases in the proposed methods are on average 4.08% and 13.5%. On the one hand, the SDs of most techniques built on country-based models increase greatly. This means that country-based models are over-fit to training sets. Presented results imply that the reconstruction of rating values by a simple operation (e.g. summation, average, or multiplication) might harm original data's intrinsic character. In fact, we used simple averages of restaurant ratings in the countries. However, as shown in Fig. 5 and Table 11, all the country model-based methods show stable performances according to the Diff. calculated by subtracting country models from user models. Moreover, as shown in Table 12, the country models provide other benefits for the sparsity problem and reduce the response time by decreasing user dimensionality. To make such benefits without damaging innate relations in data models, consensus functions (e.g. average without misery and least misery strategies (Amer-Yahia, Roy, Chawla, Das, & Yu, 2009; Hong, Jung, & Camacho, 2017)) of group recommender systems can be exploited by considering countries as user groups.

We also analyze the model sparsities and densities of the proposed and other methods. Traditional CF methods use the same user- (or country-) item matrix, while the AMCR, CIC, and CCC use four metrics by multi-criteria. Therefore, since the restaurant item is the same, their sparsity and density depend on the "user" factor's dimensionality. As a result, the sparsities (and densities) of their user and country models are 99.9515% (0.0485%) and 98.6015% (1.3685%). Although these sparsities are twice as big as for the proposed tensor models, they are similar to a real-world situation. 'Dif. density' denotes the subtraction of a user model's density to a country model. Its positive value means that a country-based model is denser than a user-based model. The comparison methods' density difference is 1.35%. As aforementioned and shown in Figs. 4 and 5, this relatively large density of the country model might positively influence predictive performances of compared techniques. 'RtD. Comp.' indicates the density ratios (i.e. d_p/d_c), where the "d_p" and "d_c" denote the densities of a proposed model and the compared model respectively. Although the user models' densities for the proposed methods are much lower than for the other techniques, the MCTu-C0R

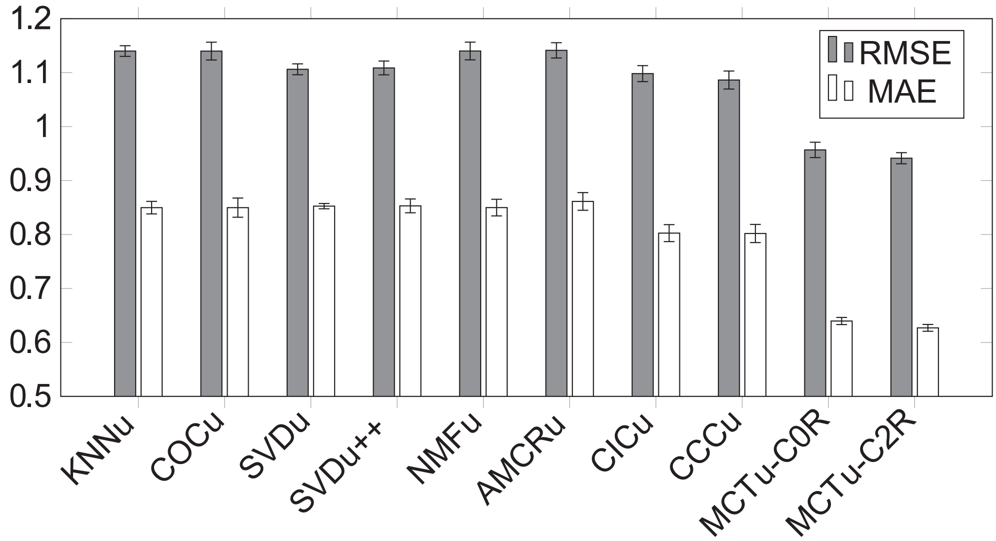

Fig. 4. Comparison of RMSE and MAE for user-based models.

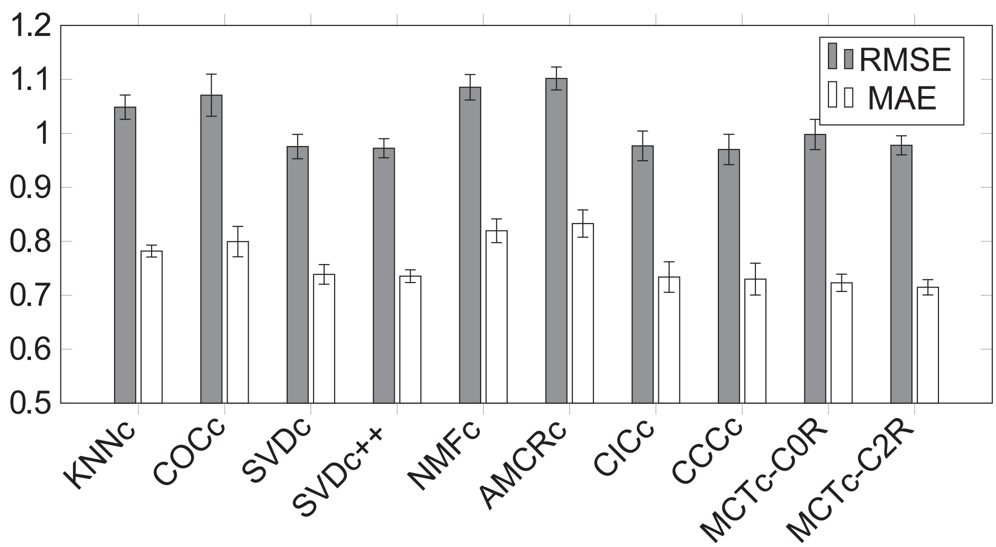

Fig. 5. Comparison of RMSE and MAE for country-based models.

Table 11
Performance difference between methods based on user and country models.

	RMSE (%)	SD_{RMSE} (%)	MAE (%)	SD_{MAE} (%)	Diff.
KNN	0.0915 (8.03)	−0.0126 (−127.27)	0.0680 (8.00)	0.0006 (5.17)	0.0235
COC	0.0692 (6.07)	−0.0225 (−136.36)	0.0504 (5.93)	−0.0103 (−57.87)	0.0188
SVD	0.1307 (11.82)	−0.0125 (−123.76)	0.1140 (13.37)	−0.0133 (−271.43)	0.0167
SVD++	0.1363 (12.29)	−0.0049 (−38.28)	0.1178 (13.81)	0.0010 (7.81)	0.0185
NMF	0.0547 (4.80)	−0.0072 (−43.90)	0.0304 (3.58)	−0.0066 (−42.86)	0.0243
AMCR	0.0396 (3.47)	−0.0071 (−50.00)	0.0285 (3.30)	−0.0089 (−54.27)	0.0111
CIC	0.1214 (11.05)	−0.0127 (−85.81)	0.0689 (8.59)	−0.0126 (−80.25)	0.0525
CCC	0.1162 (10.70)	−0.0114 (−68.26)	0.0721 (8.99)	−0.0127 (−75.60)	0.0441
MCT-C0R	−0.0411 (−4.30)	−0.0138 (−96.79)	−0.0833 (−13.03)	−0.0095 (−143.80)	0.0422
MCT-C2R	−0.0364 (−3.86)	−0.0074 (−71.82)	−0.0877 (−13.98)	−0.0079 (−123.77)	0.0513
Avg. others	0.0950 (8.53)	−0.0114 (−84.21)	0.0688 (8.20)	−0.0079 (−71.16)	0.0262
Avg. MCT	−0.0387 (−4.08)	−0.0106 (−84.30)	−0.0855 (−13.50)	−0.0087 (−133.78)	0.0468
Avg. all	0.0682 (6.01)	−0.0112 (−84.23)	0.0379 (3.86)	−0.0080 (−83.69)	0.0303

Table 12
Sparsity of tensor models and comparison methods.

MCTu-	C2R	C4R	C5R	C6R	C7R	C0R	Comp.
Sparsity (%)	99.9831	99.9916	99.9932	99.9944	99.9952	99.9662	99.9515
Density (%)	00.0169	00.0084	00.0068	00.0056	00.0048	00.0338	00.0485
RtD. Comp. (%)	34.85	17.32	14.02	11.55	9.90	69.69	
MCTc-	C2R	C4R	C5R	C6R	C7R	C0R	Comp.
Sparsity (%)	99.5207	99.7603	99.8083	99.8402	99.8631	99.0414	98.6015
Density (%)	00.4793	00.2397	00.1917	00.1598	00.1369	00.9586	01.3985
RtD. Comp. (%)	34.25	17.13	13.70	11.42	9.79	68.50	
Dif. density	0.4624	0.2313	0.1849	0.1542	0.1321	0.9248	1.3500

and MCTu-C2R showed lower RMSEs and MAEs than the others. Furthermore, this brings as benefit for the time complexity of tensor factorization. Indeed, the average response time (68.50s) for factorizing country models was much less than that (291.66s) for user models. Besides, the MCTc–C0R and MCTc–C2R have better MAEs (0.7230 and 0.7147) than the CICc (0.7337) and CCCc (0.7298). Hence, the methods for group recommender systems based on country models are also worth studying.

5.3. Additional discussion

There are still three open challenges regarding the multi-criteria recommendation combined with a cultural factor. As preliminary research, we simply considered geopolitical cultural groups, despite the importance of the "cultural group" factor in the proposed model. As studied by Hong et al. (2019), clustering is one of the useful techniques to identify more diverse and sophisticated levels of cultural groups.

In the context of the multi-criteria recommendation, it is important to note that some multi-criteria ratings are often partial or missing. In turn, it leads to a sparser model and reduction of the predictive performance. As recently many researchers (Zhang et al., 2018; Sun, Guo, & Zhu, 2019) have used sentiment analysis for recommender systems to improve recommendation performance, we are also able to complement the partial preferences with sentiment scores from the sentiment analysis.

Lastly, the temporal element has been studied a lot as one of the significant factors for tourism recommendations (Zhang et al., 2016; Ezin, Alcaraz-Herrera, & Iván, 2019). According to Gao (2016), the reviews or opinions of users change over time and by users' cultures. Therefore, temporal and cultural factors need to be investigated and analyzed in recommender systems. In this regard, the proposed tensor is one of the appropriate models to reflect these factors simultaneously, since it keeps their inherent structure and relations.

6. Conclusion

Recommender systems extract travelers' preferences regarding tourism facilities such as restaurants, hotels, and museums by analyzing their explicit or implicit feedback containing their interests. Recently, famous online review platforms for tourism items often gather multi-criteria ratings from their users who come from different cultures. However, as shown in our experiments, it is not a trivial task to reflect the multi-criteria and cultural factors into recommendation services due to their interdependencies and the unique feature of multi-aspect reviews. Furthermore, despite the significant influence of these factors, few studies have considered them in recommendation processes.

In this paper, we proposed single tensor models that consist of four dimensions (users or countries, restaurants, multiple ratings, and cultural groups) to take inter-relations of the various factors into account. The HOSVD is applied to predict missing values in the models. Several variants of the proposed methods were assessed to analyze the factors' influences on restaurant recommendation. Moreover, the predictive performances of HOSVD-based tensor factorization based on the proposed models were evaluated by comparing with well-known collaborative filtering techniques and multi-criteria recommendation methods. For the evaluations, we used a real-world dataset gathered from Tripadvisor. The dataset includes 36,795 user reviews to 2,000 restaurants in London, and the numbers of users and countries are respectively 13,620 and 120. We discussed various issues such as datasets, optimal settings, sparsity, and density, and measures for rigorous evaluations and equitable comparisons.

Experimental results showed that the proposed methods MCTu-2C and MCTc-2C outperform the compared ones (i.e. **21.31%** and **7.11%** improvements in MAE). Besides, it was shown that cultural factor have positive synergies with multi-criteria. Moreover, we learned several lessons about the risk of model reconstruction and the interrelations between multi-criteria ratings and cultural groups from systemic and various experiments. Also, the potential and benefits of the proposed method as a group recommender system were discussed. Lastly, it was revealed that considering two cultural groups (i.e., Western and Eastern cultures) is an appropriate way for improving predictive performances and their stabilities. It is also worth pointing out that the other models of cultural groups often showed better performances. Therefore, it is important to analyze and apply cultural differences in future research on recommender systems.

CRediT authorship contribution statement

Minsung Hong: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Jason J. Jung:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2017R1A2B4010774, NRF-2018K1A3A1A09078981).

References

Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems, 22*, 48–55.
Adomavicius, G., Manouselis, N., & Kwon, Y. (2011). Multi-criteria recommender systems. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 769–803). Springer.
Amer-Yahia, S., Roy, S. B., Chawla, A., Das, G., & Yu, C. (2009). Group recommendation: Semantics and efficiency. *PVLDB, 2*, 754–765.
Bakhshi, S., Kanuparthy, P., & Gilbert, E. (2014). Demographics, weather and online reviews: A study of restaurant recommendations. In *Proceedings of the 23rd international conference on World wide web* (pp. 443–454). New York, US: ACM.

Baltrunas, L., Ludwig, B., & Ricci, F. (2011). Matrix factorization techniques for context aware recommendation. In Proceedings of the 2011 ACM Conf. Recommender Systems, RecSys'11 (pp. 301–304). Chicago, IL, USA.

Berkovsky, S., Taib, R., Hijikata, Y., Braslavski, P., & Knijnenburg, B.P. (2018). A cross-cultural analysis of trust in recommender systems. In T. Mitrovic, J. Zhang, L. Chen, & D. Chin (Eds.), Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP 2018 (pp. 285–289). Singapore: ACM.

Bojanic, D. C., & Rosen, L. D. (1994). Measuring service quality in restaurants: an application of the servqual instrument. *Hospitality Research Journal, 18*, 3–14.

Cai, G., Lee, K., & Lee, I. (2018). Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos. *Expert Systems with Applications, 94*, 32–40.

Chen, L., & Pu, P. (2008). In P. Pu, D. G. Bridge, B. Mobasher, & F. Ricci (Eds.), *A cross-cultural user evaluation of product recommender interfaces Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys 2008* (pp. 75–82). Lausanne, Switzerland: ACM.

Chen, L., & Pu, P. (2014). Experiments on user experiences with recommender interfaces. *Behaviour & IT, 33*, 372–394.

Choi, J., Lee, H. J., Sajjad, F., & Lee, H. (2014). The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change, 86*, 65–79.

Chu, W., & Huang, W. (2017). Cultural difference and visual information on hotel rating prediction. *World Wide Web, 20*, 595–619.

Esmaeili, L., Mardani, S., Golpayegani, S. A. H., & Madar, Z. Z. (2020). A novel tourism recommender system in the context of social commerce. *Expert Systems with Applications, 149*, Article 113301.

Ezin, E., Alcaraz-Herrera, H., & Iván. (2019). Balancing preferences, popularity and location in context-aware restaurant deal recommendation: A bristol, cardiff and brighton case study. In *Proceedings of the 13th ACM Conference on Recommender Systems and workshop on recommender in tourism (RecTour) 2019, Copenhagen, Denmark* (p. 38).

Farokhi, N., Vahid, M., Nilashi, M., & Ibrahim, O. (2016). A multi-criteria recommender system for tourism using fuzzy approach. *Journal of Soft Computing and Decision Support Systems, 3*, 19–29.

Frolov, E., & Oseledets, I. V. (2017). Tensor methods and recommender systems. WIREs. *Data Mining and Knowledge Discovery, 7*.

Fu, Y., Liu, B., Ge, Y., Yao, Z., & Xiong, H. (2014). User preference learning with multiple information fusion for restaurant recommendation. In M. J. Zaki, Z. Obradovic, P. Tan, A. Banerjee, C. Kamath, & S. Parthasarathy (Eds.), *Proceedings of the 2014 SIAM International Conference on Data Mining* (pp. 470–478). Philadelphia, Pennsylvania, USA: SIAM.

Gao, X. (2016). Cultural differences between east asian and north american in temporal orientation. *Review of General Psychology, 20*, 118–127.

George, T., & Merugu, S. (2005). A scalable collaborative filtering framework based on co-clustering. In Proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005) (pp. 625–628). Houston, Texas, USA: IEEE Computer Society.

Hassan, M., & Hamada, M. (2017). A neural networks approach for improving the accuracy of multi-criteria recommender systems. *Applied Sciences, 7*, 868.

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization, 10*, 15–41.

Hofstede, G., & Bond, M. H. (1988). The confucius connection: From cultural roots to economic growth. *Organizational Dynamics, 16*, 5–21.

Hong, M., Akerkar, R., & Jung, J. J. (2019). Improving explainability of recommendation system by multi-sided tensor factorization. *Cybernetics and Systems, 50*, 97–117.

Hong, M., An, S., Akerkar, R., Camacho, D., & Jung, J. J. (2019). Cross-cultural contextualisation for recommender systems. Journal of Ambient Intelligence and Humanized. *Computing*, 1–12.

Hong, M., & Jung, J. J. (2018). Multi-sided recommendation based on social tensor factorization. *Information Sciences, 447*, 140–156.

Hong, M., Jung, J. J., & Camacho, D. (2017). GRSAT: A novel method on group recommendation by social affinity and trustworthiness. *Cybernetics and Systems, 48*, 140–161.

Jung, T. H., Lee, H., Chung, N., & tom Dieck, M. C. (2018). Cross-cultural differences in adopting mobile augmented reality at cultural heritage tourism sites. *International Journal of Contemporary Hospitality Management, 30*, 1621–1645.

Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010). User preference learning with multiple information fusion for restaurant recommendation. In X. Amatriain, M. Torrens, P. Resnick, & M. Zanker (Eds.), *Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010* (pp. 79–86). Barcelona, Spain: ACM.

Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Y. Li, B. Liu, & S. Sarawagi (Eds.), *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 426–434). Las Vegas, Nevada, USA: ACM.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling* (volume 26). Springer.

Lakiotaki, K., Matsatsinis, N. F., & Tsoukiàs, A. (2011). Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems, 26*, 64–76.

Lee, L. Y.-S. (2016). Hospitality industry web-based self-service technology adoption model: a cross-cultural perspective. *Journal of Hospitality & Tourism Research, 40*, 162–197.

Li, M. (2014). Cross-cultural tourist research: A meta-analysis. *Journal of Hospitality & Tourism Research, 38*, 40–77.

López, C., & Farzan, R. (2015). Exploring the mechanisms behind the assessment of usefulness of restaurant reviews. In *Proceedings of the 7th International Conference on Communities and Technologies* (pp. 19–27). Limerick, Ireland: ACM.

Luo, X., Zhou, M., Xia, Y., & Zhu, Q. (2014). An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics, 10*, 1273–1284.

McArthur, T., Lam-McArthur, J., & Fontaine, L. (2018). *Oxford companion to the English language*. Oxford University Press.

McColl, R.W. (2014). Encyclopedia of world geography volume 1. Infobase Publishing.

Pu, P., & Chen, L. (2006). Trust building with explanation interfaces. In C. Paris, & C. L. Sidner (Eds.), *Proceedings of the 11th International Conference on Intelligent User Interfaces, IUI 2006* (pp. 93–100). Sydney, Australia: ACM.

Salehan, M., Zhang, S., & Aghakhani, N. (2017). A recommender system for restaurant reviews based on consumer segment. In Proceedings of the 23rd Americas Conference on Information Systems, AMCIS 2017. Boston, MA, USA: Association for Information Systems.

Singh, M. (2018). Scalability and sparsity issues in recommender datasets: a survey. In *Knowledge and Information Systems* (pp. 1–43).

Sun, L., Guo, J., & Zhu, Y. (2019). Applying uncertainty theory into the restaurant recommender system based on sentiment analysis of online chinese reviews. *World Wide Web, 22*, 83–100.

Tang, T. Y., Winoto, P., & Ye, R. Z. (2011). Analysis of a multi-domain recommender system. In *Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications* (pp. 280–285). Macao, China: IEEE.

Turk, A. M., & Bilge, A. (2019). Robustness analysis of multi-criteria collaborative filtering algorithms against shilling attacks. *Expert Systems with Applications, 115*, 386–402.

Viktoratos, I., Tsadiras, A., & Bassiliades, N. (2018). Combining community-based knowledge with association rule mining to alleviate the cold start problem in context-aware recommender systems. *Expert Systems with Applications, 101*, 78–90.

Walek, B., & Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications, 158*, Article 113452.

Wang, L., & Yi, B. (2019). Research on O2O take-away restaurant recommendation system: taking ele.me APP as an example. *Cluster Computing, 22*, 6069–6077.

Willmott, C. J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science, 20*, 89–102.

Yüksel, A., & Yüksel, F. (2003). Measurement of tourist satisfaction with restaurant services: A segment-based approach. *Journal of Vacation Marketing, 9*, 52–68.

Zhang, F., Yuan, N.J., Zheng, K., Lian, D., Xie, X., & Rui, Y. (2016). Exploiting dining preference for restaurant recommendation. In J. Bourdeau, J. Hendler, R. Nkambou, I. Horrocks, & B.Y. Zhao (Eds.), Proceedings of the 25th International Conference on World Wide Web, WWW 2016 (pp. 725–735). Montreal, Canada: ACM.

Zhang, S., Salehan, M., Leung, A., Cabral, I., & Aghakhani, N. (2018). A recommender system for cultural restaurants based on review factors and review sentiment. In Proceedings of the 24th Americas Conference on Information Systems, AMCIS 2018. New Orleans, LA, USA.

Zheng, Y. (2017). Criteria chains: A novel multi-criteria recommendation approach. In G. A. Papadopoulos, T. Kuflik, F. Chen, C. Duarte, & W. Fu (Eds.), *Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI 2017* (pp. 29–33). Limassol, Cyprus: ACM.

Zheng, Y., Mobasher, B., & Burke, R.D. (2015). Carskit: A java-based context-aware recommendation engine. In Proceedings of the IEEE International Conference on Data Mining Workshop, ICDMW'15 (pp. 1668–1671). Atlantic, NJ, USA.