Developing tourism users profiles with data-driven explicit information

Rasoul Norouzi[a], Hamed Baziyad[a], Elham Akhondzadeh Noughabi[b,*], Amir Albadvi[c]

[a] *M.Sc., Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran*
[b] *Assistant Professor, Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran*
[c] *Professor, Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran*

Abstract

Recommender Systems (RS) are widely used in the tourism industry to cope with information redundancy. Nevertheless, the cold-start problem is a major problem for such systems, and hybrid methods and user profile development with explicit and implicit information are typically deployed to address it. Asking many questions or obtaining sensitive demographic information about a new user can be an annoying and challenging process. As a result, in this study, we present a new method for obtaining explicit data-driven information that is based on the available items for users and free of demographic information. Users' interest in tourism activities is used to identify seven categories of tourism. The mapping between extracted categories and activities is established with a multi-label classification (MLC) algorithm. The user's interest in 18 tourism activities is predicted by rating only seven tourism categories. Common MLC algorithms with different classifiers were used to evaluate the proposed method. The best result relates to binary relevance with the naive bayes classifier, which also outperforms the two state-of-the-art algorithms in collaborative filtering (CF) systems as baseline model. The proposed method can capture users' interests and develop their profiles without receiving demographic information. Also, compared to CF, in addition to a slight advantage in metrics, it only requires seven ratings to predict user interest in 18 activities. In contrast, CF algorithms require at least 15 user ng records to predict user interest in unknown activities (3-4 activities) to achieve a performance close to the proposed method.

Keywords: Recommender System, Multi-Label Classification, Cold Start, Explicit information, Tourism

*Corresponding author
Email address: elham.akhondzadeh@modares.ac.ir (Elham Akhondzadeh Noughabi)

1. Introduction

A typical trip plan consists of several steps, including selecting tourist attractions, choosing accommodations, deciding which routes to take, etc.(Huang and Bian, 2009). Tourist attractions are often the main motivation for a tourist to choose a particular destination for traveling (Huang and Bian, 2009). Therefore, selecting which tourist attractions to visit is one of the most crucial decisions in trip planning, and it is strongly connected to the tourists' travel preferences and interests, which are not explicitly known (Hsu et al., 2012). Tourists are often confused about where to go when reaching new and unfamiliar places as there could be a wide variety of choices for consideration (Yang and Hwang, 2013). Besides, they typically have a limited amount of time and budget available; thus, it is almost impossible to visit all tourist attractions during a trip, especially to large cities (Abbaspour and Samadzadegan, 2011). As a result, tourists have to select the most compelling points of interest (POIs) according to their preferences. Then, they plan an itinerary, taking into account the time available to reach the POIs concerning their accessibility and opening hours (Vansteenwegen et al., 2011). Over the last decade, the rapid development of Information and Communication Technologies (ICTs) and the global expansion of the Internet strongly influenced the tourism sector as it affected all the sectors of the economy (Büyüközkan and Ergün, 2011).

People have realized the advantages of ICTs for planning an agenda of recreational and leisure contexts, to the extent that the Internet has become an integral and inseparable part of tourism (Neidhardt et al., 2015). However, the volume of information related to travel and tourism available on the web increases at a tremendous rate, and users are usually confronted with too many options to choose from (Gao et al., 2010). In trip planning, the most posted and searched information is concerning travel destinations and their associated resources, such as tourist attractions, accommodations, restaurants, local gourmet food, etc., to appeal to the tourists. Although all this information may be beneficial for users who plan to visit an unfamiliar place, the evaluation of this long list of options to select the one that fits better with a particular tourist needs is overwhelming and time-consuming (Borràs et al., 2014). Besides, around one-third of such users cannot explicitly express their travel needs and expectations (Zins, 2007). Hence, they require help for exploring the data and filtering out irrelevant information based on their specific preferences and needs identified as personalization.

Personalization is the ability to provide tailored content and services to users based on the knowledge about their preferences and tastes (Gao et al., 2010). Personalization techniques are mainly related to recommender systems (RS), which aim to filter irrelevant information and to provide personalized information to each particular user (Borràs et al., 2014). RS can be defined as a personalization tool that provides people with a list of items that best fit their individual preferences, restrictions, or tastes (Sebastia et al., 2009). The most crucial feature of RS is that it can guess a user's preferences and interests by analyzing the behavior of the individual and/or others to generate personalized recommendations (Lu et al., 2015). The RSs can automatically learn the user's preferences by analyzing their explicit or implicit feedback. Explicit data might be given by the user in different ways, for instance by requiring them to fill a questionnaire about their preferences and interests. The system can infer implicit interests through the analysis of the user's behavior (Borràs et al., 2014). The main aim of developing RSs is to counteract the risk of information overload by assisting users in searching for relevant information from a vast amount and variety of information (Grün et al., 2017). In recent years, RSs increasingly have caught the attention of scholars in the tourism field as a powerful tool for supporting information searching and decision making in this context. Various RS techniques in tourism have been applied standalone or in combination with heuristic algorithms, machine learning, and context-aware methods (Batet et al., 2012; García-Crespo et al., 2009; Ricci et al., 2002; Lucas et al., 2013; Chiang and Huang, 2015; Kolahkaj et al., 2020; Zhu et al., 2019; Hsu et al., 2012; Huang and Bian, 2009; Loh et al., 2003; Sarkar et al., 2020; Moreno et al., 2013; Neidhardt et al., 2015; Venturini and Ricci, 2006; Nilashi et al., 2015; Rivolli et al., 2017; Sebastia et al., 2009; Yang and Hwang, 2013).

Commonly used recommendation techniques are knowledge-based (KB), content-based (CB), collaborative-filtering (CF), and Hybrid RS. The latter one uses a combination of different methods to overcome the weaknesses of each (Lu et al., 2015). In KB systems, item recommendations are done based on understanding users' and items' features and their underlying relationship (Borràs et al., 2014). In CB, the system measures the degree of similarity between the user and the items. This process is done by analyzing product features concerning user preferences. Therefore, it is assumed that both the user and the items selected for the recommendation have common features. The analysis process output usually displays the overall performance score, which indicates how much the user profile matches with the recommended items. The selected items that get more points have better performance with a higher matching rate. Sometimes, this approach also deals with the user's scoring history. In this approach, the system must have an accurate knowledge of the user to provide a recommendation (Lu et al., 2015). CF-based systems help people make their choice based on the opinions of those similar to them. The similarity between users is calculated based on the scores they have given to the list of items. When the system finds out which people are closer to each other based on their interests and choices, other "similar" users' favorites are suggested to the intended user. In this approach, to find out which recommendation is favorable and which is not, obtaining feedback is necessary. CF systems use the user-item matrix to predict users' interest in items. In such matrices, each row, column and cell respectively represents a user, an item, and a user giving rate to an item (Ricci et al., 2015). Each RS has some drawbacks that prevent them from being suitable for all scenarios and conditions.

Two of the most important problems, especially in the CF-based system, are sparsity and cold-start. Sparsity happens when many user-item matrix cells suffer from the lack of rates given by users. This makes training of machine learning models, especially in memory-based algorithms, challenging. By growing the number of items and users, the sparsity and the dimension of the user-item problems become more severe and more problematic. The cold-start problem occurs when entering a new user into the system; since there is no record of the user interests and rates to the items, it is impossible to predict what the user would be attracted to. The same problem can exist for newly added items, which in literature, it refers to items cold-start. The cold-start problem is usually handled by using hybrid systems or expanding users and item profiles through gathering explicit and implicit information. (Sun et al., 2019).

The explicit information is often preferred over implicit information because it is more accurate than the predicted or implicit information, i.e. the user can directly enter information about his interest, and then the system will generate accurate recommendations for him (Khalid and Wu, 2016). However, receiving explicit information could be challenging for a system. Users might feel uncomfortable providing explicit demographic information, and extremely common questions, such as one's race, gender, income, or age, could cause bias and unfair recommendations (Mehrabi et al., 2021).To this end, in this study, we have proposed a method in which the information collected from a new user does not contain demographic information, and the enquired explicit information is data-driven. In this method, tourism activities are categorized by using exploratory factor analysis (EFA). New users with no rating record of tourism activities are asked to rate each of these categories on a scale of one to five points. The data, rated categories, is then mapped to the activities by a multi-label classification algorithm (MLC), which predicts what activities the user is likely to enjoy; in other words, it will develop a tourism profile for the user. MLC is somehow a supervised machine learning algorithm where each instance can belong to multiple classes or, in other words, can have multiple labels (Ganda and Buch, 2018).

This study brings up two important questions, each of which has been addressed in evaluation section. The first question is how is the performance of proposed method at predicting users' interests and which MLC algorithm performs better. As part of the internal evaluation phase, we compared different MLC algorithms based on different criteria in order to answer this question. The results show that proposed method, which is evaluated with three metrics, can appropriately capture and predict users' interests. The second question concerns how this method compares with the baseline model and whether it is worthwhile to take two additional steps to extract tourism categories and collect data with a questionnaire. We addressed this question through external evaluation by converting the problem into a CF problem as a baseline model. Our external evaluation was based on a scenario where a user is randomly introduced to some tourism activities and is asked to indicate his interest in each one on a binary scale (interested or not interested). Then, the user's other interests are predicted by applying the state-of-the-art CF algorithm to the obtained data. Our method results outperform those achieved by CF algorithms in the defined scenario. This shows that it can properly map explicit information received from the user to activities and reduce the amount of received explicit information compared to the baseline model.

The rest of the paper is structured as follows: The next section provides an overview of the multi-label classifiers and their applications in RS; the research methodology section deals with how to identify tourism activities, how to extract tourism categories, proposing algorithms to predict the user's favorite activities, and how to evaluate the presented method. In the result section, we review and analyze the method's ability in capturing and predicting user interests. Lastly, in the conclusion section, we review our method and discuss this paper's achievements, research limitations, and future research suggestions.

2. Multi-label Classification and Related Works

In machine learning, single-label classification is one of the commonly used methods in which each instance in the dataset associates with a unique class label from a set of disjoint class labels L. Depending on the number of these classes, the problem can be either a binary classification (when $|L| = 2$) or a multi-class classification (when $|L| > 2$). However, in the multi-labeling problems, each instance can be associated with multiple classes. In such algorithms, the goal is to learn from a set of instances to label each instance's class or classes in L (Sorower, 2010). MLC approaches are categorized into a) problem transformation and b) algorithm adaptation methods.

In problem transformation, the MLC problem transforms into one or more single-label classification problems. Therefore, it does not need any change or adaptation to traditional algorithms, and those algorithms can be applied to the problem (Tsoumakas and Katakis, 2007). Problem transformation methods are divided into three main algorithms: Binary Relevance (BR), Label Power Set (LP), and Classifier Chain (CC). Using these three algorithms, this study applies five classifiers, namely Support Vector Machine (SVM), Decision Tree (DC), Random Forest (RF), Naïve Bays (NB), and K-Nearest Neighbor (KNN). In adaptation algorithms, instead of transforming the problem, the algorithms are changed and modified to handle multi-label data. We used two adaptation algorithms, namely Binary Relevance KNN (BRKNN) and Multilabel K Nearest Neighbors (MLKNN) (Spyromitros et al., 2008; Tsoumakas and Katakis, 2007). Besides these approaches, ensemble learning algorithms can learn from multi-label data natively without any transformation in the base algorithms or the problem. Ensemble methods are learning algorithms that construct a set of classifiers before classifying new data points by taking a (weighted) vote of their predictions (Rokach et al., 2014). This study used Random Forest (RF) and Extra Tree classifiers (ET) as ensemble algorithm candidates.

MLC has many applications in various domains including text classification (Schulz et al., 2016; Omar et al., 2021), image classification (Wang et al., 2016), bioinformatic (Zhang et al., 2018), genre classification (Sanden and Zhang, 2011), and social media analysis (Chen et al., 2014). More details could be found in Tidake and Sane (2018) and Tsoumakas et al. (2009) papers. Moreover, MLC has leveraged its power into RSs world too. Carrillo et al. (2013) demonstrated the MLC ability to recommend items and deal with RS common problems including data sparsity. Zheng et al. (2014) have used MLC to recommend users' contexts in such a way that instead of recommending the item to the user, the user-related contexts are predicted based on the items selected by the user and the ratings given to each item. To this end, they transformed the problem into an MLC problem and showed that MLC algorithms are more capable of recommending and predicting than the base algorithms. Rivolli et al. (2017) used the MLC algorithms to recommend track foods. They obtained a set of data using a questionnaire comprised of two stages. In the first stage, the user answers to 21 questions, which are the attributes describing the user. These questions are considered as predictive attributes. The second stage of the questionnaire includes 12 food alternatives in which the user is asked to specify their preferences to each of them. These alternatives associate with classes' labels or target attributes. The results indicate that the adopted method showed a weaker performance in comparison to the transformation methods. Elhassan et al. (2018) used MLC to provide remedial actions to address students' shortcomings in Learning Outcome Attainment Rates. In their model, each instance is a student described by a set of characteristics such as field of study, academic level, grades, and so on. Moreover, the related tags for each student is equal to their remedial actions. The results show that the chain classification method with the decision tree algorithm gives the best outcome for the given dataset.

However, despite the wide range of studies done about MLC applications in RS, there is still insufficient attention and evaluation of MLC capabilities. one of those capabilities is using MLC to address the cold-start problem and reduce the amount of received explicit information of a new user. This study is an attempt to fill mentioned gaps and show the performance of MLC algorithms in comparison with CF algorithms as base models. To the best of our knowledge, this is the first work in addressing the cold-start problem in developing tourism users' profiles with explicit data-driven and MLC algorithms.

3. Research Methodology

From a data-driven viewpoint, there are two main steps in building the proposed recommendation: the first step is to extract tourism categories by measuring users' interests in tourism activities. The second step is to associate categories and activities with a data-driven connection. To establish such a connection, data collection and training the MLC algorithm are required. Therefore, this connection allows the prediction of the user's interest in activities using his ratings in extracted categories. The tourism sites and activities of Tehran, the capital city of Iran, have been selected as the case study of the presented method. In this section, we respectively address the identification process of tourism activities, the first phase of data collection using the questionnaire, applying factor analysis on the first dataset to extract tourism categories, second phase data collection using the questionnaire, training, reporting the performance of diverse MLC algorithms on the second dataset, and finally in external evaluation comparing our method best result with CF algorithms as a baseline model. The stages of the proposed method are shown graphically in Fig.1.

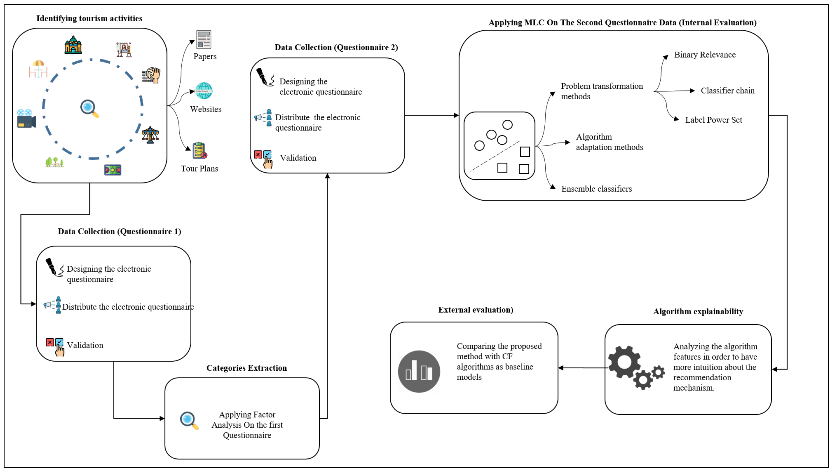

Figure 1: A schematic representation of the stages in this study

3.1. Identifying tourism activities

To identify tourism activities in Tehran, we used previous research (Moreno et al., 2013; Barta et al., 2009; Prantner et al., 2007), analytical reports of the British Tourism Organization[1], and the content of the Tripadvisor website[2]. The point to keep in mind was that many of the tourism activities mentioned in the papers and British Tourism Organization's analytical reports, such as nightclubs or beach tours, do not exist in Tehran. Therefore, by combining and modifying the activities mentioned in the mentioned sources, 18 types of tourism-related activities in Tehran were identified, namely going to cinema, theaters, museums, holy sites, historical sites, sports events, sports activities, art and book exhibitions, music events, malls, public gardens, restaurants, cafe, zoo, rural places, rivers and lakes, and mountains.

3.2. Data Collection (Questionnaire 1)

Reviewing and rating each of these 18 tourism activities can be a difficult and tedious task for a user. Thus, reducing these 18 activities to few and more interpretable categories makes the user more comfortable in recognizing and categorizing the content. To build the category layer, data was required; therefore, based on the Likert scale (a scale between one and five where one indicates the slightest interest in an activity, and five denotes the most), a questionnaire was designed to measure users' interest in each of the 18 activities. To better guide the users, we introduced several POIs in Tehran as instances for each of the mentioned activities. As an example, Saadabad Palace and Negarestan Mansion were mentioned as instances for the historical sites. After designing

[1] `www.visitbritain.org/archive-great-britain-tourism-survey-overnight-data`
[2] `www.tripadvisor.com/Attractions-g293998-Contexts-Iran.html`

the questionnaire, it was distributed randomly on social media platforms, such as Facebook, Twitter, and messenger applications. The total number of 272 questionnaires were collected, and the reliability of the designed questionnaire was proved by calculating Cronbach's alpha equal to 0.846, which was more than the cutoff required of 0.7 (Kopalle and Lehmann, 1997).

3.3. Extracting Tourism Categories (Factor Analysis of Questionnaire 1)

Factor analysis (FA) primary goal is to summarize data for revealing relationships and patterns by regrouping variables into a limited collection of clusters based on shared variance. FA utilizes mathematical methods to simply interrelate measures for discovering patterns in a set of variables. FA was applied in various types of fields, such as behavioral and social sciences, medicine, economics, and geography (Yong et al., 2013), and is divided into two main classes, namely exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is used when the research goal is to discover the number of influencing variables or to find variables that go together. FA is useful for studies such as questionnaires based on a few to hundreds of variables which can be reduced to a smaller set to simplify interpretations. Therefore, not only is focusing on a smaller set of variables easier than considering too many keys but also, by clustering them into some categories, it makes variables meaningful. In this paper, EFA was applied for accessing meaningful categories of variables. The determinant score for our data is 0.0000135, which is more than 0.00001, and indicates a violation in the assumption of correlation of variables; in such a case, to extract the factors, it is recommended to use Principal Axis Factor (Yong et al., 2013). We used the Varimax rotation method with 30 iterations based on the default value in SPSS software for rotation. To check the adequacy and suitability of the dataset for EFA, Kaiser-Meyer-Olkin measure (KMO) and Bartlett's test of sphericity were applied. The minimum value of the KMO index for the factor analysis is 0.5, which in our research is 0.76. The Bartlett test takes a statistical hypothesis, and its null hypothesis states that the correlation matrix is an identity matrix, so there is no significant relationship between the variables. As can be seen in Table.1, the p-value is not in the rejection area (the value of sig must be less than 0.05, which is zero for our data). To determine

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.76
Bartlett's Test of Sphericity	Approx. Chi-Square	1833.255
	Df	325
	Sig.	0

the number of significant factors, Kaiser's criteria states that only factors with Eigenvalues of one or more should retain. According to Fig.2 (Scree Plot), the best number of factors after rotation for this dataset is 7.

As factor naming does not follow a specific rule, here we named each factor based on the associated variables that describe the factor (Table.2).

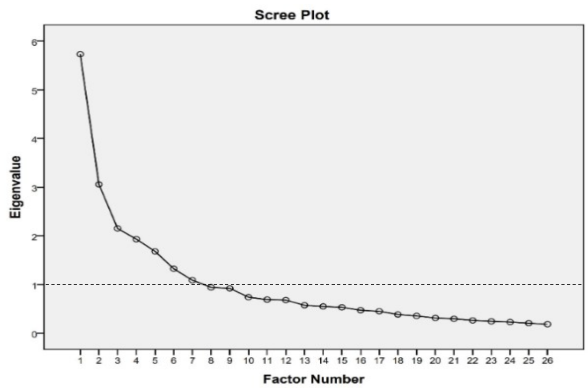

Figure 2: Scree Plot for determining the optimum number of factors, factors below eigenvalue one are dropped out

Table 2: Extracted factors name and their associates' variables

Factor	**Descriptor Variables**
Historical	Museums, Historical Sites
Fun	Restaurants, Cafe, Male
Ecotourist	Rivers and Lakes, Mountains, Rural Places
Sportive	Sport Activities, Sport Events
Cultural	Music Events, Art and Book Exhibitions, Cinema, Theaters
Religious	Mosques and Churches, Holy Sites
Urban-related	Zoo, Public Gardens, Parks

3.4. Data Collection (Questionnaire 2)

For mapping the connection between the categories and activities, data is required to train MLC algorithms. The advantage of this connection is that different states can be considered, and the new user's interest in tourism activities can be predicted only by rating the seven categories extracted by the factor analysis (Fig.3). Therefore, a dichotomous questionnaire was designed. The first part asked the users to determine their interest rate for every category on a five-point Likert scale. The second part requested to express their fondness toward all 18 activities by binary values of 0 for being not interested, and one for being interested in the activity. This electronic questionnaire was randomly broadcast through social media and messengers' applications for Tehran residents. Finally, 578 questionnaires were collected, and the calculated Cronbach's alpha was equal to 0.859.

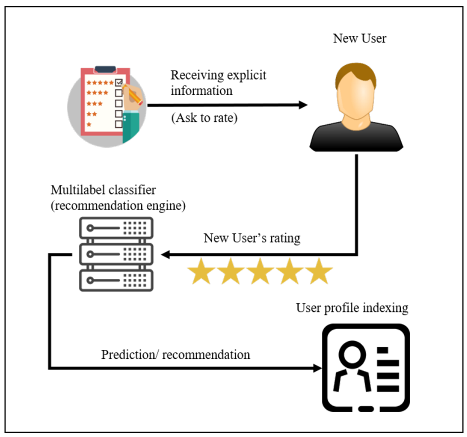

Figure 3: The schema of proposed method for developing new users profile

3.5. Multi-Label Classification Problem Definition

Let χ be the Users-Category matrix, and $\mathcal{L} = \lambda_1, \lambda_2, \ldots, \lambda_k$ be a finite set of labels or activities. A user $x \in \chi$, is represented in terms of the features vector $x = (x_1, x_2, \ldots, x_m)$. This features vector is referred to a user's given rates to each of the extracted categories; therefore, the user x is associated with a subset of labels $L \in 2^{\mathcal{L}}$. Notice that if L be the set of relevant labels of x, then the complement $\mathcal{L} \setminus L$ would be the set of irrelevant labels of x. Let denote the set of relevant labels L with a binary vector $Y = \ (y_1, y_2, \ldots, y_k)$, where $y_i = \ 1 \Leftrightarrow \lambda_i \in L$, then $\mathcal{Y} = \{0, 1\}^k$ is the set of all such possible labeling (Fig.4).

Therefore:

Given a training set, $= \ (x_i, Y_i), 1 \leq i \leq n$, consisting n training instances,($x_i \in \mathcal{X}, Y_i \in \ \mathcal{Y}$) i.i.d[3] drawn from an unknown distribution D, the goal of the multi-

[3]independent and identically distributed

Features vector or given rates to each category — *Associated subset of labels*

	$category_1$	$category_2$	...	$category_m$	$activity_1$	$activity_2$	...	$activity_k$
$user_1$	3	5		1	1	0		1
$user_2$	2	4		5	0	0		1
.	.	.		.	.	.		.
.	.	.		.	.	.		.
.	.	.		.	.	.		.
$user_n$	3	2		5	1	1		0

Figure 4: Part of a multi-label problem matrix

label learning is to produce a multi-label classifier $h : \mathcal{X} \rightarrow \mathcal{Y}$ (in other words, $h : \mathcal{X} \rightarrow 2^{\mathcal{L}}$) that optimizes specific evaluation functions (i.e. loss function) (Sorower, 2010).

This study uses the second questionnaire data as a training data set for MLC algorithms. As a result, after entering a new user to the system, only by rating each of the seven categories in the range of 1 to 5, his/her interest in the 18 activities will be predicted. In the transformation approach, all three algorithms (BR, LP, CC) with LR, DT, RF, SVM, KN classifiers are used. In the adaptation algorithm, BRKNN and MLKNN and in the ensembles algorithm, RF and ET classifiers are utilized. To implement MLC algorithms, Python version 3.5 with scikit-learn and scikit-multilearn packages are used. All the classifiers' hyperparameters in this study are the package's default values.

An issue that should be considered is imbalanced labels' problems; as shown in Fig.5, the number of classes is not equal in any of the labels, and this could cause problems in some algorithms' learning process. To solve this problem, we used the scikit learn package class weight balancing feature in all classifiers except the NB, MLKNN, and BRKNN; because the technique does not apply to such classifiers.

3.6. Evaluation

In this research, the evaluation stage is vital in two ways. First, in internal evaluation, in response to the first research question, we measured and compared the performance of different MLC algorithms in capturing and predicting users' interests. Second, in external evaluation to addressing the second research question, the proposed method performance was compared to state-of-the-art CF algorithms as a baseline model. For this purpose, 5-fold cross-validation with three metrics was used. Cross-validation can train algorithms with little data. Moreover, since all samples are used both as training data and test data in the algorithm learning process, it is a favorable way to compare different algorithms' performance in classification problems. There are some different metrics for MLC evaluation, however, it should be notice that metrics should be chosen that are usable for CF methods too. As mentioned, one of our goals in the evaluation stage is to compare the proposed framework with state-of-the-art

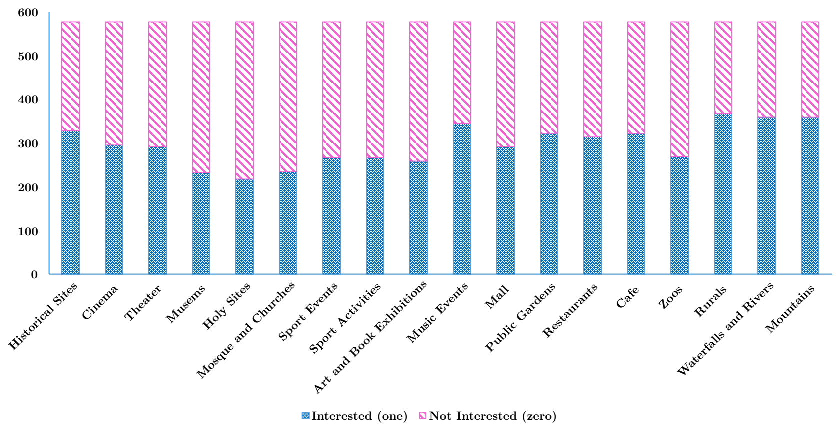

Figure 5: Labels imbalance, vertical pivot represents the number of samples, and the horizontal pivot shows the label names

CF techniques. Thus, in this study, we selected Precision, Recall, and F1-Score metrics for evaluation. Since the weight-balancing technique is not applicable for some classifiers (NB, MLKNN, BRKNN), we used the macro-average criteria that do not take label imbalance into account.

3.6.1. Metrics

Let T be a multi-label dataset consisting n multi-label examples $(x_i, Y_i), 1 \leq i \leq n, \left(xi \in \mathcal{X}, Yi \in \ \mathcal{Y} = \{0,1\}) \ ^k \right)$, with a label set $\mathcal{L}, |\mathcal{L}| = k$. Let h be a multi-label classifier and $Z_i = h(x_i) \ = \{0,1\}^k$ be the set of label memberships predicted by h for the example x_i.
Therefore:
Precision (P): is the proportion of correctly predicted labels to the total number of actual labels, averaged over all instances. In our case, Precision indicates how many of the predicted activities are actually selected by the user.

$$precision, P = \frac{1}{n} \sum_{i}^{n} \frac{|Y_i \cap Z_i|}{|Z_i|}$$

Recall (R): is the proportion of correctly predicted labels to the total number of predicted labels, averaged over all instances. In our case, Recall indicates how much the algorithm has been able to predict the user's favorite activities.

$$recall, R = \frac{1}{n} \sum_{i}^{n} \frac{|Y_i \cap Z_i|}{|Y_i|}$$

F1-Score: Naturally, the definitions for Precision and Recall lead to the following definition for F1-score: (Sorower, 2010; Ganda and Buch, 2018):

$$F1 = \frac{1}{n}\sum_{i}^{n}\frac{2 * |Y_i \cap Z_i|}{|Y_i| + |Z_i|}$$

3.6.2. The Proposed Method versus Baseline Model

In this part, the presented framework should be compared with CF algorithms as a baseline model where a user's interests in an activity is predicted from his direct interaction with other activities. To evaluate the CF's prediction ability with our proposed method, the problem should be transformed into a CF problem. We defined a scenario in which users are given a set of activities at random and asked to indicate their interest in each one in a binary manner. After that, the user's scoring record is fed into a CF algorithm, which predicts his other interests. In order to do this, we used the second questionnaire data. We eliminated users' ratings to categories for CF problem and only kept users' binary ratings of activities. Finally, our method's best score was compared to that of two state-of-the-art CF algorithms. CF systems generally have memory-based and model-based techniques for the recommendation. There are no assumptions on data in the memory-based technique and it essentially depends on the nearest neighbors' search to find the closest pairs of items or users. When the recommendation is based on measuring the similarities between items, it is an item-item method and when it is based on measuring users' similarities, it is a user-user method. For our problem, we focus on the user-user method because the number of items (activities) is few. When items are few, the variance in the item-item method is low and its bias is high. This could cause less personalization. On the other side, model-based technique tries to assume data and learn a model that explains user-item matrix interactions (Fig.6).

	$activity_1$	$activity_1$	...	$activity_k$
$user_1$	?	0		1
$user_2$	0	?		1
.	.	.		.
.	.	.		.
.	.	.		.
$user_n$	?	1		?

Figure 6: An example of user-item matrix

To formulate our problem for CF systems, only the user activity matrix (users' interest in activities) without the user-category (their rating to each category) is required. Thus, having N users and K activities, the user-activity matrix $A \in \{0,1\}^{N\times K}$:

$$\begin{cases} u_{ri} & \text{if such rating exist} \\ ? & \text{otherwise} \end{cases}$$

CF systems try to replace all the "question marks" in A_{ui} by some optimal guesses; the goal is to minimize the RMSE (root mean square error) when predicting the user interests on a test set (which is, of course, unknown during the training phase):

$$rmse = \sqrt{\frac{1}{|S_{test}|} \sum_{(i,u)\in S_{test}} (r_{ui} - \hat{r}_{ui})^2}$$

Where $(i,\ u)\ \in\ S$ test if user u interest activity i in the test set, $|S_{test}|$ is its cardinality, r_{ui} is the true rating, and $\hat{r}_{ui}$ is the prediction based on the recommendation system (Wen, 2008). The CF systems were implemented using the surprise package in the Python programming language (Hug, 2020); Furthermore, the benchmark results were used to select the top two CF algorithms[4].

In a memory-based technique, we choose BSKNN taking into account a baseline rating; A baseline estimate for an unknown rating r_{ui} is denoted by b_{ui} and accounts for the user and item effects:

$$b_{ui} = \mu + b_i + b_u$$

The parameters b_u and b_i indicate the observed deviations of user u and activity i, respectively, from the average, and μ denotes the overall average rating. To predict $b_{u,i}$:

$$\min \sum_{(u,i)\in ||} (r_{ui} - \mu - b_u - b_i)^2 + \lambda_1 (\sum_u b_u^2 + \sum_i b_i^2)$$

Therefore, the prediction $\hat{r}_{ui}$ is set as:

$$\hat{r}_{ui} = \mu_u + \frac{\sum_{v\epsilon N_i^k(u)} sim\,(u,v)\,.(r_{vi} - \mu_v)}{\sum_{v\epsilon N_i^k(u)} sim\,(u,v)}$$

Where $sim\,(u,v)$ denotes, similarity measurement between user u and v, and $N_i^k(u)$ only include neighbors for which the similarity measure is positive (Koren, 2010).

In the **model-based technique**, we used Singular Value Decomposition (SVD), a matrix factorization algorithm that tries to decompose the original sparse matrix to low-dimensional matrices with latent factors (q_i,p_u).
The prediction of $\hat{r}_{ui}$ is set as:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u$$

If user u is unknown, then the bias b_u and the factors p_u are assumed to be zero. The same applies for item i with b_i and q_i.
To estimate the unknown parameters, the problem should be minimized:

$$\sum_{r_{ui}\in R_{train}} (r_{ui} - \hat{r}_{ui})^2 + \lambda(b_i^2 + b_i^2 + ||q_i||^2 + ||p_u||^2)$$

[4] http://surpriselib.com/

For details could be found in Surprise package's document and Koren (2010).

Another problem is that the output of the mentioned algorithm is in the range of $[0, 1]$, however, the desired output should be binary with one and zero respectively denote a user interest in and dislike of an activity. To solve this problem, we suggest using a threshold α where:

$$Output = \begin{cases} 1 & \hat{r}_{ui} > \alpha \\ 0 & \text{else} \end{cases}$$

For both selected algorithms, to choose the best threshold value, the F1-score was calculated for different α values and one having the best F1-score is chosen as the threshold.

4. Result and Discussions

This section includes two parts of evaluation; in the first part, MLC algorithms' results is reviewed, and in the next part, the results of the best MLC algorithm is compared with the CF problem-solving approach.

4.1. MLC Results (Internal Evaluation)

The performance of MLC algorithms is shown in table 3. The three columns represent the Precision, Recall, and F-1 score; the higher the value, the better the result. For MLC transformation algorithms, we used "Algorithm-Classifier" to denote the algorithm, e.g., BR-NB denotes the use of binary relevance (BR) as transformation algorithm and Naïve Bayes as a classifier. Moreover, for Ensemble methods, we used the "Ensemble-Classifier" form. The results demonstrate that the proposed method can capture and predict the user's interests with just a few explicit inputs.

According to the metrics results, the best performance is related to BR algorithms, while the LP algorithm shows lower metrics values than others. The closeness of the precision score and the Recall in most classifiers indicates that the class imbalance has not affected the learning process. Since there are 18 tourism activities, BR algorithms transform the problem into 18 separate problems regardless of the interdependence of labels. This algorithm's satisfactory performance indicates that the detected activities are distinctive from each other, and the algorithm has been able to map the category layer to the activity layer space well. For example, the RF classifier with the BR algorithm showed better results than other algorithms with the RF classifier. Nevertheless, the LP algorithms' disappointing results are due to their problem-solving approach. LP converts the MLC problem into a multi-class classification problem with 2^L possible class values. Since the dataset used in this study has few records and many labels, it is difficult to train such algorithms on such a data set. Nevertheless, algorithms' outcome and superiority may change as the number of data increases. Recall's high score refers to the classifier's success rate in identifying and proposing the user's favorite activities. The higher this score is,

Table 3: Comparison of the MLC algorithms' performance

	Precision	**Recall**	**F1-Score**
BR-RF	0.748	0.722	0.725
BR-SVM	0.726	0.715	0.712
BR-LR	0.733	0.721	0.718
BR-NB	**0.75**	**0.74**	**0.733**
BR-DT	0.657	0.647	0.647
CC-RF	0.727	0.633	0.658
CC-SVM	0.724	0.692	0.699
CC-LR	0.718	0.694	0.692
CC-NB	0.677	0.663	0.656
CC-DT	0.657	0.673	0.66
LP-RF	0.62	0.615	0.61
LP-SVM	0.657	0.659	0.652
LP-LR	0.648	0.601	0.616
LP-NB	0.629	0.654	0.594
LP-DT	0.641	0.655	0.643
BRkNNa	0.731	0.666	0.674
MLkNN	0.708	0.693	0.683
Enssemble-ET	0.697	0.698	0.69
Enssemble-RF	0.715	0.699	0.696

the more the recommender has recommended the user's favorite tourism activities. On the other hand, Precision indicates what percentages of the activity recommended to the user were actually the user's favorite activities. Of course, sometimes, a precision error can also be welcome, i.e., when an activity outside of the user's favorite activities is recommended, and the user feedback to that is positive; therefore, it can be seen as a way to prevent over-personalization. If a recommender offers all the activities to the user, its recall score will be 100%, but its precision score will be low. Also, if it tries to suggest a smaller number of activities to the user, its Recall will be low, and its precision score will be high. In such cases, the F1-score, which is a harmonic average of the two mentioned metrics, can be a useful criterion for comparing classifiers' performance. To select the best algorithm based on the F-score results, BR-NB has the best performance among the categories. Its Precision, Recall, and F1 scores are 0.75, 0.74, and 0.733, respectively. The evaluation results of the two adaptive algorithms, BRKNN and MLKNN, were similar and did not have a significant advantage over each other, as are the ensemble algorithms.

4.1.1. Feature Importance Explanation

To have better intuition on how MLC predict users' interested activities, in table 4 a breakdown of features importance for each activity is presented. In order to perform the feature importance analysis, we chose the BR-RF algorithm since our best algorithm, the Naive Bayes algorithm, does not support this. Though the results in the table 4 seem quite clear, since our case study concern Tehran, some of the feature importance analysis results need for further explanation.

Cultural and fun features are important for both cinema and theater activities, but there is a slight difference in the importance of these features for the two activities. Compared to cinema, the fun feature of theater has decreased in importance, and the cultural feature has increased. In fact, the theater is a more cultural experience for individuals than films. In visiting museums activity, aside from cultural and historical features, the religious feature is also very effective because of the many religious museums in Tehran, such as the Quran Museum. For activities such as visiting public gardens and parks, in addition to urban-related features, cultural and historical features are also important. Some public gardens of Tehran have classical architecture buildings, which give the gardens a cultural and historical significance, e.g., Negarestan garden. The urban-related feature is the second most effective for all activities such as visiting rural areas, mountaineering, lakes, and rivers. Many tours and tourism categories introduce these activities as weekend tourism activities. As Tehran has easy access to the mentioned points of interest and many of them are within walking distance of residential areas, it is understandable for people to have a similar urban view of such activities.

Table 4: Feature importance analysis of activities

	Historical	Fun	Ecotourist	Sportive	Cultural	Religious	Urban-related
Historical Sites	0.411	0.084	0.074	0.050	0.199	0.093	0.089
Cinema	0.044	0.211	0.042	0.084	0.420	0.138	0.061
Theater	0.083	0.128	0.099	0.118	0.468	0.064	0.042
Musems	0.228	0.071	0.085	0.071	0.363	0.141	0.041
Holy Sites	0.085	0.088	0.028	0.132	0.040	0.465	0.162
Mosques and Churches	0.284	0.052	0.140	0.033	0.090	0.333	0.068
Sport Events	0.054	0.039	0.094	0.544	0.086	0.075	0.108
Sport Activities	0.064	0.096	0.100	0.541	0.043	0.079	0.077
Exhibitions (Art and Book)	0.116	0.124	0.044	0.104	0.432	0.132	0.048
Music Events	0.088	0.318	0.035	0.080	0.300	0.055	0.124
Mall	0.042	0.543	0.035	0.042	0.102	0.127	0.109
Public Gardens and Parks	0.106	0.097	0.214	0.032	0.154	0.063	0.334
Restaurants	0.043	0.560	0.070	0.038	0.088	0.137	0.065
Café	0.041	0.552	0.032	0.074	0.160	0.085	0.056
Zoo	0.042	0.162	0.074	0.115	0.081	0.229	0.297
Rural Places	0.155	0.060	0.414	0.046	0.071	0.079	0.175
Lakes and Rivers	0.057	0.076	0.366	0.080	0.055	0.068	0.297
Mountains	0.114	0.080	0.396	0.124	0.092	0.061	0.132

4.2. BR-NB VS BSKNN and SVD (External Evaluation)

The BSKNN and SVD algorithms' output is in the range of zero to one, so a threshold was used to convert it to a binary output. Fig.7 shows each of the algorithms' F1-Score for different α values. As expected from the performance of these algorithms, with the increase of α, the process of recommending activities to the user becomes more rigorous, consequently, the amount of recall score decreases and the score of precision increases. For both algorithms, the best equilibrium point is the α value that gives the highest F-Score. As Fig.7 shows, the best value of F1-Score for both algorithms is in the range of 0.45 to 0.55, and the difference of F1 in this limit is negligible for both. Therefore, an α value of 0.5 was chosen for both algorithms. To compare the proposed method

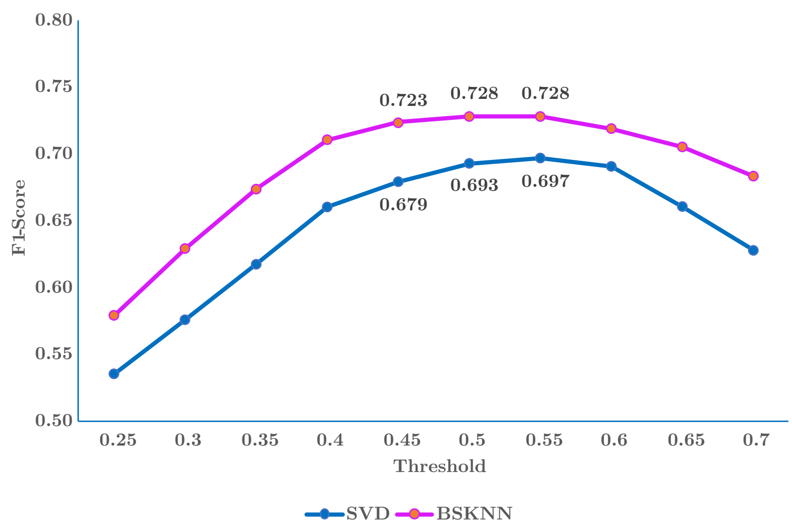

Figure 7: BSKNN and SVD F1-Scores for different α values

with the state-of-the-art CF algorithms, BSKNN and SVD as baseline models, we selected the BR-NB classifier, which has the best performance among the MLC algorithms. Fig.8 Shows the Precision, Recall, and F1 scores for each of them. The weakest performance is related to the SVD algorithm as its learning process usually uses gradients decent to estimate their parameters, while this technique requires many data. The performance of BR-NB in the proposed method is superior to the BSKNN algorithm by nearly three, two and one percent in Precision, Recall and F1-score, respectively. The proposed method's performance was slightly better than that of the CF algorithms. Yet, it should be noted that in the CF systems, only one step of data collection is needed and not the extraction of the tourism categories. Nevertheless, our method advantage is that the user interacts with only seven categories; in other words, we have reduced the problems dimensions by mapping user input to activities

Another critical point is that we also evaluated CF approach algorithms with the 5-fold cross-validation method. The total number of activities is 18. In each step of cross-validation, 4-fold (for each user, nearly 15 given rate records) is considered as training data and one-fold as test data (for each user, about three given rate records). This means that the CF algorithms' results are based on having about 15 records of user interaction with tourism activities and predicting user interests to nearly three activities. However, in the proposed method, the user interests in all 18 tourism activities are predicted by rating $[1, 5]$ scale (explicit data) to each of the seven tourism categories without having any in-

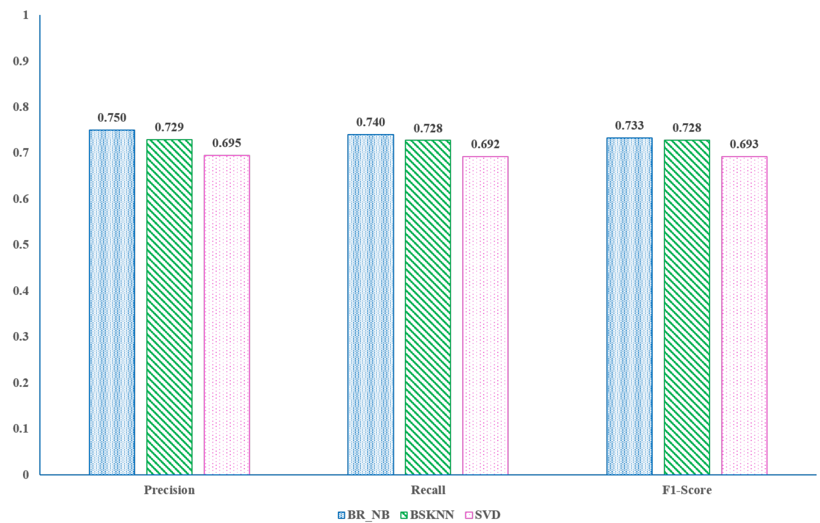

Figure 8: Comparison of BR-NB results in the proposed method with two state of the art algorithms, BSKNN and SVD, in CF systems

teraction records.

5. Conclusion and Future Research

In this study, we present a method of receiving explicit information that addresses the cold-start problem of tourism users without including sensitive or demographic information, and the enquired explicit information is data-driven. For this purpose, several tourism activities were identified in city of Tehran. Then tourism categories were obtained by applying FA on a questionnaire data that measured users' interests in the identified tourism activities. Accordingly, a MLC problem was formulated, in which new users' ratings of each category represented explicit input that is mapped to identified activities. This decoding process by MLC predicts that the user is interested to an activity or not. We used a second questionnaire to collect the required data for training and testing the MLC algorithms. In the internal evaluation phase, we compared the results of different MLC algorithms, and the BR-NB results performed best compared to the other classifiers. According to the literature review and the performance outcome of different MLC algorithms in this study, they can have different performance given to the problem and data. In other words, there is not any definite superiority for any of the algorithms. In external evaluation phase, we also compared the best classifier result in our proposed method with state-of-the-art CF algorithms as a baseline model, i.e., BSKNN and SVD. Our method outperformed the two mentioned algorithms, although this was a slight advantage. In addition to slight superiority in metrics, reducing the problem space from 18 activities to seven tourism categories is another advantage meaning that the user does not need to interact with all 18 activities or have rating records in the CF method for the profile developing. According to the results, our proposed method can capture and predict users' interests by receiving a few explicit information of new users. Comparatively, CF algorithms require more users rating records to achieve a close performance to our method.

As with any research, the present study has some limitations. The lack of a feedback mechanism is one of these limitations, which we will address in future studies. Another limitation was that this study focused on developing user profiles with a minimum amount of explicit and insensitive information from a new user. In other words, it does not recommend any POI to the user, and it tries to identify what type of tourist activities the user is interested in. Therefore, we intend to develop a touristic POIs recommender for future study by combining the proposed method with context-aware techniques.

References

Abbaspour, R.A., Samadzadegan, F., 2011. Time-dependent personal tour planning and scheduling in metropolises. Expert Systems with Applications 38, 12439–12452.

Barta, R., Feilmayr, C., Pröll, B., Grün, C., Werthner, H., 2009. Covering the semantic space of tourism: an approach based on modularized ontologies, in: Proceedings of the 1st Workshop on Context, Information and Ontologies, pp. 1–8.

Batet, M., Moreno, A., Sánchez, D., Isern, D., Valls, A., 2012. Turist@: Agent-based personalised recommendation of tourist activities. Expert Systems with Applications 39, 7319–7329.

Borràs, J., Moreno, A., Valls, A., 2014. Intelligent tourism recommender systems: A survey. Expert Systems with Applications 41, 7370–7389.

Büyüközkan, G., Ergün, B., 2011. Intelligent system applications in electronic tourism. Expert systems with applications 38, 6586–6598.

Carrillo, D., López, V.F., Moreno, M.N., 2013. Multi-label classification for recommender systems. Trends in Practical Applications of Agents and Multiagent Systems , 181–188.

Chen, X., Vorvoreanu, M., Madhavan, K., 2014. Mining social media data for understanding students' learning experiences. IEEE Transactions on learning technologies 7, 246–259.

Chiang, H.S., Huang, T.C., 2015. User-adapted travel planning system for personalized schedule recommendation. Information Fusion 21, 3–17.

Elhassan, A., Jenhani, I., Brahim, G.B., 2018. Remedial actions recommendation via multi-label classification: a course learning improvement method. International Journal of Machine Learning and Computing 8, 583–588.

Ganda, D., Buch, R., 2018. A survey on multi label classification. Recent Trends in Programming Languages 5, 19–23.

Gao, M., Liu, K., Wu, Z., 2010. Personalisation in web computing and informatics: Theories, techniques, applications, and future research. Information Systems Frontiers 12, 607–629.

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R., Gómez-Berbís, J.M., 2009. Speta: Social pervasive e-tourism advisor. Telematics and informatics 26, 306–315.

Grün, C., Neidhardt, J., Werthner, H., 2017. Ontology-based matchmaking to provide personalized recommendations for tourists, in: Information and Communication Technologies in Tourism 2017. Springer, pp. 3–16.

Hsu, F.M., Lin, Y.T., Ho, T.K., 2012. Design and implementation of an intelligent recommendation system for tourist attractions: The integration of ebm model, bayesian network and google maps. Expert Systems with Applications 39, 3257–3264.

Huang, Y., Bian, L., 2009. A bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the internet. Expert Systems with Applications 36, 933–943.

Hug, N., 2020. Surprise: A python library for recommender systems. Journal of Open Source Software 5, 2174.

Khalid, H., Wu, S., 2016. Reducing the cold-start problem by explicit information with mathematical set theory in recommendation systems. International Journal of Engineering and Computer Science 5, 17613–17626.

Kolahkaj, M., Harounabadi, A., Nikravanshalmani, A., Chinipardaz, R., 2020. A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining. Electronic Commerce Research and Applications 42, 100978.

Kopalle, P.K., Lehmann, D.R., 1997. Alpha inflation? the impact of eliminating scale items on cronbach's alpha. Organizational Behavior and Human Decision Processes 70, 189–197.

Koren, Y., 2010. Factor in the neighbors: Scalable and accurate collaborative filtering. ACM Transactions on Knowledge Discovery from Data (TKDD) 4, 1–24.

Loh, S., Lorenzi, F., Saldaña, R., Licthnow, D., 2003. A tourism recommender system based on collaboration and text analysis. Information Technology & Tourism 6, 157–165.

Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G., 2015. Recommender system application developments: a survey. Decision Support Systems 74, 12–32.

Lucas, J.P., Luz, N., Moreno, M.N., Anacleto, R., Figueiredo, A.A., Martins, C., 2013. A hybrid recommendation approach for a tourism system. Expert systems with applications 40, 3532–3550.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A., 2021. A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR) 54, 1–35.

Moreno, A., Valls, A., Isern, D., Marin, L., Borràs, J., 2013. Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. Engineering applications of artificial intelligence 26, 633–651.

Neidhardt, J., Seyfang, L., Schuster, R., Werthner, H., 2015. A picture-based approach to recommender systems. Information Technology & Tourism 15, 49–69.

Nilashi, M., bin Ibrahim, O., Ithnin, N., Sarmin, N.H., 2015. A multi-criteria collaborative filtering recommender system for the tourism domain using expectation maximization (em) and pca–anfis. Electronic Commerce Research and Applications 14, 542–562.

Omar, A., Mahmoud, T.M., Abd-El-Hafeez, T., Mahfouz, A., 2021. Multi-label arabic text classification in online social networks. Information Systems 100, 101785.

Prantner, K., Ding, Y., Luger, M., Yan, Z., Herzog, C., 2007. Tourism ontology and semantic management system: State-of-the-arts analysis. IADIS International Conference: IADIS .

Ricci, F., Arslan, B., Mirzadeh, N., Venturini, A., 2002. Itr: a case-based travel advisory system, in: European Conference on Case-Based Reasoning, Springer. pp. 613–627.

Ricci, F., Rokach, L., Shapira, B., 2015. Recommender systems: introduction and challenges, in: Recommender systems handbook. Springer, pp. 1–34.

Rivolli, A., Parker, L.C., de Carvalho, A.C., 2017. Food truck recommendation using multi-label classification, in: EPIA Conference on Artificial Intelligence, Springer. pp. 585–596.

Rokach, L., Schclar, A., Itach, E., 2014. Ensemble methods for multi-label classification. Expert Systems with Applications 41, 7507–7523.

Sanden, C., Zhang, J.Z., 2011. Enhancing multi-label music genre classification through ensemble techniques, in: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp. 705–714.

Sarkar, J.L., Majumder, A., Panigrahi, C.R., Roy, S., 2020. Multitour: a multiple itinerary tourists recommendation engine. Electronic Commerce Research and Applications 40, 100943.

Schulz, A., Mencía, E.L., Schmidt, B., 2016. A rapid-prototyping framework for extracting small-scale incident-related information in microblogs: Application of multi-label classification on tweets. Information Systems 57, 88–110.

Sebastia, L., Garcia, I., Onaindia, E., Guzman, C., 2009. e-tourism: a tourist recommendation and planning application. International Journal on Artificial Intelligence Tools 18, 717–738.

Sorower, M.S., 2010. A literature survey on algorithms for multi-label learning. Oregon State University, Corvallis 18, 1–25.

Spyromitros, E., Tsoumakas, G., Vlahavas, I., 2008. An empirical study of lazy multilabel classification algorithms, in: Hellenic conference on artificial intelligence, Springer. pp. 401–406.

Sun, Z., Guo, Q., Yang, J., Fang, H., Guo, G., Zhang, J., Burke, R., 2019. Research commentary on recommendations with side information: A survey and research directions. Electronic Commerce Research and Applications 37, 100879.

Tidake, V.S., Sane, S.S., 2018. Multi-label classification: a survey. International Journal of Engineering and Technology 7.

Tsoumakas, G., Katakis, I., 2007. Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM) 3, 1–13.

Tsoumakas, G., Katakis, I., Vlahavas, I., 2009. Mining multi-label data, in: Data mining and knowledge discovery handbook. Springer, pp. 667–685.

Vansteenwegen, P., Souffriau, W., Berghe, G.V., Van Oudheusden, D., 2011. The city trip planner: an expert system for tourists. Expert Systems with Applications 38, 6540–6546.

Venturini, A., Ricci, F., 2006. Applying trip@ dvice recommendation technology to www. visiteurope. com. Frontiers in Artificial Intelligence and Applications 141, 607.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., Xu, W., 2016. Cnn-rnn: A unified framework for multi-label image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2285–2294.

Wen, Z., 2008. Recommendation system based on collaborative filtering. CS229 lecture notes .

Yang, W.S., Hwang, S.Y., 2013. itravel: A recommender system in mobile peer-to-peer environment. Journal of Systems and Software 86, 12–20.

Yong, A.G., Pearce, S., et al., 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology 9, 79–94.

Zhang, J., Zhang, Z., Wang, Z., Liu, Y., Deng, L., 2018. Ontological function annotation of long non-coding rnas through hierarchical multi-label classification. Bioinformatics 34, 1750–1757.

Zheng, Y., Mobasher, B., Burke, R., 2014. Context recommendation using multi-label classification, in: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE. pp. 288–295.

Zhu, G., Wu, Z., Wang, Y., Cao, S., Cao, J., 2019. Online purchase decisions for tourism e-commerce. Electronic Commerce Research and Applications 38, 100887.

Zins, A.H., 2007. Exploring travel information search behavior beyond common frontiers. Information Technology & Tourism 9, 149–164.