IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 11, 2022, accepted May 21, 2022, date of publication May 27, 2022, date of current version June 2, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3178439

Normalized Attraction Travel Personality Representation for Improving Travel Recommender Systems

TURKI ALENEZI[1,2] **AND STEPHEN HIRTLE**[1]

[1]School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260, USA
[2]Department of Management Information Systems, College of Business Administration, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

Corresponding author: Turki Alenezi (tga7@pitt.edu)

ABSTRACT Travel recommender systems (TRSs) aim to reduce travel-related search overload. A significant part of a TRS is representing attractions in a way that reflect the explicit and implicit features of attractions. However, traditional attraction representation methods may not provide a complete image of attractions. Building on the notions of user travel styles (UTSs) and the wisdom of crowds, we propose a method derived from topic-model-based models to represent travel attractions, called the Normalized Attraction Travel Personality (NATP) representation. This approach attempts to leverage the semantics of attraction reviews to model user travel personalities (UTPs), which collectively can construct the attraction travel personality (ATP) representation. Furthermore, we regularize and normalize the ATP representation to obtain our proposed representation. This NATP-based attraction representation could capture implicit characteristics of attractions revealed by the wisdom of crowds. Our experiments show that our representation method gained better results when evaluated against comparative approaches in terms of rating prediction and recommendation ranking quality, indicating the effectiveness of the proposed attraction representation. Lastly, we qualitatively investigate how our attraction representation surpasses the state-of-the-art representation methods.

INDEX TERMS Content-based filtering, attraction representation, knowledge discovery, travel styles, travel recommender systems.

I. INTRODUCTION

One of the first things people think about when planning a vacation is choosing the location. Some people have their hearts set on a destination beforehand. Others are indecisive and seek help to select a destination that reflects their interests via various information sources, primarily online-based information sources. To choose the ideal destination, tourists might go through an information search, which can be overwhelming due to the choice overload and amount of information available on the Internet. Moreover, people choosing a destination from a large choice set were more uncertain and less satisfied with their choice than people choosing from a smaller choice set [1].

Researchers have utilized various recommendation techniques to develop travel recommender systems (TRSs) in the past two decades, aiming to recommend interesting attractions to users for marketing purposes or reduce cognitive information overload when searching for a particular item on the Internet. However, TRSs still borrow recommendation techniques from other domains hoping for similar efficiency. That might not be achieved since travel items seem more complex, and user travel preferences are more challenging to elicit [2].

Attraction representation and user profiling are essential components to build an effective TRS. Attractions should be appropriately represented so that the TRS can exploit their implicit and explicit characteristics in the user profiling process. A user profile models the preferences and interests of a certain user from the user interaction with the system, such as past trips and past user history actions. Most TRSs utilize user ratings, check-ins, meta-data, and reviews in the modeling process. It is significant to understand tourists' preferences and efficiently represent items to make relevant and diverse recommendations that the user may like. Similar to other domains, several approaches [3], [4] have utilized

The associate editor coordinating the review of this manuscript and approving it for publication was Vijay Mago.

user interaction with the system to implicitly elicit user preferences for those who have records in the system and used demographic information for new users to mitigate the cold-start problem.

Furthermore, users usually differ in their background knowledge, mental models, and capabilities to express their preferences [2]. Especially in the first stages of travel decision-making, users may not even be aware of their needs and preferences, so implicitly eliciting user preferences could be more promising [5]. Once attractions and users are modeled, recommendations can be made by various traditional recommendation algorithms, such as content-based filtering, collaborative filtering, and hybrid filtering [6].

Nevertheless, these traditional techniques have been criticized in the tourism literature due to the complexity of the tourism domain. Therefore, applying these methods directly to the tourism domain may not be the best solution. For instance, if two persons bought the same book, it is reasonable to infer that they like the same type of book. However, if two persons visited a museum, they may not necessarily be history buffs. With travel, it is even less likely that two users experienced the same trip [7]. Moreover, since traveling is a costly, time-consuming activity, travel-related activities are less frequent than watching movies or buying products in less complex domains, making attraction representation and user preferences elicitation challenging tasks [2], [5].

Several attempts in the literature of TRSs have based their methods on user travel styles (UTSs) to understand users' travel preferences [5], [8]. A relationship between UTSs and attractions has been identified, suggesting a predictive power of UTSs [9], [10]. UTSs are believed to provide a better representation of users' travel preferences and, therefore, more relevant travel-related recommendations [9]. However, to our knowledge, there is a lack of research focusing on the representation of attractions using methods explicitly developed for the tourism domain despite its applications in many TRSs. Therefore, this paper aims to introduce an attraction representation method called the Normalized Attraction Travel Personality (NATP), which accommodates the complexity of the tourism domain by incorporating UTSs into the representation of attractions.

This representation is inspired by the notions of the wisdom of crowds and UTSs. Here, we make an important assumption as follows: not only can attractions be described from their user-generated content (reviews) but also from the collective UTSs. To illustrate, if an attraction is visited mostly by beachgoers, it is reasonable to argue that this place is, in fact, a beach or water-related activity. This is consistent with the standard user-generated content modeling. However, if a more diverse population with different user travel personalities (UTPs) visited the attraction, this could reveal interesting patterns, such as a theme park that is highly visited by history buffs. Although this attraction is explicitly recognized as a theme park, it can also be described as a historical place from the perspective of its users' collective travel personalities. This type of information is revealed implicitly by the type of users who visit a place. Thus, we believe they have a role in describing attractions from a touristic standpoint rather than a linguistic one.

We summarize our contributions as follows:

1) This paper introduces an attraction representation method particularly devised for recommender systems in the tourism domain.
2) We propose a normalization function that can regularize and convert a vector of real numbers to a probability distribution.
3) We show the effectiveness of our representation over state-of-the-art attraction representation methods by building a prototype content-based recommender system in addition to comparing similar attractions to example attractions using a real-world dataset.

II. RELATED WORK

A. TRAVEL RECOMMENDER SYSTEMS

We can broadly categorize TRSs into two categories. The first category is TRSs that employ traditional recommendation filtering techniques. Traditional recommendation systems are basically based on the user history, including reviews and ratings of past visits, to indicate the degree of interest of the attractions for content-based approaches and to determine similarities among users for collaborative approaches [3], [4], [8], [11]. Mainly, most methods of traditional recommendation systems fall under one of four categories: content-based filtering, collaborative filtering, demographic filtering, or hybrid filtering [6].

Content-based methods make recommendations based on the content of an item, typically textual content, such as reviews, comments, tweets, and tags. Classical text vectorization methods (e.g., bag-of-words and TFIDF) have been used to represent attractions [12]. Furthermore, natural language processing (NLP) methods (e.g., topic-model-based methods and word/document embedding) have been utilized to represent attractions, showing some improvement over classical text vectorization methods [13]–[16]. Several works attempt to exploit the content of external sources (e.g., Wikipedia and OpenStreetMap) to build a concept network and knowledge base [17]. However, the main problem with content-based approaches is the filter bubble problem [18], where the system gets over-tailored to the user's interest and hides possibly other relevant, interesting items. As a result, the recommendations become less interesting over time, predictable, and sometimes annoying since recommendations are too similar to previous items in the users' history.

Collaborative filtering methods try to solve this problem by recommending items that like-minded users are interested in. They make use of ratings as the major source of information about users and items [19]. However, the data sparsity and the cold start problems make it difficult to measure similarity and perform decent classification in memory-based and model-based algorithms, respectively [6]. Collaborative filtering pays less attention to an accurate representation of items to avoid the limitations of content-based filtering methods.

Nevertheless, attraction textual representation can play a role in collaborative filtering by basing the attraction similarity matrix on the similarity of the textual representation instead of ratings [12].

Demographic information is usually used in TRSs, especially to solve the cold-start problem. One approach tried to make recommendations only based on demographic information about users to predict their ratings [20]. However, their results were so poor, and they advised utilizing the rich content of online reviews to represent items and understand users. Demographic information may enhance the quality of recommendations when incorporated as a part of a hybrid recommender system [21].

Hybrid approaches [4], [8], [14], [22] have been proposed in the literature to overcome the downsides of the aforementioned recommendation methods and improve recommendation performance. To address the data sparsity and the cold-start problems, Ameen *et al.* [16] integrates convolutional neural network into weighted matrix factorization, yet they use a standard attraction representation method, LDA [23]. The quality of recommendations of such systems remains questionable. This might be due to the fact that user profiling and item representation are more difficult to establish in the tourism domain, and researchers are still using standard item representation methods that might not fit the tourism domain.

The second category of TRSs is location-based recommender systems, including points of interest (POIs), travel destination, and context-aware recommender systems (CARSs). Most of POIs recommender systems tend to recommend nearby locations to the most recent check-in or the active user's current location [21], [24]–[26]. This can be helpful in the late stages of the travel decision. Less work has been published in the travel domain, where the task is to recommend attractions in different cities to help in the destination choice problem.

Conversational recommender systems [7], [27]–[29] attempt to elicit user preferences via conversational dialogues in order to recommend possible interesting destinations. The construction of user profiles is dependent on travel preferences entered explicitly by the user. Although such systems could recommend highly relevant places, they require a great deal of travel needs data entry which can be overwhelming and impractical.

CARSs focus on the representation of various types of the user's context information, such as current location, time, weather, and mood. Many context-aware recommenders prioritize the effectiveness of the context information representation over the attraction representation [30], [31]. CARSs and POIs recommender systems may not be practical when the user needs are travel destination recommendations.

B. USER TRAVEL STYLES

In the tourism literature, the notion of UTSs has been discussed widely and considered a significantly influential factor in understanding differences in traveler behavior [10]. Plog [32] was one of the first who related the psychology of people with destination and attraction preferences. He found that the population from allocentrism to psychocentrism is normally distributed. Allocentric travelers show high activity levels and prefer non-touristic and novel areas, while psychocentric travelers show low activity levels and prefer familiar touristic destinations.

Gretzel *et al.* [9] investigated whether UTSs can predict preferred attractions that could be recommended by a TRS. Participants were asked to choose from a list of 12 UTSs (e.g., sight seeker, boater, and history buff) the one that best describes them, and the results showed a close correspondence between UTSs and attractions in the last touristic trip. They found that a single UTS (e.g., culture creature) is often related to more than one attraction (e.g., museums and festivals). They argue that this kind of user engagement in the recommendation process is an effective strategy to capture tourist preferences, and it can serve as a preference structure providing a valid shortcut to more personalized recommendations.

Huang and Bian [33] used a Bayesian network to estimate the user's preferred activities. Two variables influenced the preferred activities, trip motivation (e.g., learning something new) and traveler type (e.g., adventurer, urban, and relaxation seeker). The latter is influenced by three variables: age, occupation, and psychographic personality types outlined in [32].

Park *et al.* [10] defined the travel persona of a user as a combination of travel interests that distinguish travel behaviors. Out of 20, participants chose up to three UTSs that best describe them and up to three travel attractions from recent trips to U.S. destinations. The results indicated that different travel personas are associated with different travel behaviors, suggesting a useful way to understand travelers and, therefore, enhancing the development of personalized destinations.

Braunhofer *et al.* [21] built a hybrid TRS combining demographic and human personality traits with collaborative filtering in the process of rating prediction. They relied on explicit personality acquisition since it has higher accuracy and takes less effort from users to provide their personality information.

Pantano *et al.* [34] used a subset of TripAdvisor's travel style tags of users who rated the Empire State Building and tried to predict whether a user would like another attraction or not based on their UTSs. The UTS feature has been discontinued at TripAdvisor since early 2019.

Neidhardt and Werthner [35] conducted a factor analysis to reduce the dimensionality of user profiles from 22 variables to a seven-factor solution. The variables are 17 UTSs and the Big Five personality traits, which are extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience [36]. They argued that a tourist profile could be interpreted as a combination of the seven factors (i.e., Sun & Chill-Out, Knowledge & Travel, Independence & History,

Culture & Indulgence, Social & Sport, Action & Fun, and Nature & Recreation).

To deal with the challenges in the tourism domain with respect to eliciting travel preferences, Neidhardt *et al.* [5] built a picture-based recommender system to implicitly elicit travel preferences by having users select three to ten travel-related pictures that are most appealing to them. Pictures were preselected and evaluated with regard to their relationship to the seven factors. In their study, they also assigned up to seven pictures to 10,835 attractions. The aim was to construct a user profile and attractions represented by the seven factors and apply a distance metric to make tourism recommendations.

Previous studies highlight the importance of understanding and eliciting tourist preferences. However, the literature pays less attention to the role of attraction representation in improving TRSs. To the best of our knowledge, the only work that has developed an attraction representation approach suitable for the tourism domain to address the unique characteristics of attractions is a season topic model based on LDA [37], adding a season layer between the topic and document layers of LDA.

In this work, we take a different approach in which we manipulate the attraction types learned by any topic model, both existing or in the future, to refine the characteristics of attractions by removing and reordering the attraction topics based on their new values obtained by our method. Since the source of information in existing attraction representation approaches is limited to the textual content of attractions, their resulting representation would inherit the drawbacks of content-based filtering even when applying state-of-the-art text topic modeling and embedding techniques. To alleviate the drawbacks of content-based filtering and accommodate the complexity of the tourism domain, our approach represents attractions by aggregating the adjusted travel personalities of users to remove noise and discover new descriptions of attractions that might not be explicit in their textual content.

III. TRAVEL ATTRACTION REPRESENTATION

A. PRELIMINARIES AND PROBLEM DEFINITION

Definition 1 (Attractions): Attractions are travel-related places or activities except restaurants and hotels. We define the set of attractions as $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$, where N is the total number of attractions in the dataset.

Definition 2 (Users): Users are the tourists who express their opinions and experiences about an attraction by posting reviews and ratings on its webpage. We define the set of users as $\mathcal{U} = \{u_1, u_2, \ldots, u_M\}$, where M is the total number of users in the dataset.

Definition 3 (Dataset): Let r_{ij} be the rating of attraction a_j given by user u_i and v_{ij} be the review posted on attraction a_j's webpage by user u_i. We define the dataset $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$, where each element is a tuple: $(u_i, a_j, r_{ij}, v_{ij})$ indicating the latest opinion of user u_i on attraction a_j.

Definition 4 (Attraction Reviews): Let v_{a_i} be a document that contains the aggregated user reviews posted on attraction a_i. We define $\mathcal{V} = \{v_{a_1}, v_{a_2}, \ldots, v_{a_N}\}$ as the collection of attractions' reviews.

Definition 5 (Attraction Type): An attraction type is an explicit characteristic of the attraction describing what the attraction is about (e.g., waterfalls).

Definition 6 (Attraction Type Representation): Let θ_{ij} indicates to what extent the jth attraction type is present in attraction a_i. We define the attraction type matrix $\boldsymbol{\Theta} \in \mathbb{R}^{N \times T}$, where $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iT})$ is a vector of T dimensions representing the types of attraction a_i.

Definition 7 (User Travel Style): A UTS is a trait that is linked to those interested in a particular attraction type (e.g., waterfalls lover).

Definition 8 (User Travel Personality): Let x_{ij} indicates to what extent the jth UTS is relevant to user u_i. We define the UTP matrix $\boldsymbol{X} \in \mathbb{R}^{M \times T}$, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})$, the UTP of user u_i, is a vector of T dimensions representing the UTSs of user u_i.

Definition 9 (Attraction Travel Style (ATS)): An ATS is a feature that measures to what extent an attraction attracts users who are represented by a particular travel style (e.g., waterfalls lovers magnet).

Definition 10 (Attraction Travel Personality (ATP)): Let y_{ij} be the jth ATS of attraction a_i. We define the ATP matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$, where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$, the ATP of attraction a_i, is a vector of T dimensions representing the ATSs of attraction a_i.

Definition 11 (Problem Definition): Given $\mathcal{D}$ and $\mathcal{V}$, our aim is to represent attractions by their NATPs. We define the NATP matrix $\boldsymbol{Z} \in \mathbb{R}^{N \times T}$, where $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iT})$ is the ATS probability distribution of attraction a_i.

We argue that this representation could uncover pieces of knowledge about attractions while preserving their explicit characteristics. For example, Central Park's extracted types using topic-model-based methods could be nature, city exploration, and outdoor activities. At the same time, its travel personality could potentially discover additional latent information, such as art, paintings, and hiking. Finally, we will evaluate the effectiveness of our proposed representation by building an attraction recommender prototype and analyzing the top similar attractions of real-world examples.

The notations and symbols used in this work are described in Table 1, and a framework of the proposed representation is shown in Fig. 1.

B. USER TRAVEL PERSONALITY

1) EXTRACTING ATTRACTION TYPES

We first extract attraction types in order to infer UTSs. Intuitively, finding the UTSs requires finding the types of attractions the user has visited and may or may not show interest in. Topic-model-based methods can be utilized to extract attraction types from $\mathcal{V}$. In particular, we choose three different topic-model-based methods, Latent Dirichlet Allocation (LDA) [23], the Embedded Topic Model (ETM) [38],

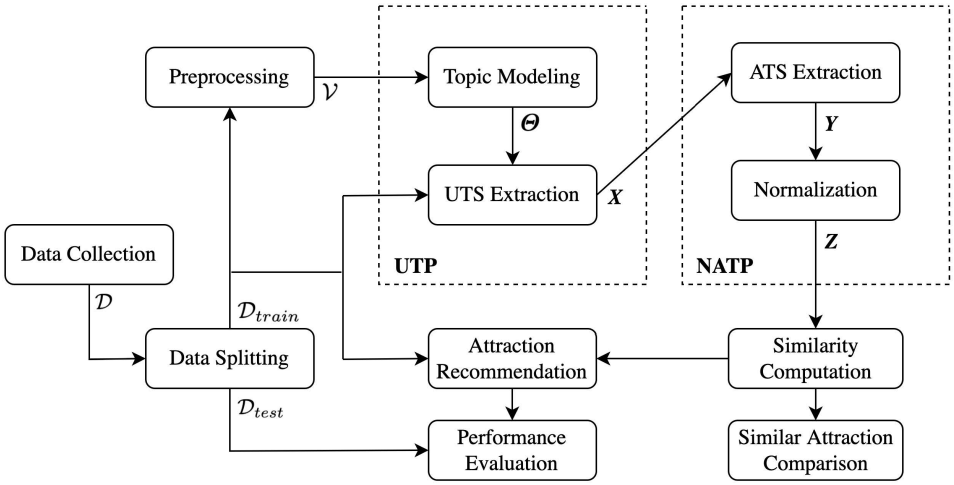


FIGURE 1. Framework of the proposed NATP-based representation for improving TRSs.

TABLE 1. Notations and symbols.

Symbol	Description
$\mathcal{A}$	set of attractions $\{a_1, a_2, ..., a_N\}$
$\mathcal{U}$	set of of users $\{u_1, u_2, ..., u_M\}$
$\mathcal{D}$	dataset of users' reviews and ratings of attractions
$\mathcal{V}$	dataset of attractions' reviews $\{v_{a_1}, v_{a_2}, ..., v_{a_N}\}$
$\boldsymbol{\theta}_i$	attraction type vector $(\theta_{i1}, \theta_{i2}, ..., \theta_{iT})$ of a_i
$\boldsymbol{\Theta}$	attraction type matrix
$\boldsymbol{x}_i$	UTP vector $(x_{i1}, x_{i2}, ..., x_{iT})$ of u_i
$\boldsymbol{X}$	UTP matrix
$\boldsymbol{y}_i$	ATP vector $(y_{i1}, y_{i2}, ..., y_{iT})$ of a_i
$\boldsymbol{Y}$	ATP matrix
$\boldsymbol{z}_i$	NATP vector $(z_{i1}, z_{i2}, ..., z_{iT})$ of a_i
$\boldsymbol{Z}$	NATP matrix
$\mathcal{A}_i$	set of attractions visited by the user u_i
$\mathcal{U}_j$	set of users who visited attraction a_j
r_{ij}	observed rating of u_i for a_j
μ_{r_i}	mean of the observed ratings of u_i
$\hat{r_{ij}}$	predicted rating of u_i for a_j
$sim(j, l)$	similarity between the representations of a_j and a_l

and Distributed Representations of Topics (Top2Vec) [39]. In this paper, we use document topics and attraction types interchangeably.

We choose LDA for its popularity and wide success in such tasks. The main assumption of LDA is that each document comes from a distribution of topics, and each topic comes from a distribution of words. Taking $\mathcal{V}$ as input, the model estimates two matrices. The first is the attraction-type matrix $\boldsymbol{\Theta}$, where each attraction is represented as a probability distribution over T topics. The second matrix is the type-word matrix $\boldsymbol{\Phi} \in \mathbb{R}^{T \times |\mathcal{W}|}$, where $\mathcal{W}$ is the set of unique words in $\mathcal{V}$, and $\boldsymbol{\varphi_i} = (\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{i|\mathcal{W}|})$ is a vector of $|\mathcal{W}|$ dimensions representing the probability distribution over words of the ith topic. $\boldsymbol{\Phi}$ allows us to examine the most probable words, or top words, in each attraction type. This can be used to interpret the various attraction types and measure their quality. However, despite its popularity, LDA suffers from limitations, including treating documents as bags of words, which ignores sentence structure, and the assumption that the optimal number of topics is given as input beforehand.

A more recent generative probabilistic model of textual documents, ETM, attempts to combine the strengths of LDA with word embeddings, namely Word2Vec [40], in an efficient manner. In lieu of the bag-of-words model used in LDA, ETM employs word embeddings, a continuous representation of words, to model the meaning of words. Topics are defined as points in the same embedding space. The main assumption of this method is that words and topics exist in the same embedding space. An entry of the word-by-topic matrix is represented by the inner product between the word's vector and its assigned topic's vector. Just like LDA, ETM requires the number of topics to be specified as a priori parameter.

On the other hand, Top2Vec does not expect prior knowledge of the number of topics. It automatically finds the number of topics by finding the dense areas of documents in the semantic space based on the assumption that the number of dense areas of documents equals the number of topics. The backbone of this method is Doc2Vec [41], a neural network-driven approach extended from Word2Vec and used to embed the textual content of documents in a low-dimensional space in which the overall meaning of each unique document is represented as a dense vector in the space. Since Doc2Vec jointly learns the embeddings of documents and words in the same vector space, similar documents, ideally, are placed close to each other. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is then applied on the dimension-reduced documents. Subsequently, topic vectors are calculated from the resulting clusters, whose total number approximates the number of topics. A topic vector is described by its closest word vectors, and a given document can be represented by a vector whose components are the distances to all discovered topic vectors.

The ijth entry of $\boldsymbol{\Theta}_{\text{LDA}}$ and $\boldsymbol{\Theta}_{\text{ETM}}$ indicates the probability that the jth attraction type is relevant to the ith attraction. However, the ijth entry of $\boldsymbol{\Theta}_{\text{Top2Vec}}$ equals the dot product between the ith attraction embedding vector learned by Doc2Vec and the jth attraction type embedding vector learned by Top2Vec.

2) EXTRACTING USER TRAVEL STYLES

Topic-model-based methods represent each attraction as a vector of T dimensions corresponding to attraction types. Let $\mathcal{A}_i$ be the set of attractions visited by the user u_i. We compute the UTS of user u_i from the attraction type vectors in $\mathcal{A}_i$ as follows:

$$\boldsymbol{x}_i = \sum_{j \in \mathcal{A}_i} \boldsymbol{\theta}_j (r_{ij} - \mu_{r_i}) \tag{1}$$

where μ_{r_i} is the mean of user u_i observed ratings.

Mean-centering the raw ratings is a crucial step that allows the transformed ratings to be negative, 0, or positive. This should adjust the UTS score depending on the user's opinion about the attraction. The assumption here is that users rate an attraction less than their average rating if it does not match their UTPs.

This mixture of positive and negative UTS scores shapes the UTP representation. Not only does this representation represent user travel preferences, but it also conveys the level of travel experience a user has. We choose not to normalize the UTP vectors because, according to (1), users who rate many attractions would have higher UTS scores, especially for their favorite travel styles, which suggests the level of experience. That could be beneficial when constructing the ATP representation.

C. NORMALIZED ATTRACTION TRAVEL PERSONALITY

Every UTP vector $\boldsymbol{x}$ contributes negatively or positively in constructing the ATP description they visit. The magnitude of contribution is dependent on the level of experience a user has. After obtaining the UTP matrix $\boldsymbol{X}$, we compute the ATP matrix $\boldsymbol{Y}$ by computing each ATP vector as follows:

$$\boldsymbol{y}_j = \sum_{i \in \mathcal{U}_j} \boldsymbol{x}_i (r_{ij} - \mu_{r_i}) \tag{2}$$

where $\mathcal{U}_j$ is the set of users who visited attraction a_j.

The resulting $\boldsymbol{y}_j$ is a spectrum of ATSs ranging from the highly irrelevant ATSs (i.e., negative components) to the highly relevant ATSs (i.e., positive components). To get our target representation of attractions $\boldsymbol{Z}$, we need to normalize $\boldsymbol{Y}$ with the least data distortion possible to maintain the underlying distribution of ATSs. There are multiple ways to achieve that, but the transformed vectors may not scale accurately. The softmax function is commonly used in such cases, but we find the resulting vectors highly distorted by setting most values to zero except the highest few values. When tuning the temperature parameter, the distortion is mitigated. However, this approach is impractical since each ATP vector has its own optimal temperature parameter value. Besides, there is no conceivable way to control the effect of the ATP vectors' negative values. Therefore, we propose a normalization method that consists of two steps. The first step is to transform $\boldsymbol{y}_j$ to a non-negative vector $\boldsymbol{\omega}_j = (\omega_{j1}, \omega_{j2}, \ldots, \omega_{jT})$ using the following formula:

$$\omega_{jn} = y_{jn} + \lambda |y_{jn}|, \quad n = 1, 2, \ldots, T \tag{3}$$

where $1 \leq \lambda \leq 2$ is is a regularization parameter controlling the influence of the negative components of $\boldsymbol{y}_j$. When $\lambda = 2$, positive components of $\boldsymbol{y}_j$ are tripled, and the negative components are converted to positive, thereby maintaining the original differences between ATSs. As we decrease λ, the effect of the originally negative components decreases, and they become zero when $\lambda = 1$. This allows us to control the influence of the irrelevant ATSs or eliminate them from the attraction representation.

The second step is to normalize the resulting vector so that its entries add up to one, resembling the representation of probabilistic topic-model-based representations:

$$\boldsymbol{z}_j = \frac{\boldsymbol{\omega}_j}{\sum_{n=1}^{T} \omega_{jn}} \tag{4}$$

We derive our NATP-based representation from three different topic-model-based methods (i.e., LDA, ETM, and Top2Vec) to obtain three variants of our representation (i.e., NATP_{LDA}, NATP_{ETM}, and $\text{NATP}_{\text{Top2Vec}}$)

IV. EXPERIMENTAL RESULTS

In this section, we investigate the quality of our NATP-based representation quantitatively and qualitatively in several ways. First, we build a prototype content-based recommender system to explore the predictive power and recommendation quality of our NATP-based representation against state-of-the-art content-based representation methods. Second, we show how our representation's performance responds to different values of λ. Finally, we retrieve the top three similar attractions to three example attractions to explain how our representation differs from traditional attraction representation methods.

A. DATASET

Tourism-related opinions and experiences are publicly available on TripAdvisor. We scraped user reviews, including ratings, from TripAdvisor's attractions located in the United States and distributed over all 50 states. As done in previous works [11], [15], we filter the data by keeping users who have more than 30 reviews in order to construct UTPs from a reasonable amount of information. Moreover, we remove some attractions from the dataset because they are either closed permanently or moved to another location. Because users can review attractions multiple times, we consider only the latest review to obtain the most up-to-date opinion of a user about a place. The resulting dataset consists of 7,798 attractions and 13,232 users, with an average of 90 reviews per attraction and 53 reviews per user. The total number of

TABLE 2. Descriptive statistics of the training and test datasets.

Descriptive measure	Attractions		Users	
	Training	Test	Training	Test
Minimum	1	1	24	7
Median	36	10	33	9
Mean	72	19	42	11
Maximum	2288	437	558	140
Std Dev	119	29	28	7

TABLE 3. Topic coherence scores for the topic-model-based methods.

Model	C_{UCI}	C_{UMASS}	C_{NPMI}	C_V
ETM	0.596	-0.772	0.075	0.510
LDA	0.652	-1.484	0.134	0.635
Top2Vec	2.612	-1.589	0.261	0.862

reviews in the dataset is 705,324, posted between 2003 and 2018. Data preprocessing is performed to prepare the dataset for our experiments and analysis, and a list of stop words is used to remove meaningless words. In addition, most frequent and rare words were removed as well, resulting in a total of 11,640 unique words.

B. EXPERIMENTAL PROCEDURES

To evaluate our approach against comparative methods, we split the dataset into training and testing sets according to the following procedure. We randomly select 20% of the observations for evaluation for each user, and all the rest serve as training data. This procedure results in roughly 20% testing set and 80% training set. Table 2 shows the data description of each split.

In order to estimate $\boldsymbol{\Theta}$ in a comparative manner, we first apply Top2Vec to infer the optimal number of attraction types T and then use it as input for LDA and ETM. The document-representation retrieval is not completely implemented in the source codes of ETM and Top2Vec, so we made slight modifications to retrieve document vectors. As a result, we find a total of 114 attraction types, meaning 114 clusters in the attractions semantic space. Furthermore, we utilize topic coherence metrics [42]–[45] to measure the quality of the discovered topics by each topic-model-based method. As shown in Table 3, most topic coherence metrics suggest that the attraction types discovered by Top2Vec have the best quality, which raises our confidence in its estimated number of attraction types. Note that lower C_{UMASS} values mean better topic coherence according to Gensim's implementation.

Next, we build a prototype content-based recommendation system to evaluate the performance of our NATP-based representation against baselines. We adopt the content-based filtering method [46] which predicts the rating of user u_i on attraction a_j in the test set as:

$$\hat{r_{ij}} = \frac{\sum_{l \in \mathcal{A}_i^N} sim(j, l)\, r_{il}}{\sum_{l \in \mathcal{A}_i^N} |sim(j, l)|} \tag{5}$$

where

$$sim(j, l) = \frac{z_j \cdot z_l}{||z_j||\,||z_l||} \tag{6}$$

and $\mathcal{A}_i^N$ is the set of the N most similar attractions to a_j in the training set of user u_i. We choose N to be 30 as this value is commonly used in such cases among researchers [47]. For the baselines, we replace z_j with the document representation vector of v_{a_j} learned by each baseline. Building this type of recommender system whose items are travel attractions represented by our method and comparative methods allows us to evaluate the effectiveness of our NATP-based representation method in terms of ratings' predictive power and recommendation quality.

C. COMPARATIVE APPROACHES

Since we utilize topic-model-based methods (i.e., ETM, LDA, and Top2Vec) in one step of our NATP-based representations (i.e., $NATP_{ETM}$, $NATP_{LDA}$, and $NATP_{Top2Vec}$), we compare our NATP-based representation methods against the topic-model-based models. As for parameter settings, we apply default parameters for all baselines.

Moreover, we compare against the classic TFIDF representation [48], one of the earliest term-weighting algorithms used by content-based recommender systems and information retrieval systems. It is used to measure the relationship between words and text documents considering common and rare words.

Another document representation method we choose to evaluate our representation against is Doc2Vec, which, as mentioned in section 3-B, is a model used in the process of creating the Top2Vec representation. It is one of the most effective methods to represent documents and is used to represent items in content-based recommender systems across several domains, such as news recommendations [49],

In addition, we represent attractions using a state-of-the-art supervised pre-trained sentence embedding model, Universal Sentence Encoder (USE) [50], with which we compare our representation method. USE is a deep learning model that uses the transformer network [51] and is pre-trained on multiple data sources (e.g., Wikipedia and discussion forums). Several recent works utilize USE to represent textual items in various recommendation system domains [52], [53]. Moreover, in a recent study, Gawinecki *et al.* [54], find USE to be the best content-based representation method for producing movie recommendations.

$$RMSE = \sqrt{\frac{\sum_{(r_{ij}, \hat{r_{ij}}) \in \mathcal{T}} (\hat{r_{ij}} - r_{ij})^2}{|\mathcal{T}|}} \tag{7}$$

D. EVALUATION METRICS

To evaluate the effectiveness of our proposed representation, we compare the performance of content-based recommendations that resulted from our NATP-based representation methods and the comparative methods. Specifically, we want to ascertain how these methods perform regarding user

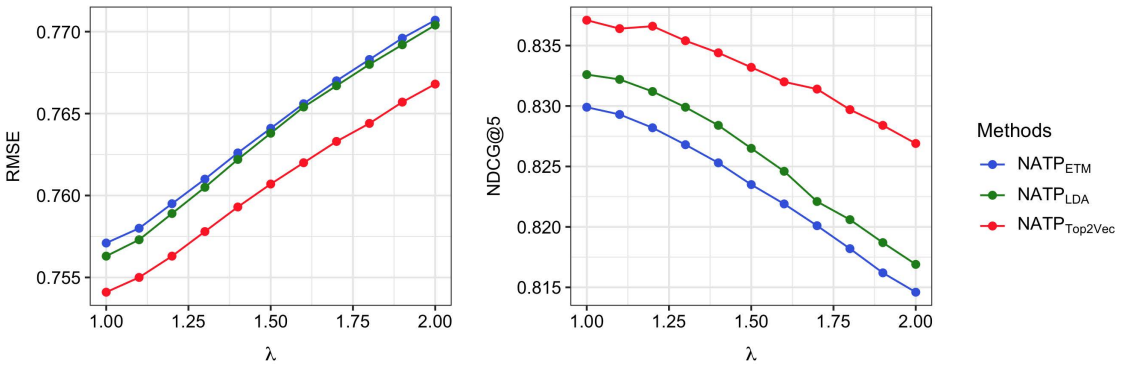


FIGURE 2. RMSE and NDCG@5 scores of recommendations using the NATP-based representation methods for different values of λ.

rating prediction and recommendations quality of the top-K recommendations using two evaluation metrics.

First, Root Mean Square Error (RMSE) measures the overall quality of predicting the ratings of a given user and attraction by computing the error between predicted ratings and observed ratings. Let $\mathcal{T}$ be the set of rating pairs of the testing dataset, where each element is a tuple: $(r_{ij}, \hat{r_{ij}})$ referring to the observed and predicted ratings for user u_i on attraction a_j. RMSE is defined as follows:

The second metric is the Normalized Discounted Cumulative Gain (NDCG@K) [55], which evaluates the ranking quality of the top-K recommended items. It measures how some items are more relevant than others by examining whether or not the highly relevant items are at the top of the recommended list. Let rel_{ij} be the true rating of the jth most relevant attraction to user u_i, and let rec_{ij} be the true rating of jth recommended attraction to user u_i. $NDCG@K$ is defined as follows:

$$NDCG@K = \frac{DCG@K}{IDCG@K} \tag{8}$$

where

$$DCG@K = \frac{1}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\sum_{j=1}^{K}\frac{2^{rec_{ij}}-1}{log_2(j+1)} \tag{9}$$

and

$$IDCG@K = \frac{1}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|}\sum_{j=1}^{K}\frac{2^{rel_{ij}}-1}{log_2(j+1)} \tag{10}$$

E. QUANTITATIVE ANALYSIS

The ATP representation consists of positive and negative ATSs before applying the normalization step. In order to construct the best-possible NATP representation, we need to find the optimal value of λ, which regularizes the influence of negative ATSs on the representation. Hence, we compare the performance of the recommendations system under different values of λ ranging from 1.0 to 2.0 with a step size of 0.1.

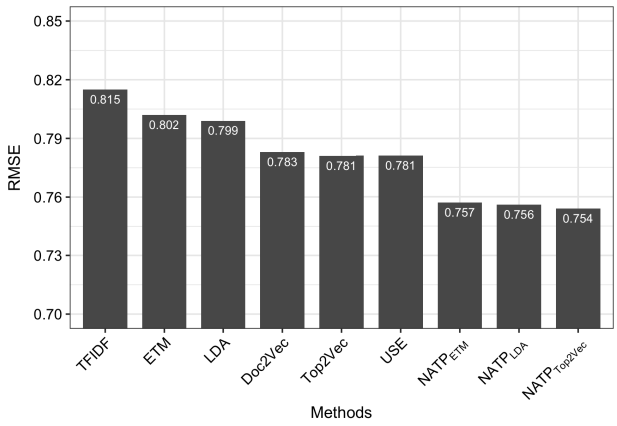


FIGURE 3. RMSE scores of the comparative models. Lower scores indicate better rating prediction.

Fig. 2 shows the RMSE and NDCG@5 scores of recommendation performance using the NATP-based representation methods. Results show that all variants obtain the best recommendation performance when $\lambda = 1.0$. This indicates that negative ATSs appear to have a deleterious effect on the representation of attractions. Therefore, we eliminate all negative ATSs by setting λ to 1. This results in a mean of 61 negative ATSs removed from NATP_{LDA} and $\text{NATP}_{\text{Top2Vec}}$, and 56 from NATP_{ETM}.

Fig. 3 and Fig. 4 display consistent results for both metrics. Fig. 3 shows that our NATP-based methods outperform the baselines in terms of the rating prediction performance. Similarly, Fig. 4 shows that the ranking quality of all variants of our NATP-based representation outperforms the comparative methods at all values of K. This indicates that our proposed method represents attractions better than state-of-the-art methods for content-based recommendation tasks.

Not surprisingly, the performance of TFIDF is relatively the worst. Interestingly, LDA outperforms ETM in terms of the recommendation task as well as the topic coherence, suggesting that ETM does not work well for our dataset. Top2Vec stands out of topic modeling methods with a

TABLE 4. Examples of attractions and their most similar ones according to the top two best performing baselines (i.e., Top2Vec and USE) versus our best performing NATP-based method (i.e., $\text{NATP}_{\text{Top2Vec}}$).

Attraction	Representation method	Top-3 similar attractions
Cedar Point *Sandusky, OH*	Top2Vec	Kings Island, *Mason, OH* Kings Dominion, *Doswell, VA* Carowinds, *Charlotte, NC*
	USE	Six Flags Magic Mountain, *Santa Clarita, CA* California's Great America, *Santa Clara, CA* Kings Island, *Mason, OH*
	$\text{NATP}_{\text{Top2Vec}}$	Kennywood Park, *West Mifflin, PA* Disney California Adventure Park, *Anaheim, CA* Disneyland Park, *Anaheim, CA*
Three Sisters Springs *Crystal River, FL*	Top2Vec	Crystal River, *Crystal River, FL* Blue Spring State Park, *Orange City, FL* Manatee Park, *Fort Myers, FL*
	USE	Blue Spring State Park, *Orange City, FL* Manatee Springs State Park, *Chiefland, FL* Manatee Park, *Fort Myers, FL*
	$\text{NATP}_{\text{Top2Vec}}$	Island Adventures, *Anacortes, WA* Island Packers, *Ventura, CA* Baxter State Park, *Millinocket, ME*
Fallingwater *Mill Run, PA*	Top2Vec	Taliesin West, *Scottsdale, AZ* Kentuck Knob, *Chalk Hill, PA* Taliesin Preservation, *Spring Green, WI*
	USE	Taliesin Preservation, *Spring Green, WI* Kentuck Knob, *Chalk Hill, PA* Taliesin West, *Scottsdale, AZ*
	$\text{NATP}_{\text{Top2Vec}}$	Boldt Castle and Yacht House, *Alexandria Bay, NY* Taliesin Preservation, *Spring Green, WI* Anderson Japanese Gardens, *Rockford, IL*

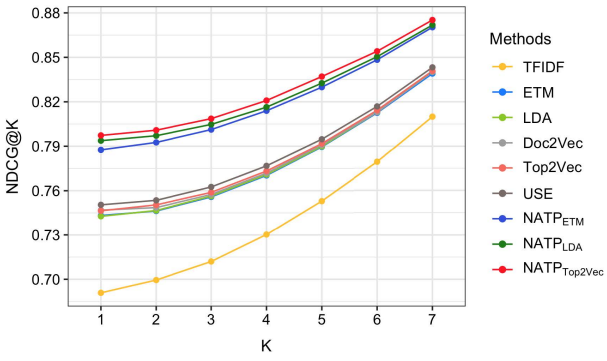


FIGURE 4. NDCG@K scores of the recommendation performance. We used $\lambda = 1$ for our NATP-based methods.

notable improvement. This can be attributed to the fact that Top2Vec utilizes Doc2Vec under the hood, which by itself shows relatively good results. Our NATP-based representation methods beat the best-performing baseline, USE, by a considerable margin on all evaluation metrics. The performance of NATP-based methods seems to be dependent on the quality of their internal topic-model-based methods that are responsible for representing UTPs. This may explain the superiority of $\text{NATP}_{\text{Top2Vec}}$ over NATP_{LDA} and NATP_{ETM}. All in all, the superior results of our representation strengthen our argument about the role of ATP in capturing attractions' latent characteristics that are not present in their textual content.

F. QUALITATIVE ANALYSIS

To examine the distinction between the baselines and our NATP-based method, Table 4 shows three example attractions and their top three similar attractions according to cosine similarity. Each example attraction and its nearest neighbors are represented by the top two best performing comparative methods (i.e., Top2Vec and USE) and the best performing NATP-based method (i.e., $\text{NATP}_{\text{Top2Vec}}$).

The first example is Cedar Point, one of the oldest theme parks in the United States, which opened in 1870. Several sites are listed on the National Register of Historic Places (NRHP) within the park's property, such as the Midway Carousel and Cedar Point Lighthouse. Cedar Point's most similar attractions retrieved by baselines as well as our method are other theme parks. By taking a closer look into the similar attractions, all baselines' theme parks were opened in the 1970s, and none of them appear to feature historic sites. By way of contrast, the most similar attraction retrieved by our NATP-based representation is Kennywood Park, which opened in 1898. The park is designated as a National Historic Landmark (NLH), and it is home to several historic structures, such as the Victoria Windmill and one of the oldest wooden roller coasters in the world that is still operating. Unlike

attraction representation baselines, our NATP-based method seems to represent Cedar Point as a theme park that has a historical significance. This extra information is what we believe brought Kennywood Park to the top of the list.

The second example is the Three Sisters Springs, a part of the Crystal River National Wildlife Refuge and one of the few places where people can view and swim with manatees. It is considered a fantastic birding site since over 100 different bird species are sighted at this attraction [56]. Furthermore, kayaking and paddleboarding are common activities in the springs. According to the baselines, its most similar attractions primarily involve bodies of water with manatees. Alternatively, we obtain more diverse attractions similar to Three Sisters Springs when represented by our NATP-based method. The first two attractions are tour companies that offer boat tours to view dolphins, whales, seals, and other marine mammals in their natural habitats in addition to birding. Baxter State Park is an expansive wilderness area with mountains and bodies of water. In addition to hiking and water sports, visitors may view a wide variety of birds and animals, such as bears, deer, and moose, swimming across a pond.

The third example is Fallingwater, an NHL and one of the most famous houses in the world. It was designed by Frank Lloyd Wright, the father of organic architecture, in the 1930s. What makes this house unique is that it was built over a waterfall in the mountains of Laurel Highlands. All similar attractions retrieved by the baselines are houses designed by Wright. On the other hand, three different attractions are found to be similar to Fallingwater by our representation. The first is Boldt Castle and Yacht House, a major attraction site in the Thousand Islands area. It is a complex of several historic structures, including the main castle nestled in Heart Island, a water gate known as the Entry Arch, a building rising out of St. Lawrence River named the Power House, and the Yacht House, which is a boathouse located on the edge of a nearby island. The second is one of Wright's famous houses and retrieved by the baselines as well. The third is Anderson Japanese Gardens, which includes streams, cascading waterfalls, koi-filled ponds, and a 16th-century Sukiya-style Guest House. Interestingly, the guest house follows the construction method of the nearby Frank Lloyd Wright-designed Laurent House.

V. CONCLUSION

In this work, we propose an NATP-based method to represent travel attractions based on their UTPs extracted from user reviews. Our representation models an attraction in such a way that conveys not only its explicit types but also the implicit types derived from visitors' travel personalities. Introducing this sort of collective intelligence to topic-model-based representations adds a significant amount of information, which modifies their original probability distribution over topics. As shown in our experimental results, the newly discovered representation may capture dense and more realistic descriptions of attractions and thus play a significant role in improving travel recommendation quality by recommending relevant attractions with some level of diversity. Our method can be used to represent attractions in a variety of recommender systems, such as context-aware, POIs, travel destination, and hybrid recommender systems, to improve their quality.

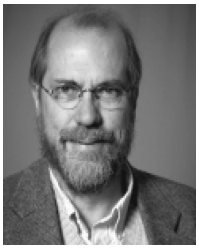
REFERENCES

[1] N. T. Thai and U. Yuksel, "Too many destinations to visit: Tourists' dilemma?" *Ann. Tourism Res.*, vol. 62, pp. 38–53, Jan. 2017.
[2] A. Felfernig, S. Gordea, D. Jannach, E. Teppan, and M. Zanker, "A short survey of recommendation technologies in travel and tourism," *OEGAI J.*, vol. 25, no. 7, pp. 17–22, 2007.
[3] Z. Y. Jia, W. Gao, and Y. J. Shi, "An agent framework of tourism recommender system," in *Proc. MATEC Web Conf.*, vol. 44, 2016, p. 01005.
[4] J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, "A hybrid recommendation approach for a tourism system," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3532–3550, 2013.
[5] J. Neidhardt, R. Schuster, L. Seyfang, and H. Werthner, "Eliciting the users' unknown preferences," in *Proc. 8th ACM Conf. Recommender Syst. (RecSys)*, 2014, pp. 309–312.
[6] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl. Syst.*, vol. 46, pp. 109–132, Jul. 2013.
[7] F. Ricci, "Travel recommender systems," *IEEE Intell. Syst.*, vol. 17, no. 6, pp. 55–57, Nov. 2002.
[8] M. E. B. H. Kbaier, H. Masri, and S. Krichen, "A personalized hybrid tourism recommender system," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2017, pp. 244–250.
[9] U. Gretzel, N. Mitsche, Y.-H. Hwang, and D. R. Fesenmaier, "Tell me who you are and I will tell you where to go: Use of travel personalities in destination recommendation systems," *Inf. Technol. Tourism*, vol. 7, no. 1, pp. 3–12, Jan. 2004.
[10] S. Park, I. P. Tussyadiah, J. A. Mazanec, and D. R. Fesenmaier, "Travel personae of American pleasure travelers: A network analysis," *J. Travel Tourism Marketing*, vol. 27, no. 8, pp. 797–811, Nov. 2010.
[11] B. Petrevska and S. Koceski, "Tourism recommendation system: Empirical investigation," *Revista Turism-Studii Cercetari Turism*, vol. 14, pp. 11–18, Dec. 2012.
[12] Y.-L. Zhao, L. Nie, X. Wang, and T.-S. Chua, "Personalized recommendations of locally interesting venues to tourists via cross-region community matching," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–26, Oct. 2014.
[13] X. Shao, G. Tang, and B.-K. Bao, "Personalized travel recommendation based on sentiment-aware multimodal topic model," *IEEE Access*, vol. 7, pp. 113043–113052, 2019.
[14] J. Shen, C. Deng, and X. Gao, "Attraction recommendation: Towards personalized tourism via collective intelligence," *Neurocomputing*, vol. 173, pp. 789–798, Jan. 2016.
[15] W. Min, B.-K. Bao, C. Xu, and M. S. Hossain, "Cross-platform multimodal topic modeling for personalized inter-platform recommendation," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1787–1801, Oct. 2015.
[16] T. Ameen, L. Chen, Z. Xu, D. Lyu, and H. Shi, "A convolutional neural network and matrix factorization-based travel location recommendation method using community-contributed geotagged photos," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 8, p. 464, Jul. 2020.
[17] C. Binucci, F. De Luca, E. Di Giacomo, G. Liotta, and F. Montecchiani, "Designing the content analyzer of a travel recommender system," *Expert Syst. Appl.*, vol. 87, pp. 199–208, Nov. 2017.
[18] E. Pariser, *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*. New York, NY, USA: Penguin, 2011.
[19] M. Ye, P. Yin, and W.-C. Lee, "Location recommendation for location-based social networks," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 458–461.
[20] Y. Wang, S. C.-F. Chan, and G. Ngai, "Applicability of demographic recommender system to tourist attractions: A case study on trip advisor," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Washington, DC, USA, Dec. 2012, pp. 97–101.
[21] M. Braunhofer, M. Elahi, and F. Ricci, "User personality and the new user problem in a context-aware point of interest recommender system," in *Information and Communication Technologies in Tourism*. Cham, Switzerland: Springer, 2015, pp. 537–549.

[22] R. Logesh and V. Subramaniyaswamy, "Exploring hybrid recommender systems for personalized travel applications," in *Cognitive Informatics and Soft Computing*. Singapore: Springer, 2019, pp. 535–544.
[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[24] C. Comito, "NexT: A framework for next-place prediction on location based social networks," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106205.
[25] C. Ma, Y. Zhang, Q. Wang, and X. Liu, "Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 697–706.
[26] Y. Si, F. Zhang, and W. Liu, "An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features," *Knowl.-Based Syst.*, vol. 163, pp. 267–282, Jan. 2019.
[27] O. Daramola, M. O. Adigun, C. K. Ayo, and O. O. Olugbara, "Improving the dependability of destination recommendations using information on social aspects," *Tourismos, Int. Multidisciplinary J. Tourism*, vol. 5, no. 1, pp. 13–34, 2010.
[28] M. Lohmann and H. Beer, "Fundamentals of tourism: What makes a person a potential tourist and a region a potential tourism destination?" *Poznan Univ. Econ. Rev.*, vol. 13, no. 4, pp. 83–97, 2013.
[29] T. Mahmood and F. Ricci, "Improving recommender systems with adaptive conversational strategies," in *Proc. 20th ACM Conf. Hypertext hypermedia (HT)*, 2009, pp. 73–82.
[30] A. Livne, M. Unger, B. Shapira, and L. Rokach, "Deep context-aware recommender system utilizing sequential latent context," 2019, *arXiv:1909.03999*.
[31] Z. Fu, L. Yu, and X. Niu, "TRACE: Travel reinforcement recommendation based on location-aware context extraction," *ACM Trans. Knowl. Discovery Data*, vol. 16, no. 4, pp. 1–22, 2022.
[32] S. C. Plog, "Why destination areas rise and fall in popularity," *Cornell Hotel Restaurant Admin. Quart.*, vol. 14, no. 4, pp. 55–58, Feb. 1974.
[33] Y. Huang and L. Bian, "A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the internet," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 933–943, Jan. 2009.
[34] E. Pantano, C.-V. Priporas, and N. Stylos, "'You will like it!' Using open data to predict tourists' response to a tourist attraction," *Tourism Manage.*, vol. 60, pp. 430–438, Jun. 2017.
[35] J. Neidhardt and H. Werthner, "Travellers and their joint characteristics within the seven-factor model," in *Information and Communication Technologies in Tourism*. Cham, Switzerland: Springer, 2017, pp. 503–515.
[36] J. P. Oliver and S. Srivastava, "The big five trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of Personality: Theory and Research*. New York, NY, USA: Guilford Press, 1999, pp. 102–138.
[37] C. Huang, Q. Wang, D. Yang, and F. Xu, "Topic mining of tourist attractions based on a seasonal context aware LDA model," *Intell. Data Anal.*, vol. 22, no. 2, pp. 383–405, 2018.
[38] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, "Topic modeling in embedding spaces," *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 439–453, Dec. 2020.
[39] D. Angelov, "Top2Vec: Distributed representations of topics," 2020, *arXiv:2008.09470*.
[40] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[41] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.
[42] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 100–108.
[43] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 262–272.
[44] N. Aletras and M. Stevenson, "Evaluating topic coherence using distributional semantics," in *Proc. 10th Int. Conf. Comput. Semantics (IWCS)*, 2013, pp. 13–22.
[45] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proc. 8th ACM Int. Conf. Web Search Data Mining*, Feb. 2015, pp. 399–408.
[46] C. C. Aggarwal, "Content-based recommender systems," in *Recommender Systems: The Textbook*. Cham, Switzerland: Springer, 2016, pp. 139–166.
[47] D. Kluver, M. D. Ekstrand, and J. A. Konstan, "Rating-based collaborative filtering: Algorithms and evaluation," in *Social Information Access*. Cham, Switzerland: Springer, 2018, pp. 344–390.
[48] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1974.
[49] R. N. Nandi, M. M. A. Zaman, T. A. Muntasir, S. H. Sumit, T. Sourov, and M. J.-U. Rahman, "Bangla news recommendation using doc2vec," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–5.
[50] D. Cer, Y. Yang, S.-Y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil, "Universal sentence encoder," 2018, *arXiv:1803.11175*.
[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[52] C. Hur, C. Hyun, and H. Park, "Automatic image recommendation for economic topics using visual and semantic information," in *Proc. IEEE 14th Int. Conf. Semantic Comput. (ICSC)*, Feb. 2020, pp. 182–184.
[53] V. Gupta and S. Kumar, "Songs recommendation using context-based semantic similarity between lyrics," in *Proc. IEEE India Council Int. Subsections Conf. (INDISCON)*, Aug. 2021, pp. 1–6.
[54] M. Gawinecki, W. Szmyd, U. Z. Uchowicz, and M. Walas, "What makes a good movie recommendation? Feature selection for content-based filtering," in *Proc. Int. Conf. Similarity Search Appl.*, 2021, pp. 280–294.
[55] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.
[56] *Three Sisters Springs*. Accessed: Mar. 22, 2022. [Online]. Available: https://www.threesistersspringsvisitor.org/sisters

TURKI ALENEZI received the B.A. degree in computer science from Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, in 2012, and the M.S. degree in information science from the University of Pittsburgh, in 2016, where he is currently pursuing the Ph.D. degree.

He was a Runner-Up at the Second Research Conference for IMSIU's Students (Progressive Exams Planning Coordination System). He was a Lecturer at Prince Sattam Bin Abdulaziz University, from 2016 to 2017. His main research interests include machine learning, recommendation systems, geographic information systems, tourism information systems, and human–computer interaction.



STEPHEN HIRTLE received the B.A. degree in mathematics and psychology from the Grinnell College, in 1976, and the Ph.D. degree in mathematical psychology from the University of Michigan, in 1982.

He is a Professor with the Department of Psychology and the Intelligent Systems Program, School of Information Sciences, University of Pittsburgh. He is the Director of the Spatial Information Research Group, University of Pittsburgh, where he conducts research on the structure of cognitive maps, navigation in real and virtual spaces, information visualization, and computational models for spatial cognition. He was the Founding Co-Editor of Spatial Cognition and Computation and the President of the Classification Society of North America. His research interests include spatial information theory with a focus on understanding how spatial concepts are represented, accessed, and utilized in a variety of spatial tasks, such as wayfinding. He hosted the Third International Conference on Spatial Information Theory (COSIT'97). He also has served on the Board of the University Consortium for Geographic Information Science and numerous review panels for the National Science Foundation, the National Institutes of Health, and the FWF, Austria.