



# A hybrid recommender system using topic modeling and prefixspan algorithm in social media

Ali Akbar Noorian Avval<sup>1</sup> · Ali Harounabadi<sup>1</sup>

Received: 13 September 2022 / Accepted: 26 December 2022 / Published online: 17 January 2023  
© The Author(s) 2023

## Abstract

Route schema is difficult to plan for tourists, because they demand to pick points of interest (POI) in unknown areas that align with their preferences and limitations. This research proposes a novel personalized method for POI route recommendation that employs contextual data. The proposed approach enhances the existing methods by considering user preferences and multifaceted tourism contexts. Due to the sparsity of the data, the proposed method employs two-level clustering (DBSCAN based on the Manhattan distance) that reduces the time to discover POI. In specific, this approach utilizes the following: first, a topic pattern model is employed to discover the users' attraction diffusion while improving the user–user similarity model using a novel asymmetric schema. Second, it has used explicit demographic information to alleviate the cold start issue, and third, it proposes a new strategy for assessing user preferences and also combined the context parameters in the form of a vector model with the Term Frequency Inverse Document Frequency technique to find contexts' similarity. Furthermore, our framework discovers a list of optimal candidate trips by involving personalized POIs in sequential patterns' mining (SPM); also, it used an adjusted forgotten function to involve the date context of each trip. Based on two datasets (Flickr and Gowalla), our methodology beats other prior approaches in F-score, RMSE, MAP, and NDCG factors in the experimental evaluation.

**Keywords** User interests · POI route recommendation · Geo-tagged photos · Topic · Personalization · POI contexts

## Introduction

In general, *Recommender Systems (RSs)* assist users in discovering the content, products, or services they need from a large amount of information on the web [1]. The tourism industry, which attempts to deliver personalized user experience and context, is one of the most prevalent implementations of RS [2, 3]. There has been an increase in the number of articles utilizing Location-Based Social Networks (LBSN) and spatial–temporal information in tourist RS during the last several years [1, 4]. Existing tourism recommender systems, on the other hand, have some drawbacks, mostly due to dynamic changes in tourists' travel habits, making the design of recommender systems for tourism purposes difficult and complex.

Conventional Collaborative Filtering (CF) methods provide suggestions relying on the travel habits of users who are acted in the same way as targeted tourists. In real-world applications, however, user similarity may differ, implying that most contemporary symmetric techniques yield lower precise findings [5–7]. Context-aware (CA) recommender systems, on the other hand, take into account the users' context and provide more accurate suggestion outputs, given that numerous tools now gather information on the users' state [8–10].

Traditional RSs may not apply to comparable individuals due to a lack of data on new users. The cold start issue is one of the most serious problems with recommender systems [11, 12]. On the other hand, the sparsity issue occurs when the number of user rankings is significantly fewer than the number of things; consequently, the RS is weak to predict significant evaluations, and the conventional techniques may result in bad suggestions [13, 14]. Given the significantly increasing number of tourist images shared on social networks, the data with them can be used to develop RSs. Geo-tagged photos may be utilized to detect real-life user

---

✉ Ali Harounabadi  
A.Harounabadi@iauctb.ac.ir

Ali Akbar Noorian Avval  
Ali.Noorianavval.Eng@iauctb.ac.ir

<sup>1</sup> Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

journey histories as a representative way of tourist visits [15, 16].

For travelers who are unknown of the diverse number of places in a new area, the planning procedure for a tailored trip can be time-consuming, because selecting points of interest (POIs) and organizing them might be difficult [17, 18]. In other words, visitors prefer having a journey by pre-arranged POIs to receiving a list of suggestions POIs; therefore, an RS that can create pre-arranged POIs is more useful for the tourist [19]. The sequence movement pattern in the POI suggestions is one of the tourist behaviors linked with the visited POIs over a specific period, and approaches like sequential pattern mining (SPM) can be used to evaluate user behavior over time [20].

Because recommender systems are complex, hybrid techniques may be utilized to improve performance, especially with the rise of social networking. Recent applications of hybrid techniques to recommender systems [21–23] have utilized strategies in various ways to benefit from complementary advantages; however, the need for more complex hybrid methodologies and data fusion remains significant. In this paper, a novel hybrid method for accumulating personalized recommendations from multiple recommendation systems is developed to predict convincing POI recommendations. Moreover, an offset of approach novelty in recommendation results is achieved by adjusting hybrid RS parameters, allowing the merging and ranking of each approach. In addition, a fusion technique is employed to combine demographic and contextual data and then generate a list of perfect-matching POIs based on the user's interests and preferences.

The current research proposes a novel hybrid technique for improving model performance and overcoming the drawbacks described above. To overcome the aforementioned shortcomings of previous techniques, our framework applied a hybrid method to CF, demographic-based (DB), and topics. This work offers acceptable tourist places and sequences based on changing tourist preferences over time. The study also manages the issue of cold starts using users' demographic information. Asymmetric schema is also utilized to solve symmetric user similarity issues and increase algorithm performance.

This study serves as an essential supplement to our previous paper [24]. This research improved the previous method by incorporating additional key factors about tourists. Also, our new framework utilizes the Topic pattern to get the subject distribution of the route records of tourists and produce tourist similarity relying on the subject distribution. Compared to the previous method, our asymmetric scheme and user preference equation are significantly improved in this study. Additionally, we employ the Manhattan formula and utilize a specific equation for user age when determining demographic similarity.

The essential contributions of this research are resumed as follows:

- Proposing a personalized POI route framework based on context and explicit demographic data using an asymmetric topic model.
- Utilizing the term frequency inverse document frequency technique to calculate the similarity between contextual factors.
- Representing a novel and improved CF method using the Markov function and user preferences to find the preferences.

The remainder of this research is organized as follows: the “[Background knowledge](#)” reviews the literature relevant to our research. The problem formulation and our proposed method, called TopicSeqHybrid are introduced in the “[The proposed method](#)”. The “[Simulations and experimental evaluation](#)” further discusses the proposed methodology and its testing and evaluation results in comparison with the existing papers. Finally, the “[Conclusions and future work](#)” provides concluding comments.

## Background knowledge

This section studies the methods available in this field. The attention is divided into three groups.

### Approaches of topic model-based recommendation

Several recently developed algorithms provide trip suggestions based on various data sources, namely geo-tagged photographs, blogs, and GPS trajectories. Specifically, collaborative filtering systems provide effective travel suggestion performance [25, 26]. Although CF-based recommendation systems produce encouraging results, they are plagued by the “data sparsity” issue. In this respect, topic model-based algorithms that permit effective tailored trip suggestions were developed to combat this issue [27, 28]. Topic models are characterized as probabilistic hierarchical models where a user is represented as a mix of themes, and the topic is represented as probability distributions over points of interest. Topic models are employed in various applications, including data retrieval and user interest modeling [25, 29]. In this context, several topic analysis approaches have been developed, including LSA (Latent Semantic Analysis), LDA (Latent Dirichlet Allocation), and PLSA (Probabilistic Latent Semantic Analysis). PLSA models primarily employ the Expectation–Maximization (EM) technique, which consists of two iterative maximization (M) and expectation (E) processes [30, 31].

Proposing a tourist RS subject pattern, the authors of [28] investigated the exploitation of trip data by constructing a User-Reign-Season Topic model. In [27], a recommender system was proposed that provided venues within a specified geospatial range. Through an offline system, they represented each individual's particular preferences utilizing a weighted category hierarchy and an iterative learning method LDA, the simplest model, is currently employed in various applications. Research by Chen et al. presented a system that directly learned patterns through photo information and utilized the Markov method to accurately assess places for unique orders of uses photographs [26]. In another study, Chen et al. evaluated photographic traces [32]. In a further study, Kurashima et al. utilized a combination of a topic and the Markov model to propose trips according to user interests and commonly frequented routes [33, 34]. Accordingly, Jiang et al. proposed Author Topic Model CF by concurrently mining travel subject categories and user topical interest [2]. Using geo-tagged Flickr images, Yin et al. analyzed the distributions of several geographical categories, including the coast, sunset, and hiking [35].

Pozdnoukhov and Kaiser examined the spatial–temporal context distribution of the thematic material by analyzing a large sample of geo-tagged tweets [36]. In another study, Zhao et al. introduced a probabilistic topic model that extracted themes from travelogues and featured places with relevant topics for the place's suggestion and summary [37]. The primary challenges of travelogues-based methods are (1) identifying whether bloggers have visited the destinations is challenging, and (2) pinpointing the exact location of travelogues is difficult, since they are typically unstructured and include noisy information. Furthermore, Jiang et al. developed a tailored vacation sequence recommendation by combining travelogues, community-contributed images, and the heterogeneous information associated with geo-tagged photos [38]. Comprising representative tags, cost distributions, visiting time distributions, and visiting season distributions for each topic are mined using topical package space to close gaps between travel routes and preferences.

Sun and Lee developed a framework for proposing top-k tours to users based on their interests and available time utilizing user-generated content from a social network for photo sharing [39]. The LDA model was employed to classify user-posted hashtags into landmark-related themes, then used these subjects to profile landmarks and users, and proposed tours to users. Their experimental findings demonstrated that their technique was superior to the Markov-Topic method concerning average score and precision. Ren et al. examined context-aware probabilistic matrix factorization modeling for recommending POIs [40]. They utilized and incorporated LDA-based topic models to compute POI ratings. A paper by Kanimozhi et al. suggested a tailored tourist-based recommendation method based on certain restrictions, time

and cost limitations, LDA-based topic modeling, and Jaccard measure [41].

## Approaches of context-aware recommendation

Across several disciplines, over 150 different meanings of the term "context" have been presented [42]. One of the best meanings of the keyword "context" is as follows: context [43] refers to all information used to define a being's condition. Context-aware recommender systems try to improve the quality of suggestions by making contextual information more accessible. The context is incorporated as a component. The three forms of context-aware recommender systems are contextual pre-filtering, contextual modeling, and contextual post-filtering [44].

A survey of pertinent papers was conducted to that goal. Memon et al. [45] developed a recommender system that produced suggestions depending on the situation. Pearson's similarity metric was used to estimate the degree of similarity between users after pre-filtering the locations of the relevant town using context data like climate and time. When collaborative filtering was employed entirely, the findings revealed that the suggestions were more precise. Sun et al. [17] illustrated a CA that generated participating recommendations tailored to the users' preferences using contextual information and geo-tagged images. When used in cold start status, this method was vastly more effective.

Our prior paper [46] described a tourist RS that considers the target phone's position and suggests that the best user accommodations rely on contextual data and trust measures. Studies and assessments revealed that introducing extra context improved the suggested method's results.

## Approaches of sequential patterns' mining recommendation

The sequential patterns' mining algorithm is an impressive approach to generating personal travel routes in route recommendation approaches. Sequential pattern mining is the process of identifying common subsets as patterns in a sequence database. Each sequence-aware RS may be categorized into four types based on application scenarios: adaptation, trend detection, repeated recommendation, and sequential patterns [38].

In the sequence-aware recommendation schema, the SPM method is valuable for creating travel routes. Because this would precisely propose the following POI to visit at the subsequent timestamp, sequential POI recommendation is nearly tenfold more difficult than conventional POI recommendation. Data mining requires the discovery of useful patterns in datasets. SPM is a favored sequence data mining method, a subsection of data mining strategies focused on discovering patterns in sequence data that may be used in several

disciplines. Some of the algorithms proposed for identifying sequential patterns are GSP, Free-Span, Prefix-Span, and SPADE [47, 48].

In [49], they represented a trip suggestion strategy as a solution to the orienteering challenge, for which researchers expressed their trip recommendation issue regarding the person's journey limits, like time limitations and the necessity that the trip begins and concludes at certain places of interest. Nevertheless, numerous essential POI-related tourist variables, such as the favored trip period and travel categories, were left out of their research. Their strategy takes into account both POI popularity and user interests when proposing acceptable POIs to visit and the amount of time to spend at each POI. They also developed a model for automatically detecting real trip sequences and estimating POI popularity and user interest utilizing geo-tagged images.

Other research [19] revealed a POI trip recommender structure that aided many tourist data sources, as well as SPM, in which they employed an organized way to develop a POI foundation of knowledge and a huge framework of POI patterns.

Table 1 provides an overview of tourist recommender system methodologies, including paradigms, and remarks.

Contrary to previous methods, contextual, geo-tagged, and tourist demographic information has been integrated to create structured POIs visit sequences. Regarding that, the Prefix-Span technique efficiently extracts POI travels visiting patterns despite taking into account various tourism contexts. Eventually, a trip retrieval method is used to develop POI trip recommendations again by taking into account the topic model and various user contexts. As a consequence, the final POI journey can incorporate traveler's limitations while also guaranteeing the route has a reasonably high user issue.

## The proposed method

Our method introduces a hybrid trip recommender system for the tourist industry that utilizes tourist demographic, contextual, and geo-tagged information to suggest a list of places in a town. This framework is separated into two phases, as illustrated in Fig. 1 (offline and online). Some calculations, including historical data preprocessing, user similarity, and discovering area of interest (AOI) and POI with clustering, were performed offline to improve speed, and the resulting data were stored for use in the online phase. The offline processes' phase can be executed on the server side or in the cloud, depending on the system's configuration. It is evident that online phase processes could be executed on mobile devices as an App.

After preprocessing the dataset offline (“Data preprocessing”), contextual data are derived from the photo time stamp and weather service. These data are added to the dataset to

make it more complete (“Enriching geo-tagged photo with contextual information”). Moreover, clustering methods are being utilized to establish the AOI. Each AOI containing one or more POI relies on the geographic coordinates of the images.

Regarding that, POIs are found by applying again the clustering technique on the AOIs' results (“Finding POIs”). Then, a profile is created for each of these POIs (L) by calculating their publicity and situational characteristics (“Producing the Profile of Points of Interests (POIs)”). After that, based on the Topic model and prior visits to each POI, the user–POI similarity with the weighted graph (“User-POI detection”) and user–user similarity with topic modeling (“Topic-based calculation of user–user asymmetric schema”) were computed by a subject distribution of tourist trip histories. These two measures stored data in their respective databases. To find prior user journeys, POI sequences were created using the user's POI visit time (“Sequence extraction”). The Prefix-Span technique was then utilized to extract suitable POIs' sequences (P set) (“SPM algorithm”). The POI sequence database was also used to hold these POI sequences. The results were cached for use during the online phase, which speeds up computation and improves system responsiveness.

Tourists can register themselves for the framework during the online phase by supplying their demographic data their age, sex, town, nation, relationship, and profession. When a user makes a query to the system, the query is enriched with the user's current contextual information, such as geographical coordinates (which specify the current user's city) and weather information for the current user's intended travel dates (“Enriching user queries by contextual data”). Next, contextual pre-filtering is also utilized to choose POIs in that place (L') based on the user's actual place (“Pre-filtering based on context”).

The next stage computes the hybrid similarity between the current tourist and those who visited the filtered POIs (L') (“Combination of the recommendations”). Both the CF and DB approaches were used in our similarity. Tourists who are more similar to the current tourist are then selected. POIs can currently be picked and suggested to rely on the top-ranking neighbor's tourists. Before actually making suggestions, the user's spatial proximity and present climate to the places are taken into account (contextual modeling). Then, the predicted list of POIs has been selected (top-N ranked POIs) (“Recommendation”).

On the other hand, for the candidate travel patterns (“Candidate trip pattern stage”), we take into account the derived top-N POIs (“Recommendation”) and the explored and saved sequential trip patterns (P set) (“SPM algorithm”). For each tourist, the trip pattern score is computed by the amount of the ranks of the POIs stored within the trip pattern using the rank function mentioned in “Candidate trip pattern stage”.

**Table 1** A survey of relevant papers

Year	Ref. no.	Recommender system paradigms							Advantages/potential limitations	
		CF	CA	DB	Personalized	Sequence	Topic	Descriptions		Dataset
2014	[27]	✓	✓				✓	A context-aware hierarchical Bayesian method Using topic regression with social matrix factorization	Epinions	Using spectral clustering for user-item Reduced complexity Not applicable for cold start No asymmetric similarity measured No personalized
2015	[45]	✓	✓					Used user preferences Employing contextual pre-filtering	Flickr	Improved accuracy Reduced complexity Small number of contextual parameters No asymmetric measure
2015	[50]	✓	✓			✓		Using geo-tagged photos Considering user interest and sequences	Flickr	Considering probability No personalized recommendation No demographic information
2015	[2]	✓			✓			Using the topics about user preference Using geo-tagged photos	Flickr	Dealing with data sparsity Consider user preference Not applicable for cold start
2015	[43]	✓	✓		✓			Based on the topic distribution of his travel histories Considering season and weather as context information	Flickr	Using geo-tagged photos Utilizing a topic model to mine the preference of a user Not applicable for cold start and sparsity problems

Table 1 (continued)

Year	Ref. no.	Recommender system paradigms							Advantages/potential limitations	
		CF	CA	DB	Personalized	Sequence	Topic	Descriptions		Dataset
2016	[38]	✓	✓	✓	✓	✓		Using travelogues and community-contributed photos and the heterogeneous metadata (geo-location, and date taken)	Flickr and IgoUgo	Recommend a travel sequence Considering tags Not applicable for topic and cold start
2017	[40]		✓				✓	Context-aware probabilistic matrix factorization Modeling the social correlations Aggregated (LDA) model	Twitter and Foursquare	Modeling the topic model Using the textual, geographical, social, categorical, and popular information Dealing with data sparsity Not applicable for cold start
2017	[51]	✓	✓		✓			A hybrid context-aware recommender system with CF and GSP Personalized	LMS	Including context pre-filtering Alleviate cold start and sparsity problems
2017	[52]	✓	✓	✓	✓	✓		Location-based recommender system Preferred time-aware route planning	Gowalla and Foursquare	Considering the geographical, social, and temporal information of users Including user preferences Not personalized

**Table 1** (continued)

Year	Ref. no.	Recommender system paradigms							Advantages/potential limitations	
		CF	CA	DB	Personalized	Sequence	Topic	Descriptions		Dataset
2017	[39]	✓	✓			✓	✓	Considering the user's interest and time frame Using Markov-topic	Flickr	Using clustering Employing the LDA Not applicable for cold start
2018	[53]	✓	✓			✓		Using sequence patterns Based on the semantic model Context-based (time and spatial data)	Flickr	Utilizing user preferences Based on geo-tagged photos No clustering Not applicable for cold start
2018	[18]				✓		✓	Personalized top-n sequential recommendation Using convolutional embedding recommendation	Gowalla Foursquare Tmall MovieLens	Dealing with data cold start Using user preferences and sequential patterns Not applicable for sparsity issue
2018	[49]	✓	✓		✓		✓	A trip recommendation for tourists Considering the interests of users and sequences	Flickr	Personalized Not applicable for cold start No demographic information

Table 1 (continued)

Year	Ref. no.	Recommender system paradigms					Topic	Descriptions	Dataset	Advantages/potential limitations
		CF	CA	DB	Personalized	Sequence				
2018	[54]	✓	✓	✓	✓	✓	A personalized an itinerary recommendation with time constraints LBSN—exploiting geographical features and social relationships	Gowalla	User-based CF with time preference Considering the visiting time of locations Not applicable for cold start	
2019	[55]	✓	✓	✓	✓	✓	Hybrid location-based travel recommender system Personalized travel recommendations	Trip advisor	Using swarm intelligence algorithms No asymmetric similarity measure	
2019	[32]	✓	✓	✓	✓	✓	Personalized itinerary recommendation Integrates POI textual contents, historical user, and POI categories	Flickr	Context-based Alleviates the cold start No asymmetric similarity measure	
2019	[19]	✓	✓	✓	✓	✓	A POI route recommender framework Using sequential pattern mining	Flickr	Dealing with data cold start and sparsity Generating different fine-grained candidate POI routes	
2019	[56]	✓	✓	✓	✓	✓	Personalized location recommendation Using Matrix factorization	Nokia mobile data	Using demographic features Personalized Dealing with data cold start	



**Table 1** (continued)

Year	Ref. no.	Recommender system paradigms						Dataset	Advantages/potential limitations
		CF	CA	DB	Personalized	Sequence	Topic		
2020	[4]	✓	✓		✓			Trip advisor	Using users' reviews on social networks to discover user's preferences Not applicable for cold start
2020	[57]	✓	✓	✓				Trip advisor	Using user preferences Dynamic contextual information Not applicable for cold start
2020	[1]	✓	✓					Trip advisor	Employing graph clustering Not applicable for cold start
2020	[58]		✓			✓		Movielens	Employing the probability density distribution Not applicable for cold start
2020	[59]				✓			Flickr	Personalized Alleviates the cold start and data sparsity conditions No demographic information No asymmetric similarity measure

Table 1 (continued)

Year	Ref. no.	Recommender system paradigms							Advantages/potential limitations	
		CF	CA	DB	Personalized	Sequence	Topic	Descriptions		Dataset
2020	[13]	✓	✓	✓	✓	✓		Tourist recommendations based on demographic and context-aware Personalized	Flickr	Using asymmetric similarity measure Consider limited context parameters
2020	[15]	✓	✓	✓	✓	✓		Personalized travel recommendation based on geo-tagged photos Using matrix factorization	Flickr	Using contextual information, text tags Not applicable for cold start
2021	[60]		✓		✓	✓		Personalized sequential pattern Considering decision tree Multi-label classification	CDNow-RFM and msdt2multi-valued	Using sequential pattern mining Deal with the cold start Not consider contextual data Not personalized
2022	[44]		✓		✓	✓		A personalized POI recommender system Using heterogeneous graphs	Trip advisor	Considering POI categories and periods No demographic information
2022	[24]	✓	✓	✓	✓	✓		A recommender system using CF and sequence patterns mining Based on geo-tagged photos	Flickr and STS	Using Demographic Data Using asymmetric similarity Deal with the cold start
TopicSeqHybrid		✓	✓	✓	✓	✓		Our framework		

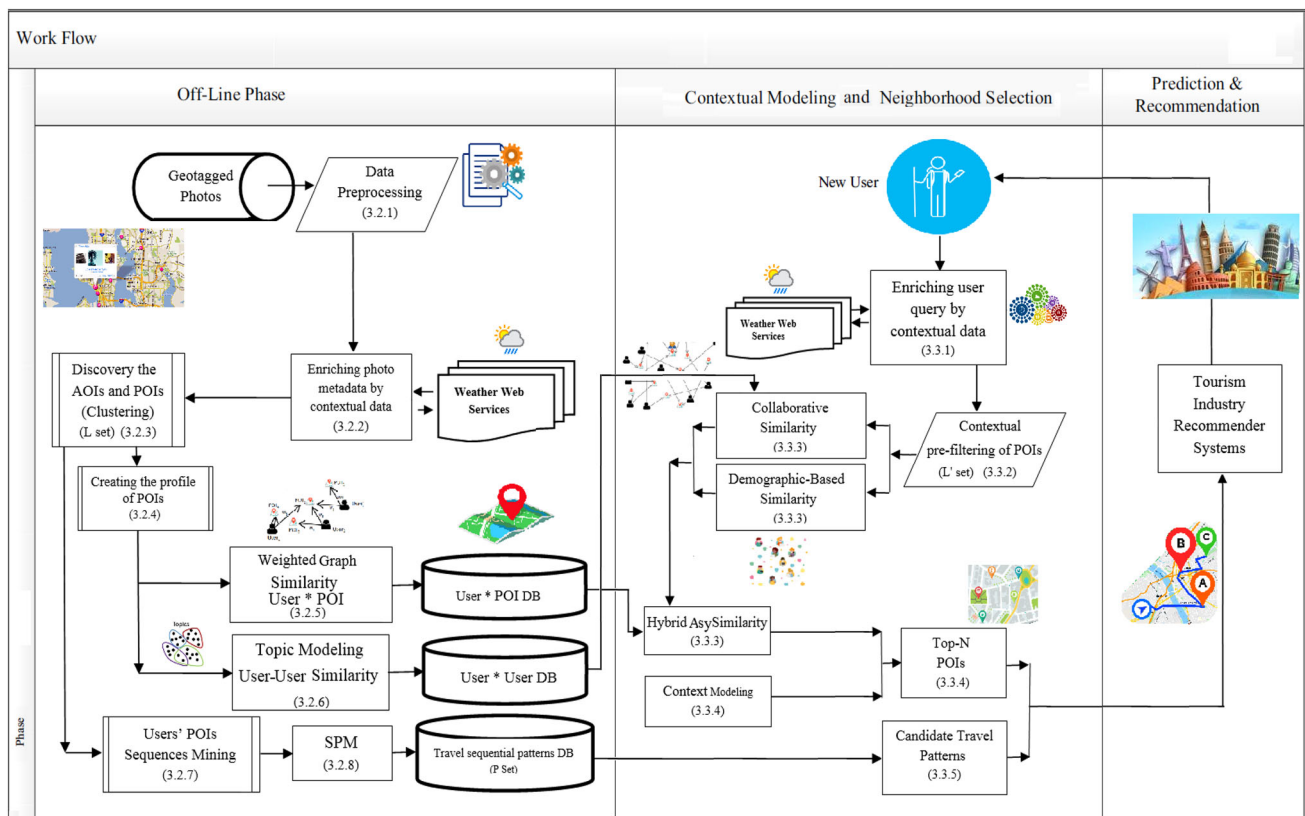


Fig. 1 The process of our method

The top-N trip sequential patterns are applicant journey patterns. Finally, the target tourist is recommended with travel suggestions based on current contexts, demographic factors, and sequential patterns of movement.

The steps that follow go over the contribution of the proposed framework as well as the algorithms that were used to create them.

### Problem identification

The following is a definition of the challenge of proposing exciting tourist destinations and POI sequences in geo-tagged social networking sites:  $P = P_1, \dots, P_n$  is a series of publicly accessible geo-tagged photographs that demonstrate the approach of locating tourist destinations in a city, assessing their attractiveness, and providing intriguing journey sequence suggestions based on prior tourist journeys and my travel sequence patterns. Interestingly, travelers’ publicity photo collections are being utilized to offer exciting tourist places and intriguing tourism sequences based on the visitors’ present context.

### Offline phase

To enhance the speed of our framework, some calculations were conducted offline, and the data obtained were preserved.

### Data preprocessing

First, the data source has to be cleaned and preprocessed due to the inclusion of some unclear and unsuitable data. This included deleting unclear information and images with insufficient parameters. It is important to mention that while visiting a POI, a person can photograph it many times. As long as the time change between a person’s initial and subsequent images is smaller than a threshold, both photos are considered one and pertain to the same visited place.

### Enriching geo-tagged photo with contextual information

The contextual data for the dataset’s image items were created and stored in this step. These data, including the time and geographic place, are paired with each image posted by users. In line with the map in Table 2, the current climate is derived using the climate application. Contextual information, such as climate, temp, season, and other date information, is included in the database.

**Table 2** Context matching

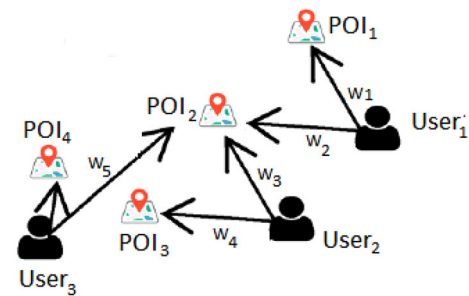
Time context	Visit day	Saturday, Sunday	Weekend	
		Monday, ..., Friday	Working day	
	Visit time	06:00–12:00	Morning	
		12:00–18:00	Afternoon	
18:00–06:00		Night		
Visit season	March, April, and May	Spring		
		June, July, and August	Summer	
	September, October, and November	Fall		
		December, January, and February	Winter	
	Weather context	Temperature	> 34 °C	Hot
			18–34 °C	Warm
< 18 °C	Cold			
Weather	Sunny, clear sky	Sunny		
	Cloudy, broken clouds, scattered clouds	Cloudy		
	Rain, fog	Rainy		
	Snow, snowfall	Snowy		

### Finding POIs

The DBSCAN approach was used to cluster geo-tagged pictures, and distance measures obtained from the Manhattan equation were used to extract spatial positions. This approach offers significant benefits over previous clustering algorithms, including the ability to discriminate clusters using arbitrary areas [61] and the demand for little scope information to determine the parameters. It also does a good job of grouping vast amounts of data. The thickness point for clusters in the DBSCAN is the same. Two factors are important: the minimum number of points needed to form a cluster (MinPts) and the radius (Eps). The size and density of clustered places can vary. AOIs were extracted from a batch of geo-tagged pictures using the DBSCAN clustering technique. After that, the algorithm was run again with the proper settings on the observed AOIs. As a consequence, a collection of POIs was created, as well as a database of important tourist places (L).

### Producing the profile of points of interests (POIs)

To evaluate the publicity and context features of each POI, Eqs. (1) and (2) were used to create a profile of the discovered POIs



**Fig. 2** The relation between Users and POIs

$$\text{Place Popularity (POI)} = \log\left(\frac{N}{N_l}\right). \tag{1}$$

Here,  $N_l$  represents the set of visits to a specific POI from a region, whereas  $N$  denotes the overall amount of visitors from that region.

A new weighting context vector structure is presented in our work as  $\vec{C}_l = \langle c_{(l,1)}, \dots, c_{(l,k)} \rangle$ , where  $c_{(l,j)}$  indicates the context ( $j$ ) of any POI, and ( $n$ ) represents the total number of contextual factors, as shown in Table 2.  $c_{(POI,j)}$  is calculated utilizing the TF-IDF algorithm formed on Eq. (2) [62]

$$c_{(POI,j)} = \text{TF}_{POI} * \text{IDF}_{POI} = \frac{w_{(POI,j)}}{w_{(0,j)}} * \log \frac{w_{(0,0)}}{w_{(POI,0)}}. \tag{2}$$

Here,  $w_{(POI,j)}$  shows the number of visitors from the POI in context ( $j$ ),  $w_{(0,j)}$  indicates the number of travels in context ( $j$ ) from all POIs in the present town,  $w_{(0,0)}$  represents the number of the journey in any context from all POIs in the present town, and  $w_{(POI,0)}$  indicates the number of the journey in each context from the POIs.

### User–POI detection

The similarity between User and POI is used to construct a weighted undirected graph  $\text{Graph}_{\text{User-POI}} = (\text{User}; \text{POI}; \text{Edge}_{\text{User-POI}}; \text{Weight}_{\text{User-POI}})$  to identify the preferences of a group of tourists  $U$  inside a collection of places  $L$ .  $\text{Edge}_{\text{User-POI}}$  and  $\text{Weight}_{\text{User-POI}}$  represents collections of sides and side weights among the User and POI, respectively, reflecting users’ visitations and the instances of visitations to a specific POI.

Figure 2 illustrates the similarity between users and POIs. Suppose  $n$  users and  $m$  POIs, an  $n$ -by- $m$  adjacency matrix  $\text{Matrix}_{\text{User-POI}} (\text{Matrix}_{\text{User-POI}} = [T_{ij}])$  is created for the network  $\text{Graph}_{\text{User-POI}}$ , where  $T_{ij}$  denotes the instances the  $j$ th POI has been visited by the  $i$ th user. If  $T_{ij} = 0$ , it indicates that the  $i$ th tourist has never seen the  $j$ th POI.

**Topic-based calculation of user–user asymmetric schema**

The similarity among prior POI visitors was computed and saved for use in online step activities. Other travellers’ experiences were utilized to produce suggestions for the present customer. In numerous applications, topic models were utilized effectively to represent user interests. Popular approaches for topic analysis are namely PLSA and LDA. As the estimation method for PLSA is superior to LDA [41], we utilize PLSA to determine the subjects of user journey histories. As a set of topics, a user can be denoted as a probability distribution throughout places. The possibility  $P(\text{POI}|h_u)$  that a user  $u$  (given POI history  $h_u$ ) visits a POI is computed using Eq. (3) in topic modeling

$$P(\text{POI}|h_u) = \sum_{t \in T} (P(t|h_u) * P(\text{POI}|t)), \tag{3}$$

where  $(t|h_u)$  and  $P(\text{POI}|t)$  denote the user  $u$ ’s probability of being interested in subject  $t$ , where POI is chosen from topic  $t$ . We use the EM technique to estimate the topic proportions  $P(t|h_u)$ .

Equation (4) is used in the E-step to compute latent topic posterior probability

$$P(t|h_u, \text{POI}) = (P(t|h_u) * P(\text{POI}|t)) / \left( \sum_{t' \in T} (P(t'|h_u) * P(\text{POI}|t')) \right). \tag{4}$$

Equations (5) and (6) are used in the M stage to bring to date the factors needed to optimize the probability

$$P(t|h_u) = \sum_{p \in \text{POI}} (N(\text{POI}, h_u) * P(t|\text{POI}, h_u)), \tag{5}$$

$$P(\text{POI}|t) = \sum_{u \in \text{User}} (N(\text{POI}, h_u) * P(t|\text{POI}, h_u)), \tag{6}$$

where  $N(\text{POI}, h_u)$  is the number of POIs occurring in history  $h_u$ , gained by retrieving the User–POI matrix  $\text{Matrix}_{\text{User-POI}}$ . Through repetition of the E stage and M stage in conjunction, the topic distribution  $P(t|h_u)$  for a journey history can be obtained, which can then be used to compute the likeness among tourists.

After retrieving the topic distribution of the user travel history, Eq. (7) is employed to compute the likeness among tourists and create the similarity matrix  $\text{Matrix}_{\text{User-User}}$ , used for tailored suggestions according to asymmetric collaborative filtering

$$\text{sim}_{CF}(u, v) = \frac{\sum_{i=1}^k f_u^i * f_v^i}{\left| \sqrt{\sum_{i=1}^k (f_u^i)^2} \right| * \left| \sqrt{\sum_{i=1}^k (f_v^i)^2} \right|}, \tag{7}$$

$f_u^i$  and  $f_v^i$  are the probability that users  $u$  and  $v$  would be attracted to subject  $(i)$ , respectively, and  $k$  denotes the number of topics in this equation.

In the actual world, user similarities are not always symmetrical and may not be identical. The similarity link between two users is valued equally in most standard similarity measurements. These techniques are founded on the premise that  $\text{sim}(u, v)$  equal  $\text{sim}(v, u)$ ; while, the impact of two different users on one another differs; therefore, asymmetric schema is given to traditional similarities in CF approaches to create a highly realistic similarity [7, 58]. This work uses asymmetric schema to bypass the limitation. The rate of similarity places among tourists, adjusted by the figures of places assessed by the present tourist, Eq. (8) is used to create the asymmetric similarity measure

$$\text{sim}_{\text{Asy-Measure}}(u, v) = \frac{1}{R + (|l_u \cup l_v| - |l_v|)} \tag{8}$$

Here,  $l_u$  represents the number of POIs visited by  $u$ . This equivalence looks at the proportion of common ratings that users have among all of their rated things, rather than the proportion of common ratings in the total number of ratings among tourists. As a result, Eq. (9) contains this parameter

$$\text{Sim}_{\text{AsyCF}}(u, v) = \text{Sim}_{\text{Asy-Measure}}(u, v) * \text{Sim}_{CF}(u, v). \tag{9}$$

We should also take into account the preferences of each visitor. Different tourists have different tastes. We utilize the median of the PlacePopularity to represent the user preference to display this conduct distinction. The following is a representation of the user PlacePopularityPreference (UPP) based on similarity metrics:

$$\text{Sim}_{\text{UPP}}(u, v) = \frac{e^{-(|r_{u,p} - r_{\text{med}}| * |r_{v,p} - r_{\text{med}}|)}}{[1 + e^{-(|r_{u,p} - r_{\text{med}}| * |r_{v,p} - r_{\text{med}}|)}]^2}, \tag{10}$$

where  $r_{u,p}$  indicates the rating of PlacePopularity by user  $u$ . The  $r_{\text{med}}$  represents the median value of two tourists,  $u$  and  $v$ , on a rating scale.

By integrating Eqs. (9) and (10), we may get at the new formalization, which we term enhanced new CF asymmetric similarity model (AsyNCF) (Eq. (11)). Hybrid RSs improve performance by integrating two or more recommendation methods. CF is frequently used in conjunction with another technique to prevent the ramp-up problem. We used the feature combination approach with multiplication, since this hybrid has two different recommendation components: contributor (in our study, UPP) and real recommender (in this study, asymmetric CF). In other words, the relationships between the product’s parts have been preserved. The genuine recommender operates on data that has been altered by

the contributor [63]

$$\text{Sim}_{\text{AsyNCF}}(u, v) = \text{Sim}_{\text{UPP}}(u, v) * \text{Sim}_{\text{AsyCF}}(u, v). \quad (11)$$

### Sequence extraction

This stage extracts the place sequences to determine the tourist travels relying on the place visit order. The period of every user’s trips to POIs is also taken into account. A single trip is formed when the time variation among two sequential POI visits is lower than a threshold level; as long as the time variation is higher than the threshold level, these distinct journeys are formed. We utilize an 8-h threshold in our strategy, as in earlier studies [49]. A factor is used to track the periodicity of each trip. The number of users that visited each trip is used to establish the sequencing frequency in this approach. Each journey has its collection of POIs, as well as its own set of POI orders.

### SPM algorithm

To assess the visitors’ sequential trip movement patterns, the Prefix-Span algorithm was employed for their journeys in our work. The sequential movement patterns of users give vital information for projecting further suggestions in the trip recommendation system, and this stage tries to build famous tourist journeys. Prefix-Span is an eminent approach for finding common item-set models in databases. The Prefix-Span method is a simple algorithm that explores the full collection of patterns [47, 64]. It is substantially quicker than both the GSP and FreeSpan techniques.

The phases of the Prefix-Span technique are as follows: calculating the support value for each trip, creating candidate sequences, and eliminating those that have a support value less than the Min-Support. The Prefix-Span method is applied to the POI sequences to the minimal support threshold, resulting in a database of sequential trip patterns.

### Online phase

The following steps are included in this stage. Our method answers the target user’s request quickly and interactively.

### Enriching user queries by contextual data

The system calculated the time requested by the user to visit as the tourist’s desired time during the online phase. The weather and temperature contexts for that location were then provided and finished according to mapping in Table 2 of the user context inquiry, utilizing the season and time of visit contexts taken from the weather web service. For the present

user, a context factors’ structure such as  $(\vec{V}_u)$  is built. When a context criterion is satisfied, it is given a value of one; or else, it is given a non-value.

### Pre-filtering based on context

The data for that city were chosen in this stage based on the current user’s geographical attributes in the enhanced query. This contextual pre-filtering creates the collection of those city locations ( $L'$ ).

### Combination of the recommendations

This phase uses Eq. (12) to compute the hybrid similarity. In terms of the present tourist and those who visited the set ( $L'$ ) of places, this equation highlights the similarities between CF and DB

$$\text{Sim}_{\text{AsyHybrid}}(u, v) = (1 - \beta) * \text{Sim}_{\text{DB}}(u, v) + (\beta)\text{Sim}_{\text{AsyNCF}}(u, v). \quad (12)$$

This compound is balanced using the coefficient ( $\beta$ ) to smooth out the linear connection [65].

Equation (13) was used to estimate the demographic similarities between the two tourists [63, 66]

$$\text{Sim}_{\text{DB}}(u, v) = \frac{|\text{num}_1(\vec{D}_u \cap \vec{D}_v)|}{|\text{Demographic feature vector}| * 1 / \left( 1 + \frac{|\text{age}_u - \text{age}_v|}{\max(\text{age}) - \min(\text{age})} \right)}. \quad (13)$$

For each user, a demographic characteristic (excluding age) vector such as  $(\vec{D}_u)$  is created. The first tourist demographic characteristic vector is compared to the second tourist demographic information vector when comparing users based on their demographic features. If two users have the same value for a certain property, such as sex, the value of one is utilized. Using the  $\text{num}_1(\vec{D}_u \cap \vec{D}_v)$  function, the number of units in the two users’ common factors vector is tallied and divided by the number of demographic factors examined by the users. The output result of this similarity is always between 0 and 1.

Given the importance we place on the aged character, we utilized the tourist age as a distance attribute model. Where  $\text{age}_u$  and  $\text{age}_v$  are the ages of the two tourists  $u$  and  $v$ .

Following that, utilizing the similarity metric presented in this equation, the present tourist’s similarity to other tourists visiting the aid area (User–User) is determined. These findings are used to choose people among those who have visited

that city who have a greater similarity score to the present user.

**Recommendation**

The level of the intention of the present tourist  $u$  to visit destinations can be determined using Eq. (14), based on the similarity among tourists

$$\text{Pred}(u, l) = \frac{\sum_{v \in U'} \text{Sim}_{\text{WT-context}}(C_u, C_l) * \text{Sim}_{\text{loc-context}}(l_u, l_l) * \text{Sim}_{\text{AsyHybrid}}(u, v) * (r_{v_l})}{\sum_{v \in U'} \text{Sim}_{\text{AsyHybrid}}(u, v)}, \tag{14}$$

where  $(r_{v_l})$  indicates the real rating of a tourist  $(v)$  for the place  $(l)$ . In this equation, when computing the place  $(l)$  score relies on the tourist visit,  $\text{Sim}_{\text{WT-context}}(C_u, C_l)$  and  $\text{Sim}_{\text{loc-context}}(l_u, l_l)$  are used as a weight.

The context factor  $\text{Sim}_{\text{loc-context}}(l_u, l_l)$  is the next to be evaluated. The farther a person is from a tourist site, the less likely they are to attend, and therefore, the less suggested the attraction is [67]. The Manhattan formula was used to get the distance factor for the site (Eq. (15))

$$\text{Distance}_{\text{Geo}(l_u, l_l)} = (|x_1 - x_2|) + (|y_1 - y_2|), \tag{15}$$

where we have the target user’s geographical location  $l_u(x_1, y_1)$  and the tourist location  $l_l(x_2, y_2)$ . To cover all points and achieve the closeness of distance, we utilize the double Laplace distribution equation (Eq. (16)) [68]

$$\text{Sim}_{\text{Loc-Context}}(l_u, l_l) = \frac{1}{2\mu} * e^{\left(-\frac{|\text{Distance}_{\text{Geo}(l_u, l_l)}|}{\mu}\right)}. \tag{16}$$

The  $\mu$  coefficient is utilized to convert the decrease rate in this case. The longer the space between the tourist’s present place and the previously visited place, the fewer suggestions are offered.

Another context aspect examined by this method is the similarity of the climate and time ( $\text{Sim}_{\text{WT-context}}(C_u, C_l)$ ). During the offline process, a profile of POIs was built, and the vector form of contextual factor values for every POI was stored.

Apart from that, contextual data were applied to the existing user query in the pattern of a vector  $(\vec{V}_u)$  in compliance with "Enriching user queries by contextual data", and on the other hand, having the vector template of the contextual metrics weight of the POIs  $(\vec{W}_l)$  enables the determination of similarity via adjusted cosine formula (Eq. (17)) [61]

$$\text{Sim}_{\text{WT-context}}\left(\vec{V}_u, \vec{W}_l\right) = \frac{\vec{V}_u * \vec{W}_l}{\left|\vec{V}_u\right| * \left|\vec{W}_l\right|}. \tag{17}$$

For the existing user context and every location context, a context vector modeling such as  $(\vec{C}_u)$  and  $(\vec{C}_l)$  is created. The output data of this similarity are always between nil and one.

The list items are sorted according to the projected points for each site, which are related tourist spots.

**Candidate trip pattern stage**

The rankings of journey sequences from the trip sequential patterns DB that contained the present town were examined first in this stage, followed by the travel sequences with the top ranking. They get their rank by summing the scores of POIs that follow Eq. (15).  $\text{Pred}(u, l)$  was determined in the stage before (Eq. 14)

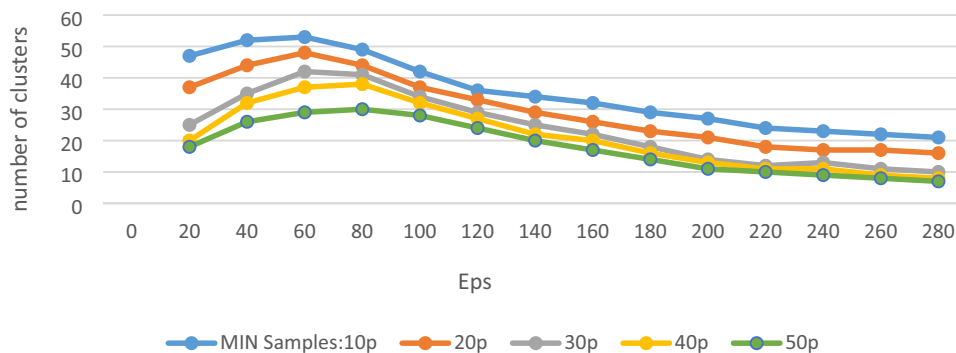
$$\begin{aligned} \text{T-Score}(\text{TravelSeq}) &= W_{\text{Time-Seq}}(T_u, T_{\text{Seq}}) \\ &* \left(\frac{1}{n} * \sum_{i=1 \& l_i \in \text{TravelSeq}}^n \text{Pred}(u, l_i)\right). \end{aligned} \tag{18}$$

Each travel sequence’s number of places is indicated by  $(n)$ .  $W_{\text{Time-Seq}}(T_u, T_{\text{Seq}})$  is used as a weight in Eq. (19) to compute the travel sequence’s value. The greater the trip sequence value, the more similar the trips are.

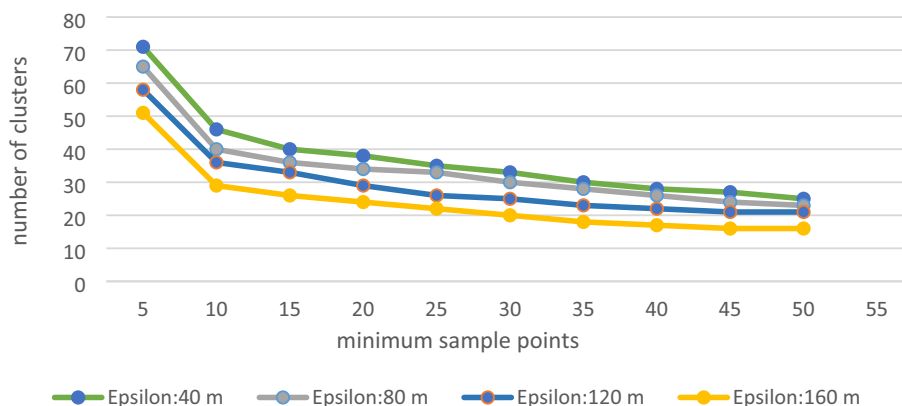
To replicate the attenuation of user preferences, we employ the forgetting function, which is an essential part of our strategy. The interests of users might change with time. This suggested strategy incorporates users’ dynamic interests by exploiting time contexts; in this scenario, journeys nearer to the user are more useful than those farther away. Equation (16) was used to determine these temporal context weights (as a penalty function, a novel adaptive combination of exponential forgetting function and exponential distribution)

$$W_{\text{Time-Seq}}(T_u, T_{\text{Seq}}) = \lambda * e^{\frac{-\text{Ln}2 * \lambda * |T_u - T_{\text{Seq}}|}{hLu}}. \tag{19}$$

**Fig. 3** Depending on the number of clusters discovered according to the Eps



**Fig. 4** Depending on the number of clusters discovered according to MinPts



Here,  $W_{\text{Time-Seq}}(T_u, T_{\text{Seq}})$  signifies the time weight reflecting how much a user’s interest has decreased;  $T_u$  indicates the current time, and  $T_{\text{Seq}}$  defines the date on which the travel was visited by subsequent users. Controlling the pace of forgetting has a half-life (in days) called  $hL_u$ . The trip’s half-life, as defined by the trip’s life cycle, is related to this context (in days). This formula governs the pace with which we forget [32, 61]. In terms of days, the suggested technique accounts for the time–space between these two times.  $hL_u$  may be considered as 15 days, considering that every journey takes an average of 1 month. The decay rate is adjusted using the time decay factor ( $\lambda$ ). In this situation, we state that  $\lambda$  equals 0.5.

Inside this step, TOP-N personalized POIs and TOP-N travel sequences were acquired, and both should be taken into account while optimizing sequential trip patterns. If the detected POI in the top list is not yet in the candidate pattern, it will be inserted based on the number of times it has been visited and the least geographical distance between two consecutive points in the candidate structure. Ultimately, relying on DB, CA, topic, and SPM, a customized trip is recommended to the present tourist.

### Simulations and experimental evaluation

In this part, many tests were conducted to show the effectiveness of the suggested strategy. For that purpose, we will go over the experimental datasets and model parameters first. The evaluation measures are discussed after that. Following that, the experiments and their results are reported and debated. Ultimately, the data are evaluated and compared to findings acquired utilizing other cutting-edge approaches, such as Flickr and Gowalla data sources. MinPts, Eps for DBSCAN, the number of topics ( $t$ ), and parameter  $\beta$  in Eq. (12) are method hyper-parameters whose values were investigated to determine how they affect the method’s performance. In the section “Evaluation dataset”, the clustering variables (MinPts, Eps) are identified and presented in Figs. 3 and 4. Then, the effect of the number of topics according to the precision measure is calculated and illustrated in Fig. 5. The weight of each DB and CF similarity between two users is then determined by the variable ( $\beta$ ) in Eq. (12) in the section “Impact of parameter  $\beta$ ”, and the result is shown in Fig. 6.



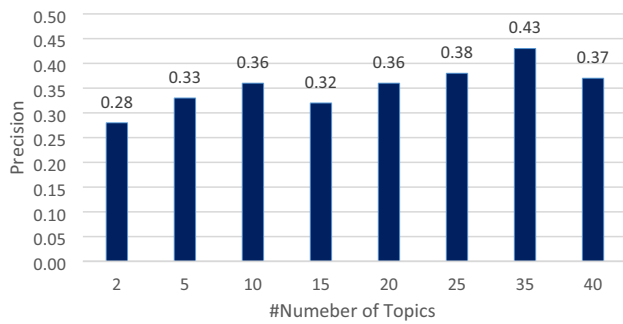


Fig. 5 Effect of the number of topics according to the precision measure

### Evaluation dataset

This study employed Flickr, one of the most popular image-uploading social networks (<https://www.flickr.com>). As a photo-based social media platform, this website has gained popularity. It was founded in 2004, and the accessibility of its vast photo database has made it a reputable data source for social science research. In addition to photo content data, Flickr photos typically contain descriptions of the photos themselves or metadata, which records supplementary information such as photo id, photographer id (owner), shooting time (time date), longitude (Lon), latitude (Lat), title, tags, and user information. This paper obtains geo-tagged Flickr images and their attribute data using the Flickr Application Programming Interface. Viewing and sharing Flickr photographs and videos does not require a Flickr account; however, sharing data does. YFCC100M Yahoo was the Flickr dataset utilized [69, 70]. This dataset is hosted at Webscope Yahoo Labs (2022). The suggested technique was tested utilizing Flickr, which includes image information.

The Application Programming Interface methods were utilized to get image information of London between 2015 and 2019. Table 3 illustrates different fields of the Flickr dataset.

In the offline state, the DBSCAN two-level clustering algorithm was employed to discover user destinations from the dataset, as defined in “Finding POIs”. We ran a study on the DBSCAN settings and then showed how the number of clusters detected changes when MinPts and Eps change. The accuracy of the method is greatly influenced by correctly determining the method’s two radius and minimum sample point factors. The size and density of clustered places can vary.

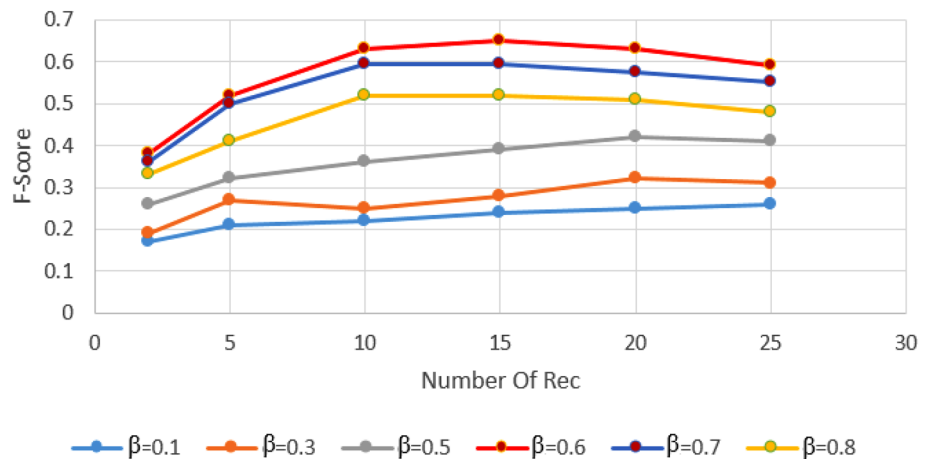
The DBSCAN settings can be modified; therefore, it is important to look them over carefully to figure out how many regions there are. In this situation, the test approach was used to discover them. Figures 3 and 4 show how the number of recognized clusters changes as these two parameters’ values alter.

The minimum sample size is a falling graph pattern when the radius reaches 120 for all attribute values, as seen in Fig. 3. For the parameter specifying the minimum sample sizes, the chart’s declining slope is decreased to a parameter of 10 in Fig. 4. The clustering variables (Eps = 120, MinPts = 10) are put to these two parameters depending on the results, resulting in a total of 36 clusters. To conduct the assessment, these data were separated into two non-overlapping halves, with 75% utilized for framework training and 25% for analyzing process.

In the topic model, the number of topics ( $t$ ) might affect the performance of our method. In this paper, the last-seen POI of each tourist is anticipated as a test. The accuracy of POI predictions is utilized as a measurement for efficiency. The outcomes are demonstrated (Fig. 5) based on their precision. As a result, the accuracy of the POI forecast maximizes at 43% for 35 topics.

We experimented with various parameter values for every formula. As a consequence, the best results were obtained with this description of the report parameter values (Eq. (9):  $\beta = 0.6$ ; Eq. (13):  $\mu = 0.5$ ; Prefix-Span: Min-Support = 0.1).

Fig. 6 The impact of DB and CF according to F1-score



**Table 3** Fields of the Flickr dataset

ID	Owner	Title	Time-taken	Tags	Latitude	Longitude
68923891584	5413236@N04	British museum	9/15/2018 16:41	British museum, London, sculpture	51.51945	– 0.12606

**Table 4** Flickr records

Images		Tourists	Location
Raw	Filtered	Filtered	
49,999	44,263	456	2957

Furthermore, the number of items evaluated in this investigation is listed in Table 4.

The second dataset used was the Gowalla dataset. This dataset is available at “<http://www.gowalla.com>”. We choose Gowalla as a real-world tourism dataset. In detail, this dataset contains 10,162 users and 24,237 POIs [71].

### The evaluation metrics

The proposed method’s accuracy and performance were evaluated using Recall, Precision, Average Precision (AP), Mean Average Precision (MAP), RMSE, F-score, and nDCG metrics.

The Recall measure is presented as the ratio of correct items proposed to the target user’s total number of relevant items (Eq. 20) [21]:

$$\text{Recall} = \frac{\text{Number of correct prediction}}{\text{Number of total relevant items}}. \tag{20}$$

As demonstrated in Eq. (21), Precision is defined as the ratio of correct item predictions to total item predictions:

$$\text{Precision} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}}. \tag{21}$$

Equation (22) determines the average Precision metric AP@N, an equation that calculates accuracy for all users:

$$\text{AP@N} = \frac{\sum_{k=1}^N (\text{Precision@k} * \text{Relevant}_k)}{M}, \tag{22}$$

$M$  is the relevant item, and  $\text{Relevant}_k$  is an index role [ $\text{Relevant}_k = 1$  if the item ( $k$ ) on the recommended list is a related POI, that anyway  $\text{Relevant}_k = 0$ ].

Equation (23) describes the mean average Precision metric for ( $m$ ) users

$$\text{MAP@m} = \frac{\sum_{k=1}^m \text{AP}(u)}{m}. \tag{23}$$

Root-mean-square error (RMSE) highlights bigger absolute error levels (Eq. (24))

$$\text{RMSE} = \sqrt{\frac{\sum_{(u,i) \in R_{u,i}} (\hat{r}_{u,i} - r_{u,i})^2}{N}}. \tag{24}$$

The user ( $u$ ) and place ( $i$ ) methods forecast the score value as ( $\hat{r}_{u,i}$ ). The genuine value of the user’s ( $u$ ) rating for place ( $i$ ) is ( $r_{u,i}$ ), and the total number of examined places is ( $N$ ).

The F-score is identified as Eq. (25)

$$\text{F – Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{25}$$

The projected suggestions’ ranking efficiency is compared utilizing the normalized discounted cumulative gain (NDCG) (Eq. (26)). The more related topics of attention shown at the head of the proposed list, the higher the NDCG score [3, 72]:

$$\text{NDCG} = \frac{\text{DCG}}{i\text{DCG}} = \frac{\sum_{i=1}^p \frac{2^{\text{rel}_i - 1}}{\log_2^{(i+1)}}}{\sum_{i=1}^{|\text{rel}_p|} \frac{2^{\text{rel}_i - 1}}{\log_2^{(i+1)}}}. \tag{26}$$

Here,  $\text{Rel}_i$  stands for the element ranked at place ( $i$ ), and also,  $\text{rel}_p$  stands for the list of related items in the relevant group in place ( $p$ ).

### Comparison approaches

Throughout this part, we evaluate our model with the following methods in Table 5.

### Experimental results

The influence of CF and DB on suggestion accuracy is discussed first in this paragraph. The impact of neighborhood numbers on suggestion quality is then investigated.

The next sections compare the efficiency of our Topic-SeqHybrid technique to previous techniques that solve cold start and data sparsity challenges based on MAP, Precision, Recall, RMSE, F-score, and NDCG criteria.

### Impact of parameter $\beta$

The weight of each DB and CF similarity among two tourists is determined by the variable ( $\beta$ ) in Eq. (12). Consequently,

**Table 5** Comparison methods

Methods	#References	Description
(CF)	[6]	Cosine similarity
(PR)	[50]	Public popularity
(Pre_CA-CF)	[73]	Contextual pre-filtering similarity measure
(CA-CF)	[74]	Jaccard measure
(ACA-CF)	[7]	Cosine similarity + asymmetric schema with Jaccard measure
(GSP-CACF)	[51]	CF and GSP
(CA-MSDT)	[60]	CA and decision tree classification
(Prefix-CSTR)	[50]	Prefix-span algorithm
(ADBCACF)	[13]	Asymmetric CF and demographic data
(SeqHybrid)	[24]	Our previous work describes a sequential recommender system that combines context-awareness, demographic-based, and asymmetric CF

the value selected for this factor has a substantial impact on the performance of the current plan. Figure 6 illustrates the results of the trials about the F-score that was used to establish the optimal value for the parameter ( $\beta$ ). The chart with a figure of  $\beta = 0.6$  fared the best among the other figures of variable  $\beta$ , as can be observed. Consequently, setting this factor's starting figure to 0.6 is a good decision. As a result, the impact of neighbors on tourism suggestions takes precedence over demographic data.

### The effect of the neighborhood numbers

By proposing 2, 6, and 12 POIs among all president's suggestions, we studied and applied the influence of neighbor size on the accuracy of anticipated suggestions for neighbors of different sized. In this scenario, the number of neighbors grew between 2 and 18. Concerning MAP, the findings are shown in Fig. 7; when the number of suggestions surpassed six, adding POIs dramatically lowered the validity of the recommendation results when using MAP, according to these studies.

The preferences of users fluctuate, and they like to visit no more than six places in each region during their journey. As a result, it appears that recommending two to six POIs for tailored suggestions is reasonable. Because of clustering, context data, and demographic information, TopicSeqHybrid exhibited a greater MAP score than previous approaches. According to the findings, asymmetric strategies outperform symmetric techniques also the topic distribution of

tourist trip records. Furthermore, the GspCACF, ADBCACF, SeqHybrid, and TopicSeqHybrid RSs combine sequential and non-sequential data. Greater neighbors for the present tourist may be located as a consequence of the suggested strategy, and the POIs produced by such neighbors are more precise.

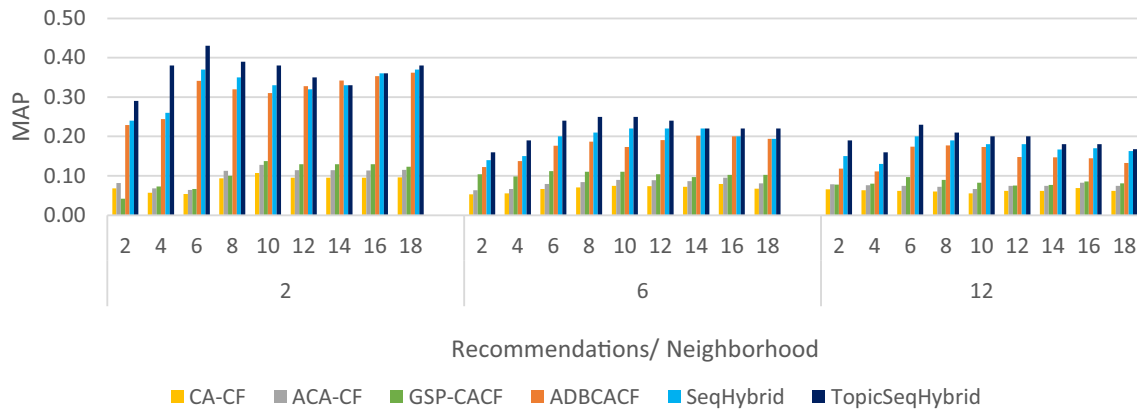
The Recall metric grew as the number of suggestions was raised, as seen in Fig. 8. This is because the top-N suggestion now includes more precise POIs. Consequently, in terms of Recall, TopicSeqHybrid outperformed previous methods. Asymmetric strategies outperform other approaches, according to the result. When compared to the other approaches, the CF and PR approaches generated the lowest accurate findings caused by the absence of clustering and disrespect for context.

### The impact of the highest suggestions

As seen in Fig. 9, as the figures for suggestions rose, the precision dropped. The cause that the highest suggestion now contains more accurate POIs was the primary driver for this enhancement. Tourists may not be able to visit all of the suggested places due to a lack of information on personal intentions. The results show that asymmetric approaches outperformed symmetric techniques, even though the suggested technique beats the other methods in terms of precision rate. Furthermore, when compared to the other approaches, the PR and CF approaches generated the lowest accurate findings caused by the absence of clustering and disrespect for context.

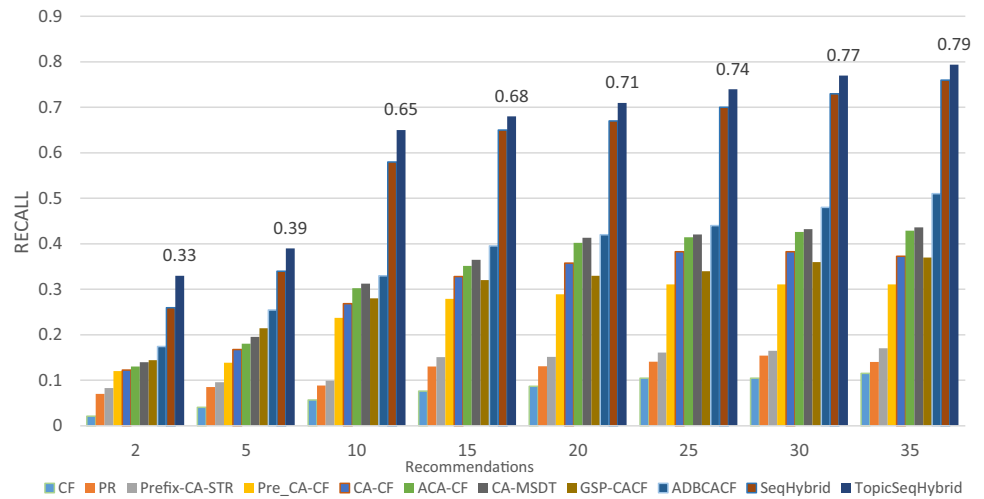
For numerous suggestions based on F-score measures, TopicSeqHybrid outperformed other techniques, as shown in Fig. 10. The suggested technique beats previous alternatives regarding cold start and data sparsity, as seen in this figure. By combining information in user profiles with an asymmetric schema technique to estimate the desired person's nearest neighbors, TopicSeqHybrid was able to produce improved outcomes. Furthermore, using demographic data on users might help forecast user preferences for future visits, so alleviating the cold start problem. Instead of depending on a single site to discover POIs, a clustering approach can assist ease the data sparsity problem. This framework suggestion may also be customized by integrating SPM with the top-N POIs.

The distance between the expected and actual rating is calculated using the RMSE measure. In recommender systems, this measure is frequently used to quantify the diversity between an item's real and projected ratings. Non-context-aware procedures, in general, have a larger mistake rate than context-aware approaches, as seen in Fig. 11. The suggested method beats other contextual approaches proposed in earlier research and had a lower error rate than non-contextual approaches. TopicSeqHybrid was able to handle the cold start

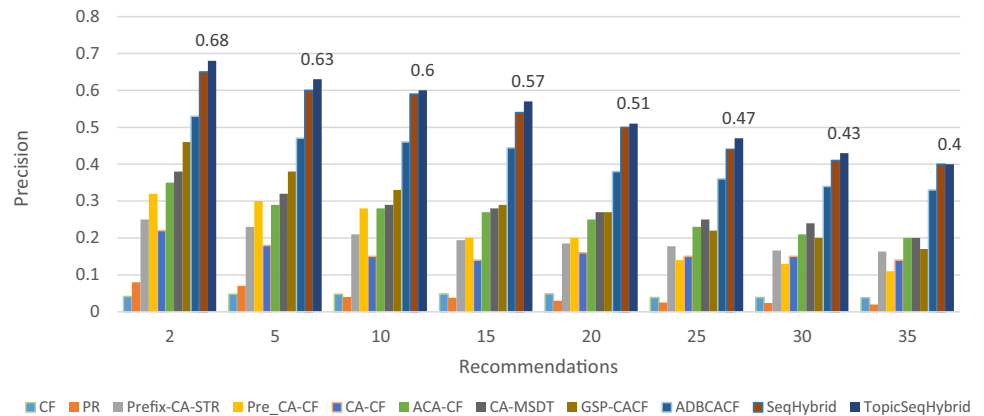


**Fig. 7** The list of recommendations is used to compare TopicSeqHybrid to others with regard to the MAP metric (Flickr)

**Fig. 8** The list of recommendations is used to compare TopicSeqHybrid to others with regard to the Recall metric (Flickr)



**Fig. 9** The list of recommendations is used to compare TopicSeqHybrid to others with regard to the Precision metric (Flickr)



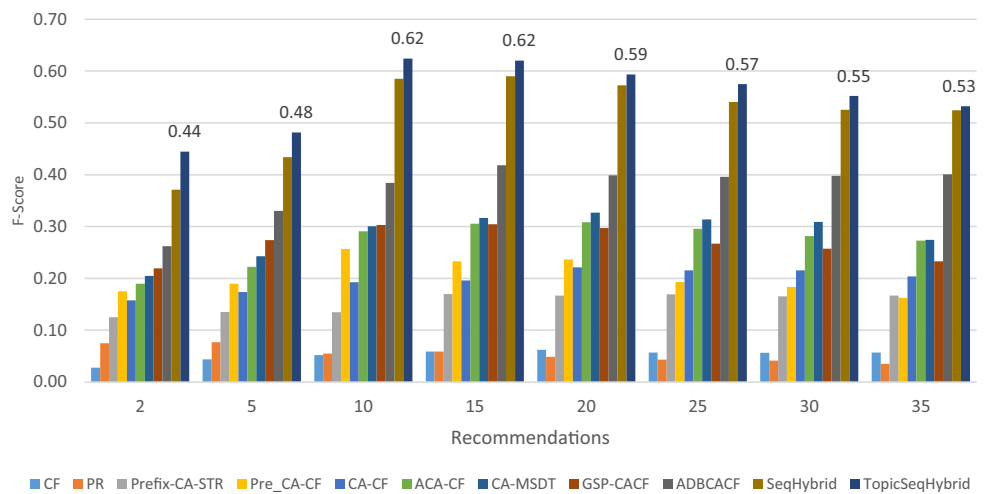
issue better than other techniques, because demographic data and the subject distribution of users were included.

**Evaluation of TopicSeqHybrid with the Gowalla**

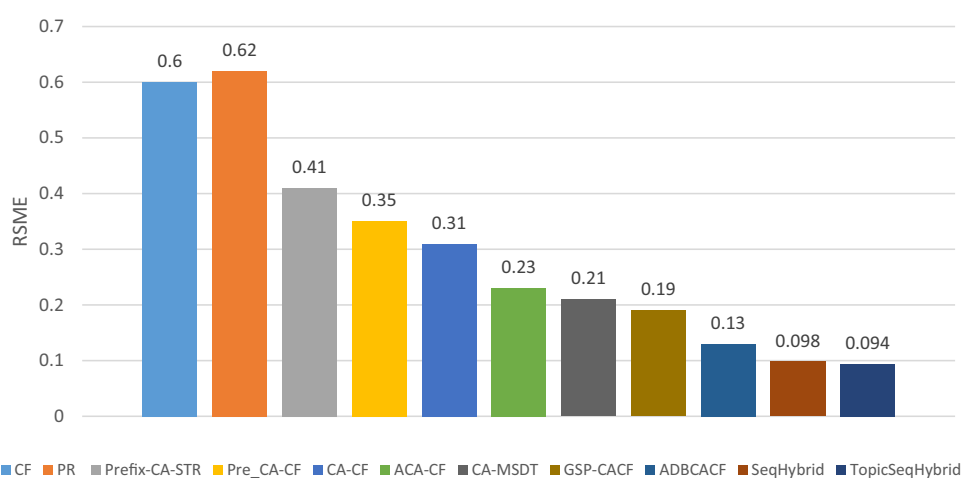
Throughout this work, the Gowalla data source was utilized as another dataset to assess TopicSeqHybrid. The outcomes

of evaluations, which are based on Recall and Precision measures, are shown in Figs. 12 and 13. Due to the incorporation of demographic data, contextual data, and an asymmetric schema, TopicSeqHybrid delivered improved results. The proposed solution proved very successful in addressing the cold start issue when contrasted to prior alternatives. When these two datasets were compared, it was discovered that

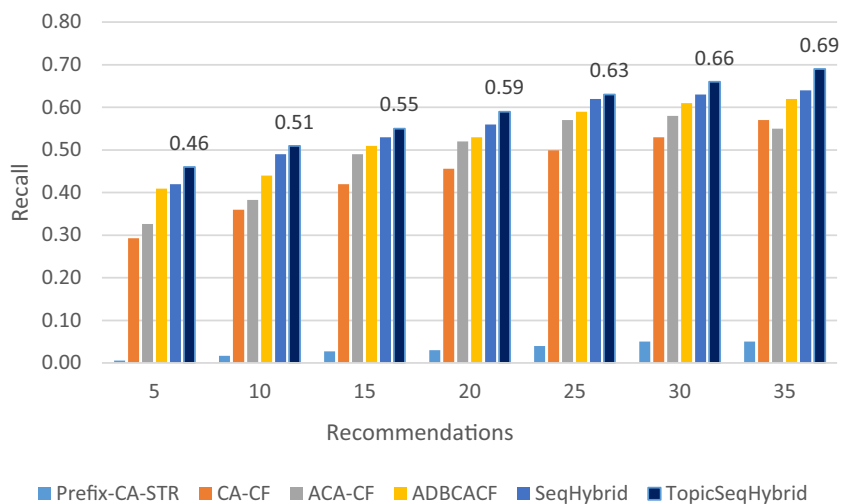
**Fig. 10** The list of recommendations is used to compare TopicSeqHybrid to others with regard to the F-score metric (Flickr)



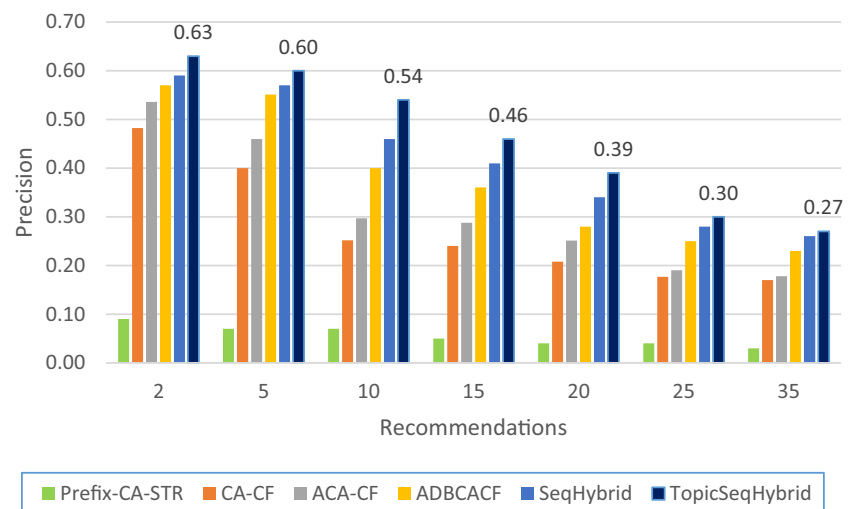
**Fig. 11** TopicSeqHybrid is compared to others according to the RMSE measure (Flickr)



**Fig. 12** TopicSeqHybrid is compared to others according to the Recall measure based on the number of recommendations (Gowalla)



**Fig. 13** TopicSeqHybrid is compared to others according to the Precision measure based on the number of recommendations (Gowalla)



the Gowalla had poorer Precision and Recall than Flickr, probably because the Gowalla had fewer demographic characteristics. The volume of the data in this data source, as well as the number of neighbors, has an impact on the outcomes.

#### Evaluation of TopicSeqHybrid by NDCG metric

In sequential approaches, the normalized discounted cumulative gain (NDCG) measure is utilized to quantify the ranking efficiency of expected suggestions. The NDCG measure is used to highlight the rating proficiency of the sequence suggestion strategy in Fig. 14. These results showed that TopicSeqHybrid projected more suitable recommendations when compared to previous approaches.

#### Example trips

As a case study, this paper investigates a visitor who visits a city. Once a recommendation is requested, this tourist's present location is dynamically determined using information from their cell phone. Next, potential locations consist of the city's AOIs and POIs. Since multiple POIs exist in each city and a tourist is unable to visit all, plus, as it is assumed that the tourist does not possess a sufficient history of visiting POIs in a given city (cold start problem), their nearest neighbors are recognized based on their current contexts, topics, and preferences. Finally, visitors are advised to list the top-N POI routes based on their activities and information about their nearby neighbors (Fig. 15). In this investigation, we consider all data received from tourist visits and present dynamic preferences based on the time and place of the targeted tourist. Therefore, with a cold start and scarce information scenarios, we may utilize all existing data to provide accurate suggestions without the direct participation of visitors.

During the case study, we conducted further trials in the Toronto metropolitan region to identify routes for active visitors (Table 6).

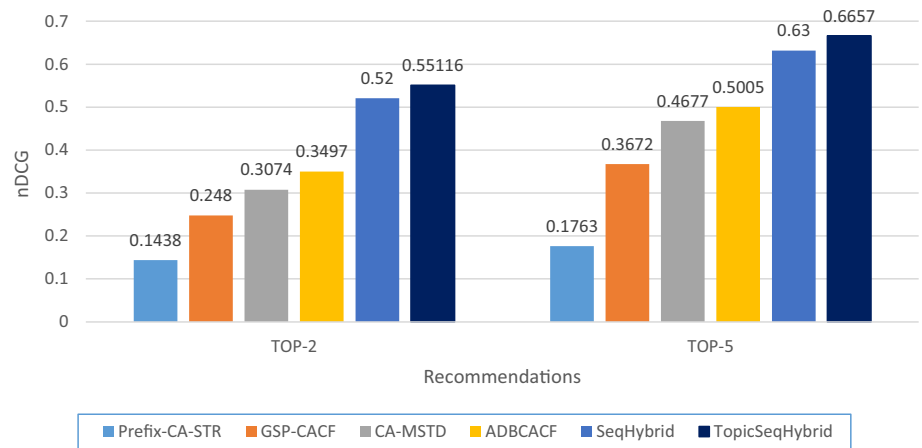
Similar neighbors are identified according to their current settings, topics, histories, and qualities once an active tourist seeks a recommendation. The top-N recommendations are then picked and presented. Performance evaluations of this stage are shown in Fig. 16. Based on the results, the recommended technique exhibits the highest F-score based on an asymmetric topic CF scheme. The approaches that varied somewhat from ours ranked second. Nevertheless, the distinction between the two methods is modest, as tourists in this city have tastes and knowledge that are more similar.

#### The complexity of the proposed approach

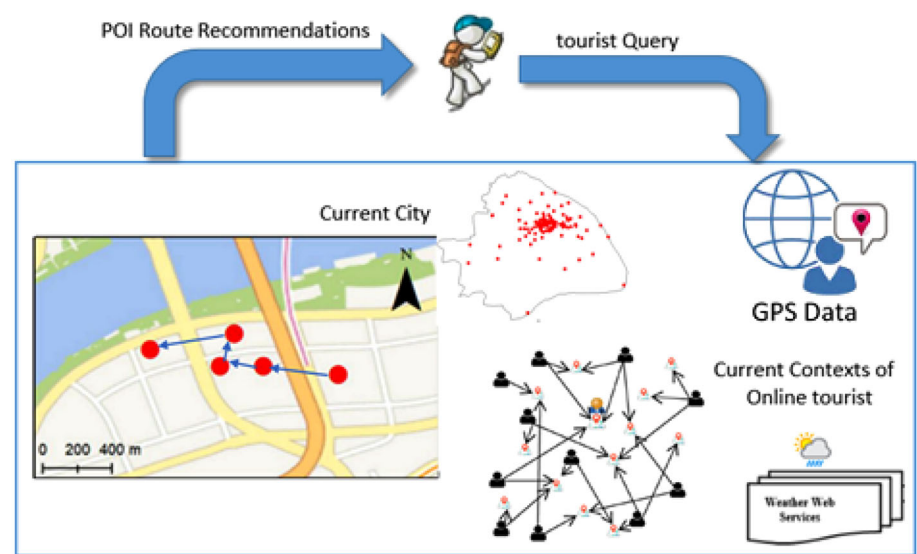
Since this paper employed the DBSCAN clustering technique for identifying AOIs and POIs, the complexity of this model was primarily derived from this technique. DBSCAN potentially visits each point of the dataset multiple times (e.g., as candidates to different clusters). The entire algorithm's complexity is  $O(n^2)$ , where  $n$  is the number of degrees in the data set. In practical terms, the time complexity is primarily determined by the number of query invocations. DBSCAN executes exactly one of these queries for each point, and if an efficient indexing structure that executes a neighborhood query in  $O(\log n)$  is used, the overall average runtime complexity is  $O(n \log n)$ . Without an accelerating index structure or degenerated data (such as all points within a distance less than  $\epsilon$ ), the worst-case runtime complexity remains  $O(n^2)$ .

It is also important to note that, due to the algorithm's high computational complexity, it is not advisable to use it online; like our method, it must run during the offline phase.

**Fig. 14** TopicSeqHybrid is compared to others according to the NDCG measure



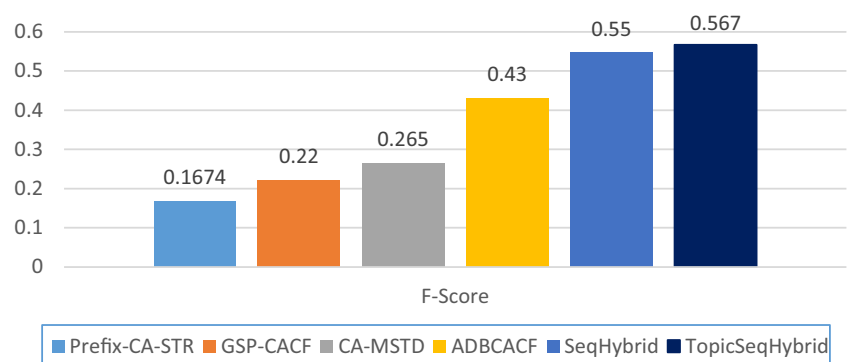
**Fig. 15** Example trip recommended by TopicSeqHybrid



**Table 6** Flickr dataset description for our case study

City	Photos	Tourists	POI visits	Trip seq.
Toronto	1,57,500	1390	39,410	6053

**Fig. 16** Results for performances of some methods



## Conclusions and future work

Using contextual and demographic data as well as geo-tagged social network photos, this article presents a unique context-aware RS for personalized tourist destinations. To construct a hybrid RS, the researchers used innovative asymmetric schema, context-aware filtering, DB, and sequential pattern mining algorithms. Because most recommender systems rely on weak data, demographic data were explored to manage the cold start problem.

The recommended technique outperformed prior approaches due to the integration of contextual information and the fact that it was employed for both contextual pre-filtering and contextual modeling. When producing tourist suggestions, it was determined that every user's context is critical. The suggested technique's personalization refers to how it makes use of the user's choices. Furthermore, the proposed technique improved from the use of DBSCAN clustering at two levels to detect POIs in any area, which made clustering detection easier and more complicated. The TF-IDF approach was used to assess context similarity.

Additionally, the Prefix-Span approach is applied, which outperformed the other approaches due to the Prefix-Span algorithm's sequential movement pattern. Two data sources were examined to evaluate the efficacy of this technique (Flickr and Gowalla). The recommended strategy, according to the results of the comparison, can offer more precise locations than other ways. The proposed technique beats all current recommendation systems in terms of discovering users who were more similar to the present tourist and showed superior outcomes while coping with data sparsity and cold start difficulties.

We incorporated tourist taste and current place data into the probabilistic manners pattern by merging topic instances and the Markov formula. In other words, we deduce that similar tourists relying on their topic distribution of trip records may get more useful individual trip suggestions. Because customers are more likely to accept an RS that makes ideas based on their likes and interests, this article used the median of the rating to indicate a unique enhanced ACF technique based on user preferences. The study's journey sequences would help travelers plan their vacations and make them more convenient. This method increases the interactivity of the tourist recommender system by deriving trip patterns from visitor behavior. This technology increases user engagement with online trip RS by recognizing and tailoring journey patterns based on users' travel behavior. Tourist behavior patterns can give insight into visitors' intentions and desires by anticipating users' future interests and activities based on recent behaviors.

## Future work

The context information utilized by this proposed method consists of geographic and time context information. Future research will attempt to incorporate additional contextual information, such as trip cost, travel companion details, and total travel time. These factors may increase the effectiveness of the tourist recommender method, as they may interact with various tourist preferences. The dataset could potentially be expanded to include additional cities. In addition, this model could be improved by incorporating user opinions extracted from their comments on social media or by utilizing bidirectional encoder representations from transformer (BERT) to find more suitable sequential trips.

**Data availability** The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** None.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Noorian A, Ravanmehr R, Harounabadi A, Nouri F (2020) Trust-based tourism recommendation system using context-aware clustering. *Tour Manag Stud*. <https://doi.org/10.22054/tms.2020.41870.2137>
- Jiang S, Qian X, Shen J, Mei T (2015) Travel recommendation via author topic model based collaborative filtering. In: *Int. conf. multimed. model.*, pp 392–402
- Vineela A, Lavanya Devi G, Nelaturi N, Dasavatara Yadav G (2021) A comprehensive study and evaluation of recommender systems. In: *Microelectron. electromagn. telecommun*. Springer, pp 45–53
- Abbasi-Moud Z, Vahdat-Nejad H, Sadri J (2021) Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst Appl* 167:114324. <https://doi.org/10.1016/j.eswa.2020.114324>
- Dakhel AM, Malazi HT, Mahdavi M (2018) A social recommender system using item asymmetric correlation. *Appl Intell* 48:527–540
- Pirasteh P, Hwang D, Jung JE (2014) Weighted similarity schemes for high scalability in user-based collaborative filtering. *Mob Netw Appl*. <https://doi.org/10.1007/s11036-014-0544-5>



7. Pirasteh P, Hwang D, Jung JJ (2015) Exploiting matrix factorization to asymmetric user similarities in recommendation systems. *Knowl-Based Syst*. <https://doi.org/10.1016/j.knosys.2015.03.006>
8. Ojagh S, Malek MR, Saeedi S, Liang S (2020) A location-based orientation-aware recommender system using IoT smart devices and Social Networks. *Future Gener Comput Syst* 108:97–118
9. Villegas NM, Sánchez C, Diaz-Cely J, Tamura G (2018) Characterizing context-aware recommender systems: a systematic literature review. *Knowl-Based Syst* 140:173–200
10. Kala KU, Nandhini M (2019) Context-category specific sequence aware point-of-interest recommender system with multi-gated recurrent unit. *J Ambient Intell Humaniz Comput* 1–11
11. Ali Z, Qi G, Kefalas P, Abro WA, Ali B (2020) A graph-based taxonomy of citation recommendation models. Springer Netherlands, Dordrecht. <https://doi.org/10.1007/s10462-020-09819-4>
12. Mirhasani M, Ravanmehr R (2020) Alleviation of cold start in movie recommendation systems using sentiment analysis of multi-modal social networks. *J Adv Comput Eng Technol* 6:251–264
13. 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. *Electron Commer Res Appl* 42:100978. <https://doi.org/10.1016/j.elerap.2020.100978>
14. Sarkar JL, Majumder A, Panigrahi CR, Roy S (2020) MULTITOUR: a multiple itinerary tourists recommendation engine. *Electron Commer Res Appl* 40:100943
15. Lyu D, Chen L, Xu Z, Yu S (2020) Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geo-tagged photos. *Appl Intell* 50:924–938. <https://doi.org/10.1007/s10489-019-01566-6>
16. Spyrou E, Mylonas P (2016) Analyzing Flickr metadata to extract location-based information and semantically organize its photo content. *Neurocomputing* 172:114–133
17. Sun X, Huang Z, Peng X, Chen Y, Liu Y (2019) Building a model-based personalised recommendation approach for tourist attractions from geotagged social media data. *Int J Digit Earth* 12:661–678
18. Tang J, Wang K (2018) Personalized top-n sequential recommendation via convolutional sequence embedding. In: *Proc. elev. ACM int. conf. web search data min.*, pp 565–573
19. Bin C, Gu T, Sun Y, Chang L (2019) A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining. *Multimed Tools Appl* 78:35135–35156. <https://doi.org/10.1007/s11042-019-08096-w>
20. Hong M, Jung JJ (2021) Multi-criteria tensor model for tourism recommender systems. *Expert Syst Appl* 170:114537
21. Aggarwal CC (2016) *Recommender systems text book*. Springer International Publishing Switzerland, Cham
22. Dara S, Chowdary CR, Kumar C (2020) A survey on group recommender systems. *J Intell Inf Syst* 54:271–295. <https://doi.org/10.1007/s10844-018-0542-3>
23. Kuanr M, Mohapatra P (2021) Recent challenges in recommender systems: a survey. In: *Prog adv. comput. intell. eng.* Springer, pp 353–365
24. Noorian A, Harounabadi A, Ravanmehr R (2022) A novel Sequence-Aware personalized recommendation system based on multidimensional information. *Expert Syst Appl* 202:117079
25. Iwata T, Watanabe S, Yamada T, Ueda N (2009) Topic tracking model for analyzing consumer purchase behaviour. In: *Twenty-first int. jt. conf. artif. intell.*
26. Chen L, Lyu D, Xu Z, Long H, Chen G (2020) A content-location-aware public welfare activity information push system based on microblog. *Inf Process Manag* 57:102137
27. Chen C, Zheng X, Wang Y, Hong F, Lin Z (2014) Context-aware collaborative topic regression with social matrix factorization for recommender systems. In: *Proc. AAAI conf. artif. intell.*
28. Liu Q, Ge Y, Li Z, Chen E, Xiong H (2011) Personalized travel package recommendation. In: *2011 IEEE 11th int. conf. data min.*, pp 407–416
29. Hofmann T (2003) Collaborative filtering via gaussian probabilistic latent semantic analysis. In: *Proc. 26th annu. int. ACM SIGIR conf. res. dev. informaon retr.*, pp 259–266
30. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet distribution. *J Mach Learn Res* 3:993–1022
31. Hofmann T (2017) Probabilistic latent semantic indexing. In: *Proceedings of the ACM SIGIR forum*, pp 211–218
32. Chen L, Zhang L, Cao S, Wu Z, Cao J (2020) Personalized itinerary recommendation: deep and collaborative learning with textual information. *Expert Syst Appl* 144:113070. <https://doi.org/10.1016/j.eswa.2019.113070>
33. Kurashima T, Iwata T, Irie G, Fujimura K (2010) Travel route recommendation using geotags in photo sharing sites. In: *Int. conf. inf. knowl. manag. proc.*, pp 579–588. <https://doi.org/10.1145/1871437.1871513>
34. Kurashima T, Iwata T, Irie G, Fujimura K (2013) Travel route recommendation using geotagged photos. *Knowl Inf Syst* 37:37–60
35. Yin Z, Cao L, Han J, Zhai C, Huang T (2011) Geographical topic discovery and comparison. In: *Proc. 20th int. conf. world wide web*, pp 247–256
36. Pozdnoukhov A, Kaiser C (2011) Space–time dynamics of topics in streaming text. In: *Proc. 3rd ACM SIGSPATIAL int. work. locat. soc. networks*, pp 1–8
37. Zhao F, Zhu Y, Jin H, Yang LT (2016) A personalized hashtag recommendation approach using LDA-based topic model in microblog environment. *Future Gener Comput Syst* 65:196–206
38. Jiang S, Qian X, Mei T, Fu Y (2016) Personalized travel sequence recommendation on multi-source big social media. *IEEE Trans Big Data* 2:43–56. <https://doi.org/10.1109/tbdata.2016.2541160>
39. Sun C-Y, Lee AJT (2017) Tour recommendations by mining photo sharing social media. *Decis Support Syst* 101:28–39
40. Ren X, Song M, Haihong E, Song J (2017) Context-aware probabilistic matrix factorization modeling for point-of-interest recommendation. *Neurocomputing* 241:38–55
41. Tsai C-Y, Lai B-H (2015) A location-item-time sequential pattern mining algorithm for route recommendation. *Knowl-Based Syst* 73:97–110
42. Gediminas Adomavicius AT, Mobasher B, Ricci F (2011) Context-aware recommender systems. *Assoc Adv Artif Intell AI Mag* 3:67–80
43. Xu Z, Chen L, Chen G (2015) Topic based context-aware travel recommendation method exploiting geotagged photos. *Neurocomputing* 155:99–107. <https://doi.org/10.1016/j.neucom.2014.12.043>
44. Yu D, Yu T, Wang D, Shen Y (2022) NGPR: a comprehensive personalized point-of-interest recommendation method based on heterogeneous graphs. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-022-13088-4>
45. Memon I, Chen L, Majid A, Lv M, Hussain I, Chen G (2015) Travel recommendation using geo-tagged photos in social media for tourist. *Wirel Pers Commun* 80:1347–1362. <https://doi.org/10.1007/s11277-014-2082-7>
46. Sun Y, Gu T, Bin C, Chang L, Kuang H, Huang Z, Sun L (2018) A multi-latent semantics representation model for mining tourist trajectory. In: *Pacific rim int. conf. artif. intell.*, pp 463–476
47. Trivonanda R, Mahendra R, Budi I, Hidayat RA (2020) Sequential pattern mining for e-commerce recommender system. In: *2020 int. conf. adv. comput. sci. inf. syst.*, pp 393–398
48. Tarus JK, Niu Z, Kalui D (2018) A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Comput* 22:2449–2461
49. Lim KH, Chan J, Leckie C, Karunasekera S (2018) Personalized trip recommendation for tourists based on user interests,

- points of interest visit durations and visit recency. *Knowl Inf Syst* 54:375–406
50. Majid A, Chen L, Mirza HT, Hussain I, Chen G (2015) A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. *Data Knowl Eng* 95:66–86. <https://doi.org/10.1016/j.datak.2014.11.001>
  51. Tarus JK, Niu Z, Yousif A (2017) A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Gener Comput Syst* 72:37–48
  52. Li C-T, Chen H-Y, Chen R-H, Hsieh H-P (2018) On route planning by inferring visiting time, modeling user preferences, and mining representative trip patterns. *Knowl Inf Syst* 56:581–611. <https://doi.org/10.1007/s10115-017-1106-5> Springer Nature
  53. Cai G, Lee K, Lee I (2018) Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos. *Expert Syst Appl* 94:32–40. <https://doi.org/10.1016/j.eswa.2017.10.049>
  54. Hsueh Y-L, Huang H-M (2019) Personalized itinerary recommendation with time constraints using GPS datasets. *Knowl Inf Syst* 60:523–544. <https://doi.org/10.1007/s10115-018-1217-7>
  55. Ravi L, Subramaniaswamy V, Vijayakumar V, Chen S, Karmel A, Devarajan M (2019) Hybrid location-based recommender system for mobility and travel planning. *Mob Netw Appl* 24:1226–1239. <https://doi.org/10.1007/s11036-019-01260-4>
  56. Shi H, Chen L, Xu Z, Lyu D (2019) Personalized location recommendation using mobile phone usage information. *Appl Intell* 49:3694–3707. <https://doi.org/10.1007/s10489-019-01477-6>
  57. Alrasheed H, Alzeer A, Alhowimel A, Shameri N, Althyabi A (2020) A multi-level tourism destination recommender system. *Procedia Comput Sci* 170:333–340. <https://doi.org/10.1016/j.procs.2020.03.047>
  58. Wang Y, Wang P, Liu Z, Zhang LY (2021) A new item similarity based on  $\alpha$ -divergence for collaborative filtering in sparse data. *Expert Syst Appl* 166:114074
  59. Zhao K, Zhang Y, Yin H, Wang J, Zheng K, Zhou X, Xing C (2020) Discovering subsequence patterns for next POI recommendation. In: *IJCAI*, pp 3216–3222
  60. Hsu C-L (2021) A multi-valued and sequential-labeled decision tree method for recommending sequential patterns in cold-start situations. *Appl Intell* 51:506–526
  61. Xu Z, Li L, Yan M, Liu J, Luo X, Grundy J, Zhang Y, Zhang X (2021) A comprehensive comparative study of clustering-based unsupervised defect prediction models. *J Syst Softw* 172:110862
  62. Ray B, Garain A, Sarkar R (2021) An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Appl Soft Comput* 98:106935. <https://doi.org/10.1016/j.asoc.2020.106935>
  63. Han J, Kamber M, Pei J (2012) *Data mining: concepts and techniques*, vol 10. Morgan Kaufman Publ., Waltham, pp 971–978
  64. Anwar T, Uma V (2019) CD-SPM: cross-domain book recommendation using sequential pattern mining and rule mining. *J King Saud Univ Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2019.01.012>
  65. Zeng T, Acuna DE (2020) GotFunding: a grant recommendation system based on scientific articles. In: *Proc. assoc. inf. sci. technol*, vol 57, p e323
  66. Hu Y, Yang B (2015) Enhanced link clustering with observations on ground truth to discover social circles. *Knowl-Based Syst* 73:227–235
  67. Zhang C, Li T, Ren Z, Hu Z, Ji Y (2019) Taxonomy-aware collaborative denoising autoencoder for personalized recommendation. *Appl Intell* 49:2101–2118
  68. Han M, Lee J (2015) Bayesian typhoon track prediction using wind vector data. *Commun Stat Appl Methods* 22:241–253
  69. Flickr (2022) <http://www.Flickr.com>. Accessed 23 Jan 2022
  70. Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, Borth D, Li L-J (2016) YFCC100M: the new data in multimedia research. *Commun ACM* 59:64–73
  71. Liu C, Liu J, Wang J, Xu S, Han H, Chen Y (2019) An attention-based spatiotemporal gated recurrent unit network for point-of-interest recommendation. *ISPRS Int J Geo-Inf* 8:355
  72. Shokeen J, Rana C (2020) Social recommender systems: techniques, domains, metrics, datasets and future scope. *J Intell Inf Syst* 54:633–667
  73. Kefalas P, Manolopoulos Y (2017) A time-aware spatio-textual recommender system. *Expert Syst Appl* 78:396–406. <https://doi.org/10.1016/j.eswa.2017.01.060>
  74. Linda S, Bharadwaj KK (2019) A genetic algorithm approach to context-aware recommendations based on spatio-temporal aspects. In: *Integr. intell. comput. commun. secur.* Springer, pp 59–70

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.