Multimedia Tools and Applications (2024) 83:77565–77594
https://doi.org/10.1007/s11042-024-18522-3

A personalized context and sequence aware point of interest recommendation

Ali Noorian[1]

Received: 16 August 2022 / Revised: 4 January 2024 / Accepted: 29 January 2024 /
Published online: 27 February 2024

Abstract
This study introduces an innovative hybrid approach for personalized trip recommendations, aiming to enhance existing recommender systems by leveraging multidimensional data. Our proposed method integrates user preferences and diverse contextual factors to address challenges related to data sparsity effectively. To overcome this hurdle, our methodology employs a clustering approach, streamlining the extraction of Points of Interest (PoI) and reducing computational complexity. The framework comprises three key components: I) a unique strategy for context assessment, achieved by combining contextual information in vector form through the Term-Frequency-Inverse-Document-Frequency technique, II) the incorporation of tourist demographic information to alleviate the Cold Start problem, and III) the implementation of an asymmetric schema that elevates the traditional similarity paradigm. Moreover, our approach utilizes personalized PoIs in consecutive travel patterns, enabling the retrieval and ranking of an optimal list of potential routes. The experimental results based on Flickr and Yelp datasets reveal that the proposed method surpasses prior work on all three metrics, achieving a significant 8% increase in precision and an 11% increase in F-Score, thereby enhancing the quality metrics of personalized trip recommendations.

Keywords Personalized recommendations · Multidimensional data · Text mining · Context-aware · Sequential · Clustering approach

1 Introduction

Recommender systems are applications that suggest desired items to users [28]. The tourism industry, which attempts to deliver personalized user experience and context, is one of the most prevalent recommender systems (RS) implementations. [42, 45]. There has been an increase in the number of articles using Location-based Social networks (LBSN) and spatial–temporal data in tourist RS during the last several years [1, 11]. On the other hand, existing tourism recommender systems have several drawbacks,

✉ Ali Noorian
anoorian@asu.edu

1 School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA

primarily 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 recommendations that rely on users' travel habits, who act the same as targeted tourists. In real-world applications, however, user similarity may differ, implying that most contemporary symmetric techniques yield lower precise findings [10, 30, 31]. On the other hand, context-aware (CA) recommender systems take into account the users' context and provide more accurate suggestion outputs, given that numerous tools now gather information on the users' state [29, 44, 58].

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 severe problems with recommender systems [4, 27]. 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 at predicting significant evaluations, and conventional techniques may result in bad suggestions [19, 34]. Given the significantly increasing number of tourist images shared on social networks, their data can be used to develop RSs. Geo-tagged photos may be used to detect real-life user journey histories as a representative of tourist visits [24].

For travelers unaware 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 [36–38]. In other words, visitors prefer having a journey by pre-arranged POIs to receiving a list of recommended POIs. Therefore, an RS that can create pre-arranged POIs is more advantageous for the tourist [7]. 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 [59].

Because recommender systems are complex, hybrid techniques may be utilized to improve performance, especially with the rise of social networking. Several hybrid techniques to recommender systems have recently been applied [3, 13, 20]. The current research proposes a novel hybrid technique for improving model performance and overcoming the abovementioned drawbacks. To overcome the aforementioned shortcomings of previous techniques, our framework applied a mixed method to CF, demographic-based (DB), and CA. And. This work offers acceptable tourist places and sequences based on changing tourist preferences. The study also manages the issue of cold starts by using users' demographic information. Asymmetric schema is also utilized to solve symmetric user similarity issues and increase algorithm performance.

Consider Sarah, a tourist exploring a vibrant city for the first time. She has varied interests, from historical landmarks to trendy coffee shops, and her preferences evolve as she navigates through the city's neighborhoods. Traditional recommender systems might struggle with the "cold start" problem, lacking sufficient information about Sarah's preferences at the beginning of her journey.

Our research addresses this challenge by intelligently combining CF, DB, and CA methods in a hybrid framework. To kickstart recommendations for Sarah, we leverage her demographic information, such as age, interests, and travel history, offering a more informed starting point. Additionally, our approach incorporates an asymmetric schema to enhance the accuracy of user similarity.

The goal is not just to suggest interesting POIs to Sarah but also to craft a personalized sequence of recommendations that align with her evolving tastes and the context of her travel. In the following sections, we delve into the technical details of our novel hybrid approach, demonstrating its superiority through an experimental evaluation using datasets from Flickr and YELP.

This study is an essential supplement to our previous paper [15]. This research improved the last method by incorporating additional key factors about tourists. 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 main contributions of this study are resumed as follows:

- Proposing an efficient synergy of asymmetric schema and SPM.
- Utilizing the Term-Frequency-Inverse-Document-Frequency(TF-IDF) algorithm for the similarity between context parameters.
- Representing an improved CF approach with user's preferences
- Applying the sequence-aware recommendation method to optimize personalized route

The remainder of this study is organized as follows: Section 2 reviews the literature relevant to our research. The problem formulation and our proposed method, called NewSeqHybrid, are introduced in Section 3. Section 4 further discusses the proposed methodology and its testing and evaluation results compared with the existing papers. Finally, Section 5 provides concluding remarks and future works.

2 Literature review

This section studies the methods available in this field. The attention is divided into two groups: Context-Aware RSs and SPM algorithms in sequence-aware RSs.

2.1 Background on the context-aware recommender systems

Across several disciplines, over 150 different meanings of the term "context" have been presented [2]. 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 [21].

A survey of pertinent papers was conducted to that goal. Memon et al. [26] 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. However, the study found that relying solely on collaborative filtering led to more precise suggestions. This finding prompts a crucial question: How can we strike a balance between collaborative filtering and context-aware techniques to enhance recommendation accuracy?

Sun et al. [37] 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. This prompts us to delve deeper into addressing the "cold start" challenge, an aspect often overlooked in existing approaches. How can we effectively utilize demographic information to mitigate the cold start problem without compromising recommendation accuracy? While these studies contribute valuable insights, our proposed hybrid method aims to address some of the limitations observed in the existing literature.

Our prior paper [28] described a tourist recommendation system that considers the target phone's position and suggests the best tourist accommodations rely on contextual data and trust measures. Studies and assessments revealed that introducing extra context improved the proposed method's results.

2.2 Sequential pattern mining algorithm in sequence-aware recommender system

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. Any sequence-aware recommender system may be categorized into four types based on application scenarios: adaptation, trend detection, repeated recommendation, and sequential patterns [57].

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 complex 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 some disciplines.

Some algorithms proposed for identifying sequential patterns are GSP, Free-Span, Prefix-Span, and SPADE [41, 50, 51].

In reference [23], a personalized route suggestion strategy was proposed by Lim et al. to address challenges related to trip orientation. While the authors tackled issues such as time limitations and the requirement for trips to start and end at specific points of interest, their approach fell short in incorporating essential variables, such as preferred trip periods and travel categories. The strategy considered both POI popularity and user interests in suggesting acceptable POIs, along with the allocation of time at each location. Notably, their model included automated detection of real trip sequences and estimation of POI popularity and user interest using geo-tagged images. However, the omission of certain critical tourist variables raises questions about the comprehensiveness of their approach. How can we integrate a broader range of tourist preferences, including trip duration and specific travel categories, to enhance the personalization of recommendations?

In another study by Bin et al. [7], a POI trip recommender structure was presented, leveraging Sequential Pattern Mining and various tourist data sources. Their organized approach aimed to construct a comprehensive foundation of POI knowledge and a substantial framework of POI patterns. While their method showed promise in utilizing diverse data sources, how does it compare to our proposed hybrid approach, which integrates collaborative filtering, DB information, and context-aware techniques? Additionally, does their approach adequately address the challenges of data sparsity and the cold start problem, which our methodology aims to mitigate?

By critically examining these works, we set the stage for a more nuanced understanding of the existing landscape and emphasize the distinctive contributions of our proposed recommendation system.

Table 1 provides an overview of tourist recommender system methodologies, including methods, descriptions, and comments.

In contrast to prior studies, our current research uniquely integrates contextual, geotagged, and user demographic data to construct substantial structured POI visit sequences. While preceding approaches often generate lists of individual POIs for tourists, our method efficiently extracts POI routes from visit sequences across diverse tourism contexts using a selective parameter. Notably, unlike methods focusing solely on the top-k POIs, our proposed approach considers additional factors such as POI popularity, time, and weather conditions, recommending a connected route that satisfies these criteria rather than individual POIs. Another key distinction is our emphasis on high-level tourism personalization in POI routes, as opposed to prior works suggesting routes where all users visit the same non-personalized POI. Additionally, our approach places a greater emphasis on user preference modeling, setting it apart from many related works. Moreover, extensive experiments conducted on two real-world tourism datasets affirm the feasibility and effectiveness of the proposed system. Finally, our route retrieval approach generates POI route recommendations by comprehensively considering various tourism contexts, including the POI's location, weather, temperature, season of visit, and day of visit. Consequently, the resulting POI route accommodates tourists' personal constraints while maintaining a notably high tourism value.

3 Proposed approach

Our method introduces a hybrid trip RS tailored for the tourist industry, utilizing a combination of tourist demographic, contextual, and geo-tagged data to suggest a curated list of places in a town. This framework is delineated into two phases, as illustrated in Fig. 1 (offline and online). In the offline phase, contextual data is derived from image timestamps and climate networks following preprocessing (3.2.1). This data is integrated into the dataset to enhance completeness (3.2.2). Additionally, clustering methods are applied to establish regions of interest, with each zone containing one or more POIs based on the geographic coordinates of the images.

In the subsequent step (3.2.3), places are identified by repeating the clustering technique on the Area of Interest (AoI) results. Then, for each of these POIs (denoted as L), a profile is created by calculating their publicity and situational characteristics (3.2.4). User-POI (3.2.5) and User-User (3.2.6) relationships are determined based on prior visits, each having its respective database.

To find prior user journeys, POI sequences are established using each user's visit time (3.2.7), and the Prefix-Span technique is employed to locate relevant POI sequences (3.2.8). These sequences are stored for use in the subsequent phase, optimizing processing speed.

In the online phase, tourists register by providing demographic data (age, sex, town, nation, relationship, and profession). When a tourist queries the system, the question includes context data such as geographic location (current place) and climate data for the travel dates (3.3.1). Contextual pre-filtering is then employed to select POIs in that place based on the user's current location (3.3.2).

The subsequent stage calculates the similarity between the current tourist and those who visited the filtered POIs (L') (3.3.3), utilizing both CF and DB approaches. Tourists more similar to the current one are selected, and POIs are suggested based on the top-ranking

Table 1 Presents an overview of approaches to tourist RSs

Methods	References	Description	Remarks
CA + CF	[26]	- Tourism RS -Used tourist preferences	-Enhanced accuracy -No asymmetric measure
CF + CA + sequence	[23]	- Trip recommendation -Supported user interests and location -Personalized	-Considering sequences -Considering the duration of visits and recent visits
CA + CF + sequence	[24]	- Travel location recommendation - Supported Geo-Tagged photos -Using Sequences - Personalized	- Context-Aware -Managing Data Sparsity and Cold Start issues
CF + CA + SPM	[25]	-A sequence-aware RS using Geo-Tagged Photos	- Context-Aware -Using Prefix-Span - No Personalized RS
CA + CF + sequence	[8]	- Supported geo-tagged photos - Supported Semantic Model	- Context-Aware -Prefix-Span - Using User Preferences -Not suitable for Cold Start issue
CF + CA	[32]	-Hybrid location-based travel RS - Contextual information -Considering the user's preferences	-Personalized -Managing Data Sparsity and Cold Start issues
DB + CA + CF	[5]	- A Multi-level tourism RS - Context-Aware	- Supported Demographic info (age, gender, country) -Not suitable for the Cold Start issue
CF + CA	[9]	- Itinerary RS - Using location text contents -Using historical tourist visits - Based on categories, the locations	- Personalization - context-based -Managing Data Sparsity and Cold Start issues
CF + CA + Semantic similarity	[1]	-A context-aware RS - Using Social Networks data -Semantically clustering	-Personalized -Not suitable for the Cold Start issue

Table 1 (continued)

Methods	References	Description	Remarks
CA + CF + SPM	[51]	-A hybrid context-awareness RS -Supported SPM and CF for e-learning	-Not suitable for the Cold Start issue - Personalized -Used GSP Algorithm for SPM
Subsequence Patterns POIs + ASPPA	[48]	-Find POI sequential patterns -Personalized	-Managing Data Sparsity and Cold Start issues - Contextual(spatial and temporal)
Sequential Mining + CNN + LFM	[38]	-A sequential RS - Using the Convolutional Neural Network (CNN) -Using the Latent Factor Model (LFM)	-Managing Data Sparsity and Cold Start issues - Personalized - No Contextual
CA + CF + SPM(GSP)	[52]	-Hybrid RS model for e-learning - Personalized -Supported ontology domain knowledge	- Contextual -Used GSP Algorithm for SPM -Managing Data Sparsity and Cold Start issues
CA + SPM	[7]	-A personalized route recommendation -Using contexts (travel duration, companion type, season, POI types) -Used Prefix-span Algorithm for SPM	- Supported the travelogues and POI tourism attributes -Contextual -Managing Data Sparsity and Cold Start issues
CA + SPM	[16]	-Personalized sequential pattern -Sequential-labeled Decision Tree -Predicted Patterns	-Contextual - Personalized -Used Classification -Managing Data Sparsity and Cold Start issues
CA + CF + DB	[19]	-Tourist RS -Asymmetric similarity measure -Clustering	-Contextual - Personalized -Geo-tagged photo

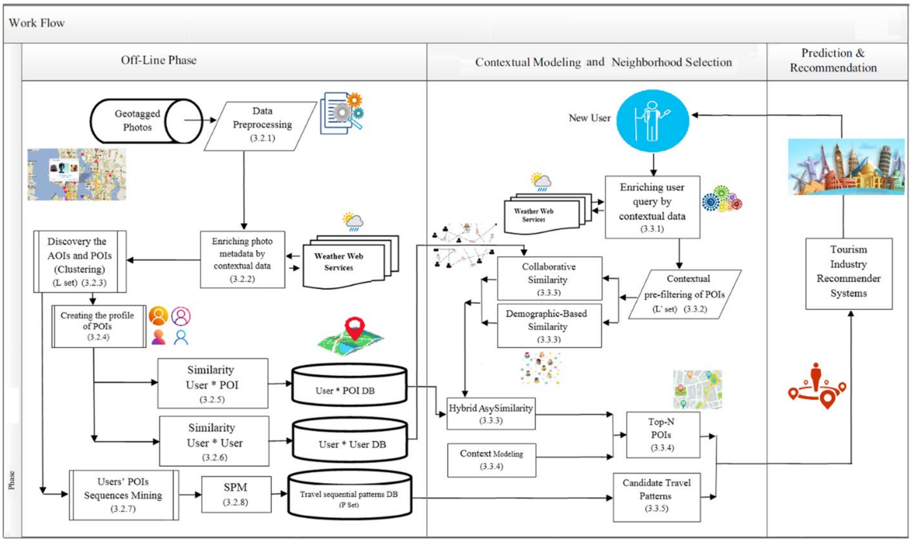

Fig. 1 The process of our method

neighbors' preferences, accounting for spatial proximity and present climate (Contextual modeling) (3.3.4).

For the nominee travel patterns phase (3.3.5), we consider the derived Top-N places (3.3.4) and the explored and saved sequential trip patterns (P set) (3.2.8). The pattern ($p \in P$) score is computed for each tourist, representing the sum of ranks of the POIs stored within the pattern using the rank function mentioned in 3.3.5. The Top-N trip sequential patterns are considered as applicant journey patterns, and the target tourist is provided 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.

3.1 Problem identification

The following is a definition of the challenge of proposing exciting tourist destinations and POI sequences in geo-tagged social networking sites: $P = P_1, \ldots, P_n$ 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' current context.

3.2 Offline phase

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

3.2.1 Data preprocessing

Firstly, the dataset 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's 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.

3.2.2 Enriching geotagged photo with contextual information

The contextual data for the dataset's image items were created and stored in this step. This data, including the time and geographic place, is 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.

3.2.3 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 [47] and the demand for little scope information to determine the parameters. It also does an excellent job of grouping vast amounts of data. The thickness point for clusters in the DBSCAN is the same. Two factors are essential: 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

Table 2 Contextual 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, May	Spring
		June, July, August	Summer
		September, October, November	Fall
		December, January, February	Winter
Weather Context	Temperature	>34℃	Hot
		18℃-34℃	Warm
		<18℃	Cold
	Weather	Sunny, Clear Sky	Sunny
		Cloudy, Broken, Clouds, Scattered Clouds	Cloudy
		Rain, Fog	Rainy
		Snow, Snowfall	Snowy

rerun with the proper settings on the observed AOIs. As a consequence, a collection of POIs was created, as well as a database of essential tourist places (L).

3.2.4 Producing the profile of 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.

$$PlacePopulartity(\mathrm{POI}) = \log(\frac{N}{N_l}) \tag{1}$$

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

A novel weighting context vector structure is defined in this work as $\vec{C}_{POI} = < c_{(POI,1)}, \ldots, c_{(POI,k)} >$, where $c_{(POI,j)}$ indicates the context (j) of each 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) [33].

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

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 a journey in any context from all POIs in the present city, and $w_{(POI,0)}$ indicates the number of a trip in any context from the POIs.

3.2.5 Measuring the similarity between users and POI

In this step, the information format technique has been used to organize information about users' previous trips to improve the effectiveness of generating and introducing visit suggestions to tourists. This method was built with the help of a matrix known as the User-POI-Matrix. Every one of its components represents the number of pictures taken by the tourist at a location in the town. This quantity is inferred as the user's implicit location rating.

3.2.6 Measuring the similarity between users

The similarity among tourists who had visited the places before was evaluated and stored for use during the online phase. The experiences of other travelers were leveraged to create suggestions for the intended consumer. Equation four, a revised prescription of Sorensen's formula (Eq. (3)), was used to accomplish this. The goal would have been to estimate tourist similarities relying on the number of photographs they saw and to develop a rational link between their attractiveness.

The suggested method is based on the concept that when two users take images of places that are less popular and visited by fewer people, this similarity is much higher than that of the other. Two visitors who visit a minor part of a town are more interested in sharing similar concerns. Consequently, while computing similarity, the suggested similarity metric tries to account for individuals' preferences for commonly visited sites [31].

$$Sim(u, \mathrm{v}) = \frac{2 * |l_u \cap l_v|}{|l_u| + |l_v|} \tag{3}$$

$$Sim_{CF}(\mathrm{u}, \mathrm{v}) = \frac{2 * \sum_{l \in l_u \cap l_v} PlacePopulartity(\mathrm{l})}{\sum_{l \in l_u} PlacePopulartity(\mathrm{l}) + \sum_{l \in l_v} PlacePopulartity(\mathrm{l})} \tag{4}$$

Here, $\mathrm{l_u}$ represents the places visited by user u, and PlacePopulartity(l) indicates the popularity of a place (l). The expression $|l_u \cap l_v|$ signifies the count of locations visited by both user u and user v.

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. In fact, these techniques are founded on the premise that sim(u,v) equal sim(v,u) while the impact of two different users on one another differs; therefore, the asymmetric schema is given to traditional similarities in CF approaches to create a highly realistic similarity [30, 46]. This work uses asymmetric schema to bypass the limitation. The term "Asymmetric CF (ACF)" implies that it's a variant of CF where the similarity measure between users is not symmetrical.

The rate of similarity places among tourists, adjusted by the figures of places assessed by the present tourist, Eq. (5), is used to create the asymmetric similarity measure.

$$sim_{Asy-Measure}(u, v) = (1 - \exp(-|l_u \cap l_v|)/|l_u|)) \tag{5}$$

Here, $\mathrm{l_u}$ represents the number of POIs visited by u. The expression $|l_u \cap l_v|$ signifies the count of locations visited by both user u and user v. This equivalence looks at the proportion of standard ratings that users have among all of their rated things rather than the proportion of standard ratings in the total number of ratings among tourists. As a result, Eq. (4) contains this parameter.

$$Sim_{AsyCF}(\mathrm{u}, v) = Sim_{Asy-Measure}(\mathrm{u}, v) * Sim_{CF}(\mathrm{u}, v) \tag{6}$$

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

$$Sim_{UPP}(\mathrm{u}, v) = \frac{e^{-(|r_{u,p} - r_{med}| * |r_{v,p} - r_{med}|)}}{[1 + e^{-(|r_{u,p} - r_{med}| * |r_{v,p} - r_{med}|)}]^2} \tag{7}$$

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

The expression $[1 + e^{-(|r_{u,p} - r_{med}| * |r_{v,p} - r_{med}|)}]^2$ represents a mathematical computation involving the ratings of a particular place p by two users u and v. The expression calculates the squared value of the sum of 1 and the exponential function e raised to the power of the negative product of the absolute differences between the ratings of each user and the median rating.

By integrating Eqs. (6) and (7), we may get at the new formalization, which we term enhanced new CF asymmetric similarity model (AsyNCF) (Eq. (8)). 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 actual 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 [56].

$$Sim_{AsyNCF}(\mathrm{u}, v) = Sim_{UPP}(\mathrm{u}, v) * Sim_{AsyCF}(\mathrm{u}, v) \tag{8}$$

3.2.7 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 [23]. 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.

3.2.8 SPM method

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 standard item-set models in databases. The Prefix-Span method is a simple algorithm that explores the entire collection of patterns [6, 41]. 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.

3.3 Online phase

The following steps are included in this stage. In fact, our method answers the target user's request quickly and interactively.

3.3.1 Enriching user queries by contextual data

The system calculated the time requested by the user to visit as the user 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 (V_u) is built. When a context criterion is satisfied, it is given a value of one; otherwise, it is given a zero value.

3.3.2 Pre-filtering based on context

The data for that city was 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').

3.3.3 Combination of the recommendations

The hybrid similarity is calculated in this step using Eq. (9). This equation illustrates the similarity between collaborative filtering and demographic-based filtering in terms of the target user and those who visited the set (L') of locations. This phase uses Eq. (9) to compute the hybrid similarity. In terms of the target user and those who visited the set (L') of places, this equation highlights the similarities between CF and DB.

$$Sim_{AsyHybrid}(\mathrm{u}, v) = (1 - \upbeta) * Sim_{DB}(\mathrm{u}, v) + (\upbeta)Sim_{AsyNCF}(\mathrm{u}, v) \tag{9}$$

This compound is balanced using the coefficient (β) to smooth out the linear connection [40].

Equation (10) was used to estimate the demographic similarities among the two tourists [17, 56].

$$Sim_{DB}(\mathrm{u}, v) = \frac{|\mathrm{num}_1\left(\vec{D}_u \cap \vec{D}_v\right)|}{|DemograpicFeatureVector|} * 1/(1 + \frac{|\mathrm{age}_u - \mathrm{age}_v|}{\max(age) - \min(age)}) \tag{10}$$

For each user, a demographic characteristic (excluding age) vector such as $(\vec{D}_u)$ is created. The first user demographic characteristic vector is compared to the second user demographic information vector when comparing users based on their demographic features. Suppose two users have the same value for a particular property, such as sex; the value of one is utilized. Using the $\mathrm{num}_1(D_u \cap D_v)$ function, the number of units in the two users' common feature vectors is tallied and divided by the number of demographic features examined by the users. The output result of this similarity is always between zero and one.

Given the importance we place on the aged character, we utilized the tourist age as a distance attribute model where age_u and 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 more excellent similarity score to the present user.

3.3.4 Recommendation

The level of intention of the present tourist to visit destinations can be determined by using Eq. (11) based on the similarity among tourists.

$$Pred(\mathrm{u}, \mathrm{l}) = \frac{\sum_{v \in U'} Sim_{WT-context}\left(C_u, C_1\right) * Sim_{loc-context}\left(l_u, l_1\right) * Sim_{AsyHybrid}(\mathrm{u}, v) * \left(r_{v_l}\right)}{\sum_{v \in U'} Sim_{AsyHybrid}(\mathrm{u}, v)} \tag{11}$$

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

The context factor $Sim_{loc-context}(l_u, l_l)$ 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 [49]. The Manhattan formula was used to get the distance factor for the site (Eq. (12)).

$$Distance_{Geo\left(l_u, l_{l_i}\right)} = \left(|x_1 - x_2|\right) + (|y_1 - y_2|) \tag{12}$$

where we have the target user's geographical location $l_u(x_1, y_1)$ and the tourist location l_{l_i}, *to* cover all points and achieve the closeness of distance, we utilize the double Laplace distribution equation (Eq. (13)) [14].

$$Sim_{Loc-Context}(l_u, l_l) = \frac{1}{2\mu} * e^{\left(-\left|Distance_{Geo(l_u, l_l)}\right|/\mu\right)} \tag{13}$$

The μ 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 ($Sim_{WT-context}(C_u, C_{l_i})$). 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 was applied to the existing user query in the pattern of a vector (V_u) in compliance with Section 3.3.1, and on the other hand, having the vector template of the contextual metrics weight of the POIs (W_l) enables the determination of similarity via adjusted cosine formula (Eq. (14)) [47].

$$Sim_{WT-context}\left(\vec{V}_u, \vec{W}_l\right) = \frac{\vec{V}_u * \vec{W}_l}{\left||\vec{V}_u|\right| * \left||\vec{W}_l|\right|} \tag{14}$$

For the existing user context and every Location context, a context vector modeling such as (C_u) and (C_l) is created. The output data of this similarity is always between nil and one. The list items are sorted according to the projected points for each site, which are related tourist spots.

3.3.5 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). Pred(u,l) was determined in the stage before (Eq. 11).

$$T - Score\left(\mathrm{Travel_{Seq}}\right) = W_{Time-Seq}\left(T_u, T_{Seq}\right) * \left(\frac{1}{n} * \sum_{i=1 \& l_i \in \mathrm{Travel\_Seq}}^{n} Pred\left(\mathrm{u}, l_i\right)\right) \tag{15}$$

Each travel sequence's number of places is indicated by (n). $W_{Time-Seq}(T_u, T_{Seq})$ is used as a weight in Eq. (15) 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 valuable 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).

$$W_{Time-Seq}\left(T_u, T_{Seq}\right) = \lambda * e^{\frac{-Ln2*\lambda*|T_u - T_{Seq}|}{hL_u}} \tag{16}$$

Here, $W_{Time-Seq}\left(T_u, T_{Seq}\right)$ signifies the time weight reflecting how much a user's interest has decreased; T_u indicates the current time, and T_{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, in fact, governs the pace with which we forget [9, 47]. In terms of days, the suggested technique accounts for the time–space between these two times. hL_u may be considered as fifteen days, considering that every journey takes an average of one month. The decay rate is adjusted using the time decay factor (λ). In this situation, we state that λ 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, and SPM, a customized trip is recommended to the present tourist.

4 Simulations and experimental evaluation

In this part, many tests were conducted to show the effectiveness of the suggested strategy. For that purpose, we'll 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 is evaluated and compared to findings acquired utilizing other cutting-edge approaches, such as Flickr and Yelp Data sources.

4.1 Evaluation dataset

This study employed Flickr, one of the most popular image-uploading social networks. As a photo-based social media platform, this website has gained popularity. Viewing and sharing Flickr photographs and videos does not require a Flickr account; however, sharing data does. YFCC100M Yahoo was the Flickr dataset utilized [12, 39]. This dataset is hosted at Webscope Yahoo Labs (2022). The suggested technique was tested utilizing Flickr, which includes image information.

The Application Programming Interface (API) 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 Section 3.2.3. 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's 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 2 and 3 show how the number of recognized clusters changes as the values of these two parameters alter.

The minimum sample size is a falling graph pattern when the radius reaches 120 for all attribute values, as seen in Fig. 2. For the parameter specifying the minimum sample sizes, the chart's declining slope is decreased to a parameter of 10 in Fig. 3. 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, this data was separated into two non-overlapping halves, with 75 percent utilized for framework training and 25 percent for analyzing process.

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).

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

The second dataset used was the Yelp dataset, which was discovered in several regions of the US (Las Vegas and Nevada). This dataset is a public multimedia collection. It is accessible via "https://www.yelp.com/dataset". After preprocessing, this dataset of 18,534 entries was collected from the transaction information of 19,800 users. These data include the Business ID, Name, geo-location, and annotation tags, among other things.

4.2 The evaluation metrics

The proposed method has been evaluated based on some criteria. The Recall measure is presented as the ratio of accurate things offered to the whole number of relevant objects for the target user (Eq. 17). [3].

$$\text{Recall} = \frac{\#(\text{Number of Accurate predictions})}{\#(\text{Number of Relevant Objects})} \tag{17}$$

As demonstrated in Eq. 18, Precision is presented below.

$$\text{Precision} = \frac{\#(\text{Number of Accurate Predictions})}{\#(\text{Number of Total Predictions})} \tag{18}$$

The Average Precision measure is presented as Eq. (19), which computes Precision for all users.

Equation (19) determines the Average Precision metric AP@N, an equation that calculates accuracy for all users.

$$\text{AP@N} = \frac{\sum_{k=1}^{N} (\text{Precision@k} * \text{Relevant}_k)}{\text{M}} \tag{19}$$

(M) is the relevant item, and (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 (20) describes The Mean Average Precision metric for (m) users [53].

Table 3 Fields of the Flickr Dataset

ID	Owner	Title	Time-Taken	Tags	Latitude	Longitude
47922891637	61013449@N07	British Museum	4/29/2018 15:50	British Museum, London, sculpture	51.51928	-0.12705

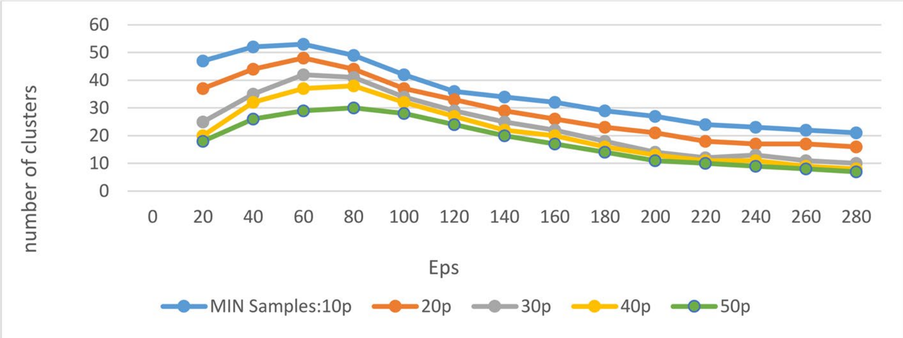

Fig. 2 Depending on the number of clusters discovered according to the Eps

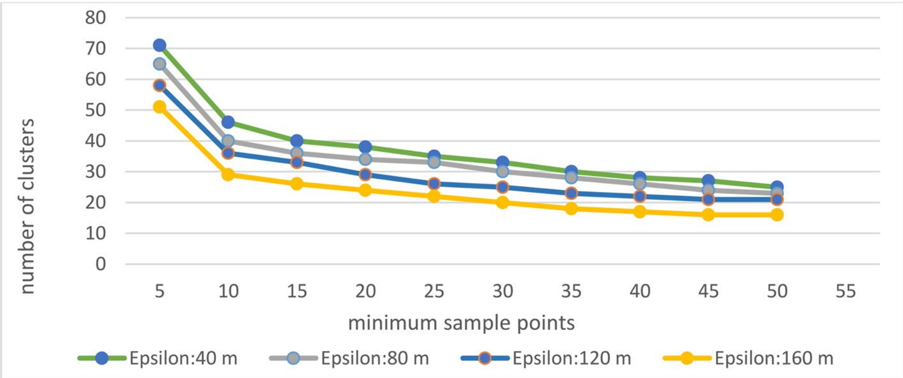

Fig. 3 Depending on the number of clusters discovered according to MinPts

Table 4 Flickr Description

Photos		Users	Location
Raw	Filtered	Filtered	
49,999	44,265	458	2960

$$\text{MAP@m} = \frac{\sum_{k=1}^{m} AP(u)}{m} \tag{20}$$

RMSE (Root Mean Square Error) highlights bigger absolute error levels (Eq. (21)) [54].

$$\text{RMSE} = \sqrt{\frac{\sum_{(\mathrm{u,i})|\mathrm{R}_{\mathrm{u,i}}} \left(\hat{r}_{u,i} - \mathrm{r}_{\mathrm{u,i}}\right)^2}{\mathrm{N}}} \tag{21}$$

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

The F-Score is identified as Eq. (22) [55].

$$\mathrm{F-Score}=\frac{2*\#(\mathrm{Precision}*\mathrm{Recall})}{\#(\mathrm{Precision}+\mathrm{Recall})} \tag{22}$$

The projected suggestions' ranking efficiency is compared utilizing the Normalized Discounted Cumulative Gain (NDCG) (Eq. (23)). The more related topics of attention shown at the head of the proposed list, the higher the NDCG score [35, 45].

$$\mathrm{NDCG}=\frac{DCG}{iDCG}=\frac{\sum_{i=1}^{p}\frac{2^{rel_{i-1}}}{log_2^{(i+1)}}}{\sum_{i=1}^{|rel_p|}\frac{2^{rel_{i-1}}}{log_2^{(i+1)}}} \tag{23}$$

Here, rel_i stands for the element ranked at the place (i), and rel_p stands for the list of related items in the relevant group in position (p).

4.3 Comparison approaches

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

4.4 Experimental results

This section initially explores the effect of CF and DB on suggestion accuracy. Subsequently, we delve into an examination of how neighborhood numbers influence the quality of suggestions.

Following these analyses, the ensuing sections provide a comparative assessment of the effectiveness of our NewSeqHybrid approach against earlier methods addressing Cold Start and Data Sparsity challenges. This evaluation is based on a range of criteria, including MAP, Precision, Recall, RMSE, F-Score, and NDCG.

Table 5 Comparison Methods

Methods	#References	Description
(CF)	[30]	Cosine similarity
(PR)	[25]	Public Popularity
(Pre_CA-CF)	[18]	Contextual Pre-Filtering Similarity Measure
(CA-CF)	[22]	Jaccard Measure
(ACA-CF)	[31]	Cosine similarity + Asymmetric Schema with Jaccard Measure
(GSP-CACF)	[52]	CF and GSP
(CA-MSDT)	[16]	CA and Decision Tree Classification
(Prefix-CSTR)	[25]	Prefix-Span Algorithm
(ADBCACF)	[19]	Asymmetric CF and Demographic Data
(SeqHybrid)	[15]	Our previous work describes a sequential recommender system that combines context awareness, demographic-based, and asymmetric CF

4.4.1 Impact of parameter β

The significance of each DB and CF similarity in the context of two tourists is determined by the variable (β) in Eq. (9). Consequently, the choice of the value for this factor significantly influences the performance of the current plan. Figure 4 presents the outcomes of trials focusing on the F-score to identify the optimal value for the parameter (β). Notably, the chart featuring a β value of 0.6 outperformed other variations of β, indicating its superiority. Thus, initiating this factor with a value of 0.6 is considered a judicious decision. This underscores that the influence of neighbors on tourism suggestions takes precedence over demographic data.

4.4.2 The effect of the neighborhood numbers

By proposing 2, 6, and 12 POIs among all the president's suggestions, we studied and applied the influence of neighbor size on the accuracy of anticipated suggestions for neighbors of different sizes. In this scenario, the number of neighbors grew between two and eighteen. Concerning MAP, the findings are shown in Fig. 5; when the number of suggestions surpassed six, adding POIs dramatically lowered the validity of the recommendation results when using MAP, according to these studies.

Users' preferences 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, NewSeqHybrid exhibited a greater MAP score than previous approaches. According to the findings, asymmetric strategies outperform symmetric techniques. Furthermore, the GspCACF, ADBCACF, SeqHybrid, and NewSeqHybrid recommender systems combine sequential and non-sequential information. 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 exhibited an increase as the number of suggestions was elevated, as depicted in Fig. 6. This growth is attributed to the inclusion of more precise POIs in the

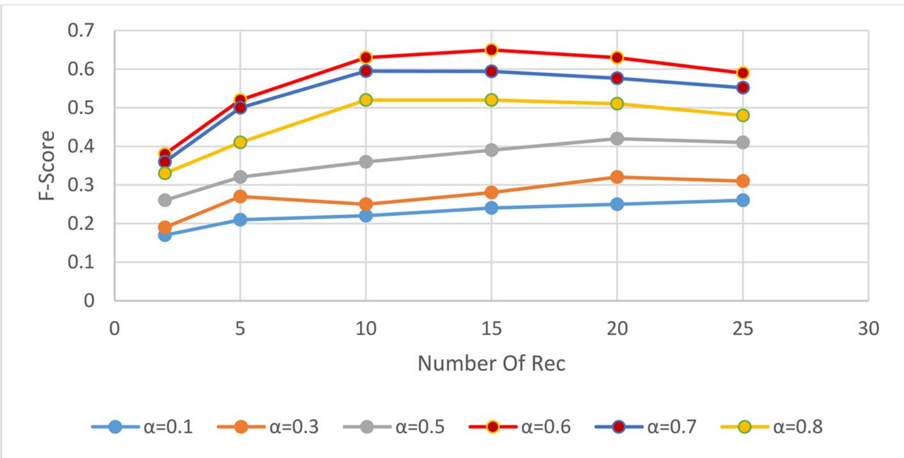

Fig. 4 The impact of DB and CF in terms of F-Score

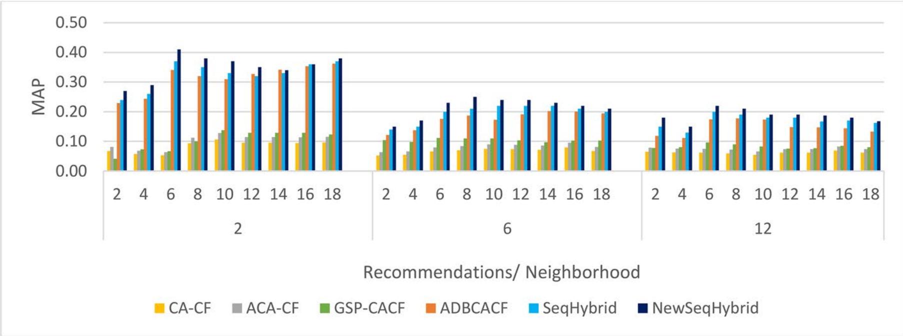

Fig. 5 The list of recommendations is used to compare NewSeqHybrid to others with regard to the MAP metric (Flickr)

Top-N suggestions. Consequently, in terms of Recall, NewSeqHybrid demonstrated superior performance compared to prior methods. The results indicate that asymmetric strategies outperformed alternative approaches. Notably, the CF and PR approaches produced the least accurate findings in comparison to other methods, which can be attributed to the absence of clustering and a disregard for context.

4.4.3 The impact of the highest suggestions

As seen in Fig. 7, as the number of suggestions rose, the accuracy dropped. The fact 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 despite the fact that the suggested technique beats the other methods in terms of accuracy 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.

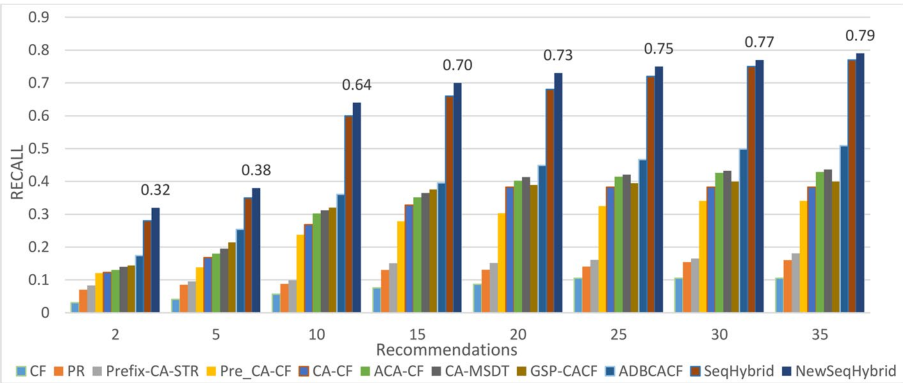

Fig. 6 The list of recommendations is used to compare NewSeqHybrid to others with regard to the Recall metric (Flickr)

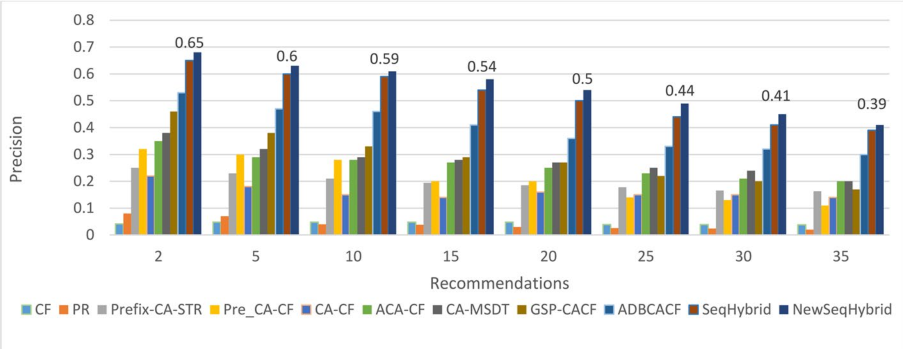

Fig. 7 The list of recommendations is used to compare NewSeqHybrid to others with regard to the Precision metric (Flickr)

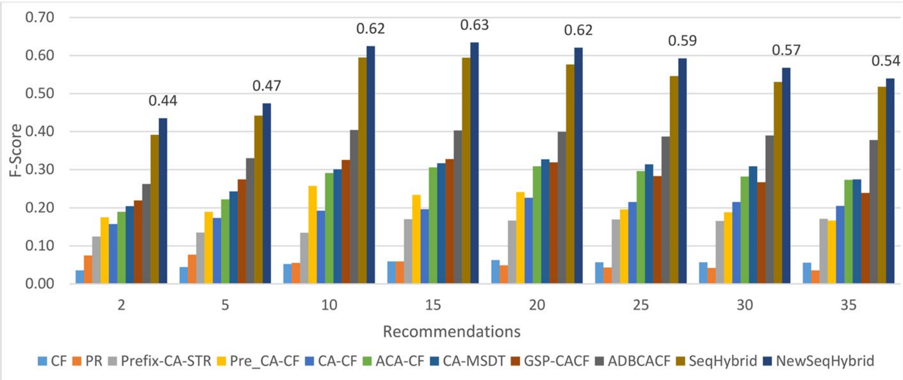

Fig. 8 The list of recommendations is used to compare NewSeqHybrid to others with regard to the F-Score metric (Flickr)

For numerous suggestions based on F-Score measures, NewSeqHybrid outperformed other techniques, as shown in Fig. 8. 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, NewSeqHybrid was able to produce improved outcomes. Furthermore, using demographic data on users might help forecast user preferences for future visits, thus alleviating the Cold Start problem. Instead of depending on a single site to discover POIs, a clustering approach can assist in easing the data sparsity problem. This framework suggestion may also be customized by integrating SPM with the Top-N POIs.

The RMSE measure calculates the difference between expected and actual ratings, serving as a common metric in recommender systems to gauge the variance between real and predicted ratings for items. As illustrated in Fig. 9, non-context-aware procedures typically exhibit a higher error rate compared to context-aware approaches. Remarkably, our proposed method outperformed previous contextual approaches, showcasing a lower error rate than non-contextual methods. Notably, NewSeqHybrid demonstrated superior Cold Start issue handling compared to alternative techniques, attributed to the inclusion of demographic data. On average, the improvement of NewSeqHybrid over the listed methods is approximately 32%.

4.4.4 Evaluation of NewSeqHybrid with the Yelp

In this study, we employed the Yelp data source as an additional dataset to evaluate the performance of NewSeqHybrid. The evaluation results, depicted in Figs. 10 and 11 and based on Recall and Precision measures, reveal the enhanced outcomes achieved by NewSeqHybrid. The incorporation of demographic data, contextual information, and an asymmetric schema contributed to these improved results. Notably, the proposed solution demonstrated significant success in mitigating the Cold Start issue compared to previous alternatives. A comparative analysis between the two datasets, Yelp and Flickr, revealed that Yelp exhibited lower Precision and Recall, possibly due to its fewer demographic characteristics. It is worth noting that the data volume in this source, along with the number of neighbors, significantly influences the observed outcomes.

4.4.5 Evaluation of NewSeqHybrid by NDCG metric

In assessing NewSeqHybrid with the NDCG metric, a key indicator of ranking efficiency in sequential approaches, notable improvements were identified. The NDCG measure, depicted in Fig. 12, showcases the heightened rating proficiency achieved by NewSeqHybrid in its sequence suggestion strategy. Notably, the substantial improvement from 0.63 to 0.66 in NDCG underscores the superiority of NewSeqHybrid, affirming its capability to project more fitting recommendations compared to prior approaches. In other words, the improvement of NewSeqHybrid over SeqHybrid is approximately 5.6% for TOP-2 and 5.1% for TOP-5.

The current method stands apart from its predecessors by integrating contextual, geotagged, and user demographic data to construct comprehensive and well-structured POI visit sequences. In contrast to previous approaches that merely compile lists of individual

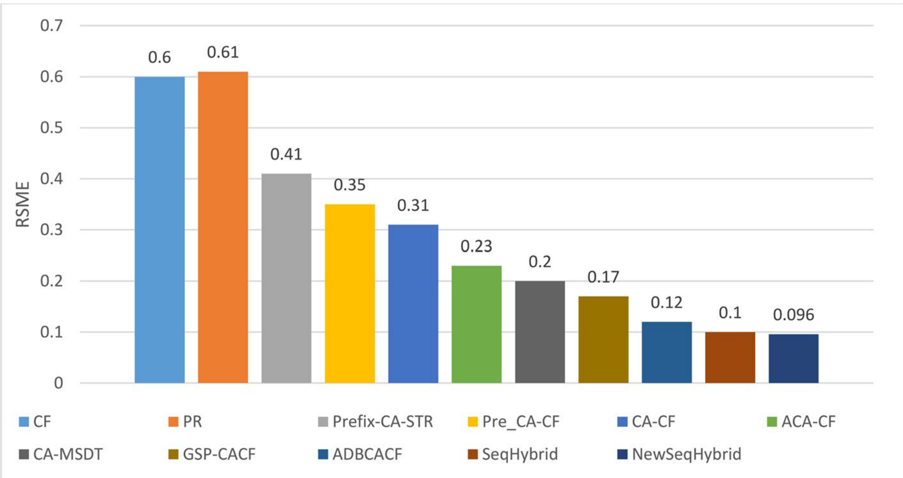

Fig. 9 NewSeqHybrid is compared to others according to the RMSE measure (Flickr)

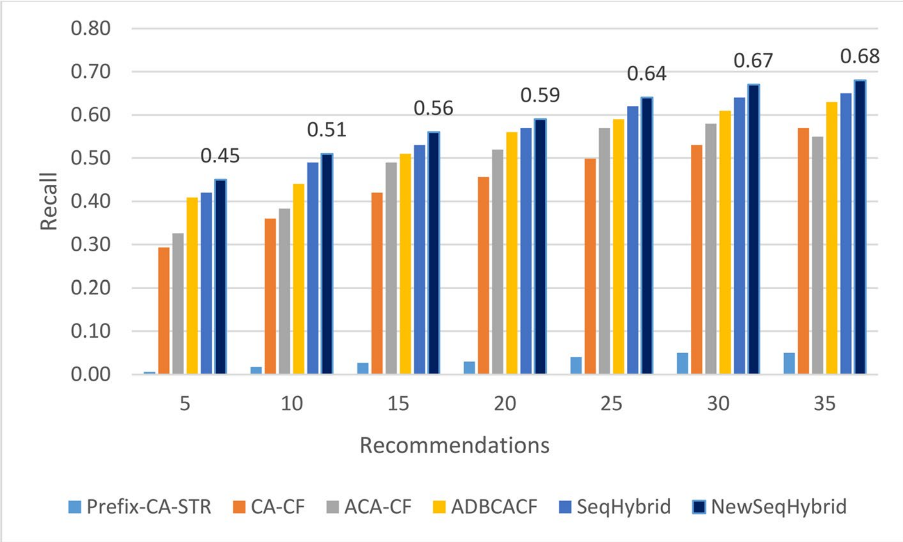

Fig. 10 NewSeqHybrid is compared to others according to the Recall measure based on the number of recommendations (Yelp)

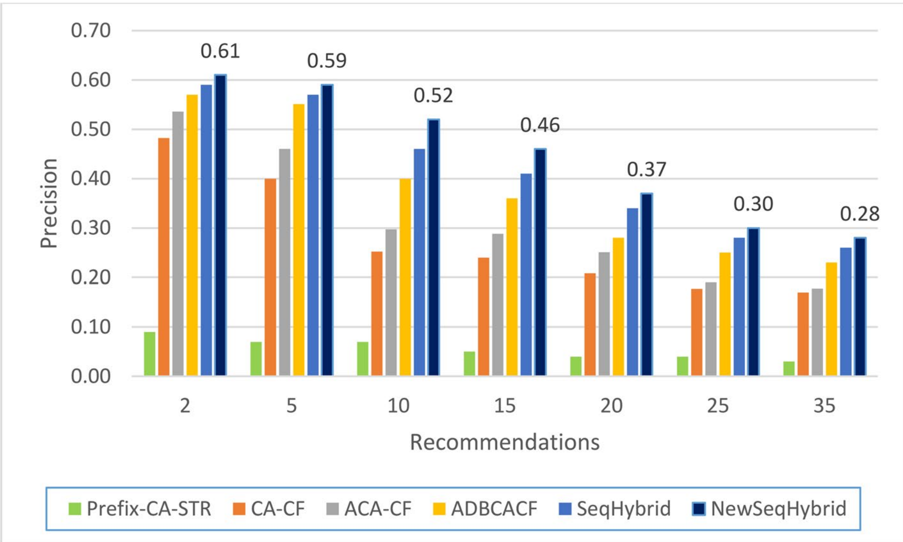

Fig. 11 NewSeqHybrid is compared to others according to the Precision measure based on the number of recommendations (Yelp)

POIs for tourists, our methodology efficiently extracts POI routes within diverse tourism contexts using a refined selective parameter. Notably, while existing methods typically concentrate on identifying Top-k POIs, our proposed solution incorporates additional criteria such as POI popularity, temporal considerations, and weather conditions. Rather than recommending individual POIs, our approach suggests a connected route that aligns with

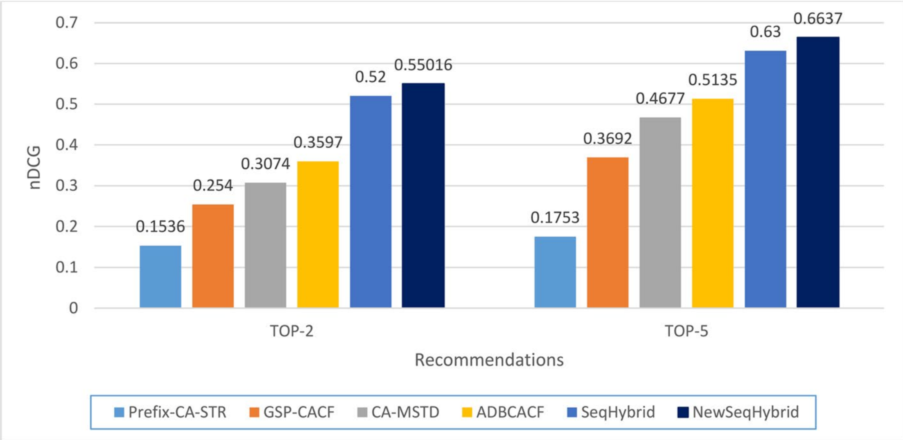

Fig. 12 NewSeqHybrid is compared to others according to the NDCG measure

these specified criteria. Another noteworthy deviation from prior works is our emphasis on high-level tourism personalization in POI route recommendations, in contrast to previous models suggesting routes that cater to all users with non-personalized POIs. Moreover, user preference modeling, often given lesser attention in related works, is a focal point in our study.

Furthermore, the effectiveness and feasibility of the proposed system are validated through extensive experiments conducted on two real-world tourism datasets. Finally, a route retrieval approach is employed to generate POI route recommendations, taking into account various tourism contexts, including POI location, weather conditions, temperature, season of visit, and the day of visit. This ensures that the resultant POI route accommodates the individual constraints of tourists while maintaining a notably high tourism value.

4.4.6 Case study

In this case study, we explore the scenario of a visitor navigating a city. When a recommendation is requested, the tourist's real-time location is dynamically determined through their cellphone data. The potential locations include both the city's AOIs and POIs) Given the abundance of POIs in each city and the impracticality for a tourist to visit them all, coupled with the assumption of the tourist lacking a sufficient history of POI visits in the city (cold start problem), their nearest neighbors are identified based on current contexts and preferences. Subsequently, visitors receive recommendations for the top-N POI routes based on their activities and information about nearby neighbors. This investigation considers all data obtained from tourist visits, dynamically adjusting preferences based on the time and place of the targeted tourist. Thus, in scenarios involving a cold start and limited information, we leverage all available data to provide accurate suggestions without direct visitor participation. During the case study, additional trials were conducted in the New York metropolitan region to identify routes for active visitors, as illustrated in Table 6.

When an active tourist seeks a recommendation, similar neighbors are identified based on their current settings, histories, and attributes. The top-N recommendations are then selected and presented. Performance evaluations of this stage are depicted in Fig. 13. According to the results, our proposed technique demonstrates the highest F-score,

Table 6 Description of the Flickr dataset used in our case study

City	#Images	#Users	#POIs
New York	1,510,000	9,387	49,400

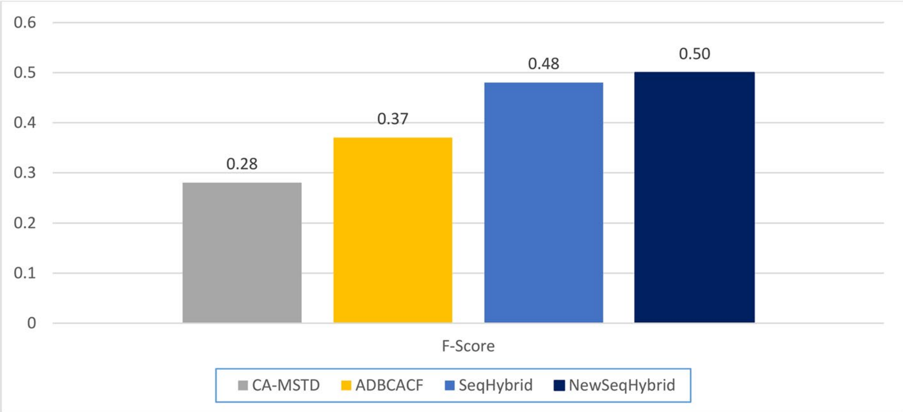

Fig. 13 Performance results for several methods

particularly based on an asymmetric topic CF scheme. While alternative approaches rank slightly lower, the distinction between the methods is relatively modest, indicating that tourists in this city exhibit tastes and knowledge that are more similar.

5 Conclusions

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 SPM algorithms. Because most recommender systems rely on weak data, demographic data was 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 using DBSCAN clustering at two levels to detect POIs in each area, making clustering detection easier and more complicated. The TF-IDF approach was used to assess context similarity.

The Prefix-Span approach is also 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 from London and Yelp discovered in the US). The recommended strategy, according to the results of the comparison, can offer more precise locations than other ways. The proposed technique beat 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.

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 travel 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.

5.1 Future work

A multitude of promising pathways beckon for extending the framework expounded in this study. While the current methodology adeptly caters to tourist recommendations within a confined set of POIs, prospective research carries the potential to seamlessly incorporate additional contextual factors, encompassing elements like trip cost and overall travel time. This integration is poised to heighten the efficiency of the tourist recommender system, adeptly accommodating a diverse spectrum of tourist preferences. Another compelling avenue for enhancement involves the expansion of the dataset to include a diverse array of cities, thereby ensuring a more comprehensive applicability of the proposed method. Multiple compelling directions for extension emerge, encompassing the exploration of user relationships and the nuanced impact of companions on recommendations to fortify precision. Additionally, considering user reviews from social network posts promises valuable insights into user experiences. These advancements harbor significant potential in amplifying the effectiveness of the tourist recommender system, catering to an even broader spectrum of tourist preferences. The evolution of neural recommendation systems, leveraging deep learning models and contextual information from social media posts, is anticipated to accentuate key aspects such as explainability; and personalization.

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest None.

References

1. Abbasi-Moud Z, Vahdat-Nejad H, Sadri J (2020) Tourism recommendation system based on semantic clustering and sentiment analysis. Expert Syst Appl 167114324–114341. https://doi.org/10.1016/j.eswa.114324
2. Noorian A (2024) A BERT-based sequential POI recommender system in social media. Comput Stand Interfaces 87:103766. https://doi.org/10.1016/j.csi.2023.103766
3. Aggarwal CC (2016) Recommender systems. The textbook, Springer, IBM T.J. Watson Research Center Yorktown Heights, NY, USA, 2016
4. Ali Z, Qi G, Kefalas P, Abro WA, Ali B (2020) A graph-based taxonomy of citation recommendation models. Artif Intell Rev 53:5217–5260. https://doi.org/10.1007/s10462-020-09819-4

5. Alrasheed H, Alzeer A, Alhowimel A, Althyabi A (2020) A Multi-level tourism destination recommender system. Procedia Comput Sci. 170(333):40. https://doi.org/10.1016/j.procs.2020.03.047
6. 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
7. Bin C, Gu T, Sun Y, Chang L (2019) A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining. Multimedia Tools Appl 78(24):35135–35156
8. 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
9. Chen L, Zhang L, Cao S, Wu Z, Cao J (2019) Recommendation: Deep and collaborative learning with textual information. Expert Syst Appl 144 (113070). https://doi.org/10.1016/j.eswa.2019.113070
10. Dakhel A, Malazi H, Mahdavi M (2018) A social recommender system using item asymmetric correlation. Appl Intell 48(3):527–540
11. Dietz L, Sen A, Roy R, Wörndl W (2020) Mining trips from location-based social networks for clustering travelers and destinations. Inf Technol Tour 22(1):131–166
12. Flickr (2022) www.Flickr.com. Accessed 23 Jan 2022
13. Dara S, Chowdary CR, Kumar C (2020) A survey on group recommender systems. J Intell Inf Syst 54(2):271–295. https://doi.org/10.1007/s10844-018-0542-3
14. Han M, Lee J (2015) Bayesian typhoon track prediction using wind vector data. CSAM 22241–253. https://doi.org/10.5351/CSAM.2015.22.3.241
15. Noorian A, Harounabadi A, Ravanmehr R (2022) A novel Sequence-Aware personalized recommendation system based on multidimensional information. Expert Syst Appl 202:117079. https://doi.org/10.1016/j.eswa.2022.117079
16. Hsu C (2021) A multi-valued and sequential-labeled decision tree method for recommending sequential patterns in cold-start situations. Appl Intell 51(1):506–526. https://doi.org/10.1007/s10489-020-01806-0
17. Hu Y, Yang B (2015) Enhanced link clustering with observations on ground truth to discover social circles. Knowl-Based Syst. 73:227–235
18. Kefalas P, Manolopoulos Y (2017) A time-aware spatio-textual recommender system. Elsevier. Expert Syst Appl 78:396–406
19. Kolahkaj M, Harounabadi A, Nikravanshalmani CR (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
20. Kuanr M, Mohapatra P (2021) Recent challenges in recommender systems: A survey. In Progress in Advanced Computing and Intelligent Engineering (pp 353–365). Springer, Singapore
21. Kulkarni S, Rodd S (2020) Context-aware recommendation systems: A review of the state of the art techniques. Comput Sci Rev 37:100255
22. Linda S, Bharadwaj K (2019) A genetic algorithm approach to context-aware recommendations based on Spatio-temporal aspects. Integr Intell Comput Commun Sec 771(2019):59–70 Springer, Singapore
23. Lim K, 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(2):375–406. https://doi.org/10.1007/s10115-017-1056-y
24. Lyu D, Chen L, Xu Z, Yu S (2020) Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geotagged photos. Appl Intell 50:924–938. https://doi.org/10.1007/s10489-019-01566-6
25. Majid A, Chena L, Mirza H, Hussain I, Chena I (2015) A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. Data Knowl Eng 95:66–86
26. Memon I, Chen L, Majid A, Lv M, Hussain I, Chen G (2015) Travel recommendation using geo-tagged photos in social media for tourists. Springer. Wirel Pers Commun 80(4):1347–1362
27. 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(4):251–264
28. Noorian A, Ravanmehr R, Harounabadi A, Nouri F (2020) Trust-based tourism recommendation system using context-aware clustering. Tour Manag Stud 15(52). https://doi.org/10.22054/tms.2020.41870.2137
29. Ojagh S, Malek M, Saeedi S, Liang S (2020) A location-based orientation-aware recommender system using IoT smart devices and Social Networks. Futur Gener Comput Syst 108:97–118
30. Pirasteh P, Hwang D, Jung J (2015) Weighted similarity schemes for high scalability in user-based collaborative filtering. Mob Netw Appl 20(4):497–507. https://doi.org/10.1007/s11036-014-0544-5
31. Pirasteh P, Hwang D, Jung J (2015) Exploiting matrix factorization to asymmetric user similarities in recommendation systems. Knowl Based Sys. 83:51–57

32. Ravi L, Subramaniyaswamy 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
33. 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
34. Sarkar J, Majumder A, Panigrahi C, Roy S (2020) MULTI TOUR: A multiple itinerary tourist's recommendation engine. Electron Commer Res Appl 40:100943. https://doi.org/10.1016/j.elerap.2020.100943
35. Shokeen J, Rana C (2019) Social recommender systems: techniques, domains, metrics, datasets, and future scope. J Intell Inf Syst 54:633–667
36. Ravanmehr R, Mohamadrezaei R (2024) Deep learning overview. In: Session-Based Recommender Systems Using Deep Learning. Springer, Cham. https://doi.org/10.1007/978-3-031-42559-2_2
37. Sun X, Huang Z, Peng X, Chen Y, Liu Y (2019) Building a model-based personalized recommendation approach for tourist attractions from geo-tagged social media data. Int J Digit Earth 12(6):661–678. https://doi.org/10.1080/17538947.2018.1471104
38. Tang J, Wang K (2018) Personalized top-N sequential recommendation via convolutional sequence embedding. Paper presented at the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA
39. Thomee B, Shamma D, Friedland G, Elizalde B, Poland D, Ni K, Borth D, Li L (2016) YFCC100M: The New Data in Multimedia Research. Commun ACM. 59(2):64–73
40. Tong Z, Acuna D (2020) GotFunding: A grant recommendation system based on scientific articles. Proc Assoc Inf Sci Technol 57(1):323. https://doi.org/10.1002/pra2.323
41. Trivonanda R, Mahendra R, Budi I, Hidayat R (2020) Sequential pattern mining for e-commerce recommender system. At the 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE. 393–398
42. Veloso B, Leal F, Malheiro B, Burguillo J (2020) A perspective on "Online guest profiling and hotel recommendation": Reliability, Scalability, Traceability, and Transparency. Electron Commer Res Appl. 40:100957. https://doi.org/10.1016/j.elerap.2020.100957
43. Viktoratos I, Tsadiras I, Bassiliades V (2018) Combining community-based knowledge with association rule mining to alleviate the cold start problem in context-aware recommender systems. Expert Syst Appl Elsevier BV. 101:78–90
44. Villegas N, Sanchez C, Dõaz-Cely J, Tamura G (2018) Characterizing context-aware recommender systems: A systematic literature review. Knowl-Based Syst. 140:173–200
45. Vineela A, Devi G, Nelaturi N, Yadav G (2021) A comprehensive study and evaluation of recommender systems. Microelectronics, Electromagnetics, and Telecommunications. Springer, Singapore, pp 45–53
46. Wang Y, Wang P, Liu Z, Zhang L (2021) A new item similarity based on α-divergence for collaborative filtering in sparse data. Expert Syst Appl 166:114074
47. 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. https://doi.org/10.1016/j.jss.2020.110862
48. Zhao K, Zhang Y, Yin H, Wang J, Zheng K, Zhou X, Xing C (2020) Discovering subsequence patterns for next POI recommendation. The Twenty-Ninth international conference on artificial intelligence 3216–3222
49. Zhang C, Li T, Ren Z, Hu Z, Ji Y (2019) Taxonomy-aware collaborative denoising autoencoder for a personalized recommendation. Appl Intell 49(6):2101–2118
50. Sayari S, Harounabadi A, Banirostam T (2023) A hybrid method for community detection based on user interactions, topology and frequent pattern mining. J Model Eng. https://doi.org/10.22075/JME.2023.29816.2402
51. Tarus J, Niu Z, Kalui D (2017) A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. Soft Comput 22(8):2449–2461
52. Tarus J, Niu Z, Yousif A (2017) A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. Futur Gener Comput Syst 72:37–48
53. Seilsepour A, Ravanmehr R, Nassiri R (2023) SSTSA: A self-supervised topic sentiment analysis using semantic similarity measures and transformers. 101142/S0219622023500736
54. Noorian A, Harounabadi A, Hazratifard M (2023) A sequential neural recommendation system exploiting BERT and LSTM on social media posts. Complex Intell Syst 1–24. https://doi.org/10.1007/s40747-023-01191-4
55. Seyfosadat SF, Ravanmehr R (2023) Systematic literature review on identifying influencers in social networks. Artif Intell Rev 56:567–660. https://doi.org/10.1007/s10462-023-10515-2

56. Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann, Boston, pp 39–82
57. Noorian Avval AA, Harounabadi A (2023) A hybrid recommender system using topic modeling and prefixspan algorithm in social media. Complex Intell Syst 9:4457–4482. https://doi.org/10.1007/s40747-022-00958-5
58. Kala KU, Nandhini M (2019) Context-Category Specific sequence aware Point-Of-Interest Recommender System with Multi-Gated Recurrent Unit. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-019-01583-w
59. Hong M, Jung JJ (2021) Multi-criteria tensor model for tourism recommender systems. Expert Syst Appl 170:114537