Hybrid Recommender System for Tourism Based on Big Data and AI: A Conceptual Framework

Khalid AL Fararni*, Fouad Nafis, Badraddine Aghoutane, Ali Yahyaouy, Jamal Riffi, and Abdelouahed Sabri

Abstract: With the development of the Internet, technology, and means of communication, the production of tourist data has multiplied at all levels (hotels, restaurants, transport, heritage, tourist events, activities, etc.), especially with the development of Online Travel Agency (OTA). However, the list of possibilities offered to tourists by these Web search engines (or even specialized tourist sites) can be overwhelming and relevant results are usually drowned in informational “noise”, which prevents, or at least slows down the selection process. To assist tourists in trip planning and help them to find the information they are looking for, many recommender systems have been developed. In this article, we present an overview of the various recommendation approaches used in the field of tourism. From this study, an architecture and a conceptual framework for tourism recommender system are proposed, based on a hybrid recommendation approach. The proposed system goes beyond the recommendation of a list of tourist attractions, tailored to tourist preferences. It can be seen as a trip planner that designs a detailed program, including heterogeneous tourism resources, for a specific visit duration. The ultimate goal is to develop a recommender system based on big data technologies, artificial intelligence, and operational research to promote tourism in Morocco, specifically in the Daraâ-Tafilalet region.

Key words: recommender systems; user profiling; content-based filtering; collaborative filtering; hybrid recommender system; e-tourism; trip planning

1 Introduction

In the field of tourism, recommender systems could be a great help when planning a trip or searching for a service among many destinations, attractions, and activities[1].

Strictly speaking, these systems are defined as information filtering systems that make recommendations of the most suitable offers (products, services, …) to customers[2,3], for example, products that are similar to other products they have already bought and enjoyed[4] or products that have already been enjoyed by other customers with similar tastes[5].

The principle is to use the interests of a user collected during his navigation as inputs, to predict the degree of interest that this user may have for a given item[6]. The approaches used to estimate these degrees of appreciation are numerous. They are traditionally classified by the literature into several categories according to the source of information used[3,7]. One of these approaches is based on ratings given by a set of users on a set of items. It consists in recommending to a given user the items that have been highly evaluated in the past by other users who have similar preferences, we speak here about collaborative
filtering\cite{8}. With the advent of social networks, research has used social information in addition to the rating matrix as input to build social recommender systems\cite{9}. These systems consist in calculating the similarities between the target user and his social space, according to a specific metric. In recent years, research has integrated contextual information (location, weather conditions, . . . ) in recommender systems due to their importance\cite{10,11}. According to Ref. \cite{12}, context is defined as “any information that can be used to characterize the situation of an entity”. An entity may be a person, place, or object considered relevant. In tourism, an object can be of different types (monuments, parks, museums, etc.).

As there is a multitude of recommendation approaches, and due to the variety of their information sources and the heterogeneous nature of tourism data, the implementation of this type of system, especially those using a hybrid approach, encounters several difficulties. The main objective of this work is then to contribute to the design of tourism recommender systems by proposing a framework that clarifies how the hybrid recommendation process works and details each of these steps. The implementation of the proposed framework, in addition to serving to gather the benefits of the different recommendation approaches, will provide a better visitor experience by recommending the most relevant items and helping visitors personalize their itinerary. Then, we give a brief description of our big data system that aims to integrate our recommendation system and opinion analysis using deep learning techniques\cite{13}.

This document is organized as follows: In Section 2, we will briefly present a literature review of the most commonly used recommendation approaches in the field of tourism. Section 3 presents a description of the proposed architecture for tourism recommendation systems and its conceptual framework. In Section 4, we describe a methodology for integrating Big Data and Artificial Intelligence (AI) in the proposed architecture, and we conclude in Section 5 on the perspectives of our work.

2 Literature Review

Current tourism recommendation techniques have several innovative aspects and can be classified in different ways, depending on how they analyze the user’s information and filter the list of items\cite{14}.

2.1 Collaborative filtering

This approach aims to offer visitors destinations they have not yet been to, but which they might like, based on the habits and tastes of similar users’ profiles\cite{15}; the similarity of taste between two users is calculated based on the similarity of their rating history.

The VISIT system\cite{11}, for example, applies sentiment analysis techniques (using the Alchemy Application Programming Interface (API)) to analyze news about a given attraction on Twitter and Facebook and determine whether users make positive or negative comments about it. This information is displayed in green and red by the system in its interface so that the user can easily identify the places that visitors enjoy most today and those are not.

However, this approach is more difficult to meet the needs of tourists, when it is almost impossible to match users’ travel history (see rating), it is very difficult to find two people on the same trip, with the same duration, the same places of interest, and the same experience.

2.2 Content-based recommender system

For making recommendations to potential visitors, content-based systems are based on the analysis of content similarities between items previously consulted by users (or examined in the present) and those have not yet been consulted\cite{16,17}.

Content-based filtering is the most popular and widely used technique in tourism recommendation systems\cite{14,18}.

The system proposed in Ref. \cite{19} defines a content-based recommendation method for cultural heritage (tangible and intangible). This method selects resources based on user preferences and item metadata, orders items using multi-criteria user feedback, and enriches the set of suggestions using semantic relationships between items.

A natural limitation of content-based filtering is the need to have a generic and rich representation of the content of the items, which is not the case for tourist items characterized by their great extent and variety. Moreover, this type of system generally suffers from the problem of overspecialization; for example, when a tourist enjoys an event or a show during a trip, it does not mean that he will want to see it again. However, using a content-based approach, the system will suggest him to come back a second time to the same place with the same type of event (even if it is not organized!!), when he might be more interested in events, he did not
2.3 Context-aware filtering

Recommender systems are called context-sensitive when they use context in their calculations to predict what is likely to be of interest to the user\cite{11}. The most commonly used context elements in tourism recommendation systems are geolocation, weather, visit history, and weather.

Today, many mobile devices connected to the Internet, also known as “connected objects”, are widely available and used to capture and provide a wealth of information that can enrich the current context and its variations.

Reference \cite{20} proposed a context-sensitive recommender system and presented a definition of the concept of context through a meta-model. An applied case study was conducted in the city of Tangier. The system developed is composed of three main modules: Context, composed of the user’s profile, Spatio-temporal and environmental information (i.e., location, characteristics of the device used to access the application, external physical environment (e.g., weather conditions) and information collected about the community (from Facebook, Twitter, and other social network applications)); the tourism content repository, which contains tourism service data; and the recommender system.

2.4 Hybrid recommender system

The hybridization of these approaches, to overcome the shortcomings of each technique used alone and take advantage of their strengths, has been the subject of several research studies\cite{21}.

Reference \cite{22} proposed a new architecture for a recommender system for individuals and groups, based on the characteristics of the works of art, the context of all users, and the social affinity between these users. The system is a hybridization of three approaches: the content-based approach, the social approach, and the context-based approach.

2.5 Discussion

It should be noted that 90% of the current solutions are generally concentrated on a single category of items (hotels, museums, tourist sites, …)\cite{14}, providing only tourist services information (inserted in the system by the administrator or by experts) to make the trip more pleasant; besides, most of these works use a single approach, with a clear predominance for content-based approaches\cite{14,18}.

For all these reasons, there is a need for a conceptual framework not only to gather the recommendation approaches but also to present the different tourism resources in a single architecture.

3 Proposal of Hybrid Tourism Recommender System Architecture

3.1 Reference architecture for tourism recommender system

Our research work consists of proposing a new architecture for tourist recommendation systems. This architecture is based on a hybrid recommendation approach, which aims to improve user access to tourism resources in information retrieval systems, such as tourism portals and service providers’ documentary Extranets.

Another innovative aspect of this architecture is that the proposed system goes beyond a list of recommended tourist attractions and can be seen as a planner that aims to build a complex and detailed program of a multi-day visit. The client will thus be offered a diversified list of tourist resources (monuments, activities, hotels, shows, …) that exactly meet their specific needs and preferences.

We propose to decompose the proposed system architecture into five main modules (Fig. 1):

1. **Visitor profiles** contain in particular information that can be used to determine user preferences in terms of items (ratings, social information, etc.).

2. **Services repository** contains information on tourist services (such as accommodation, restaurants, tourist sites, transport, …) as well as associated multimedia content.

3. **A contextual meta-model** takes into account multiple factors involved in manipulating context, such as time, space, location, the distance between two places,
routes, tourist travel history, etc., to make a specific recommendation.

(4) The hybrid filtering process returns a list of items with the degrees of appreciation that the target user can give to each item.

(5) A trip planner selects items considered relevant to the user, and uses operational research techniques to correlate these choices in the form of a trip.

3.2 Conceptual framework of the proposed architecture

The conceptual framework of the proposed architecture is made up of three main sub-processes, which are a process known as user profiling, a process for selecting content (filtering) that best matches user profiles, and a trip planning process. These processes take place at the intersection of different areas of computer science research, including artificial intelligence and operational research.

For example, in artificial intelligence, the profiling process can be expressed as a learning problem that exploits users’ past knowledge. Often, the system should learn the user’s profile rather than requiring the user to provide it. This usually involves the application of Machine Learning (ML) techniques.

The purpose of the filtering process is to learn how to categorize new information based on previously seen information that has been implicitly or explicitly labeled as interesting or uninteresting by the user. With these labels, ML methods can generate a predictive model that, given a new item, will help to decide the degree of interest the user may have in it.

In operational research, the trip planning process leads to the formal definition of a combinatorial optimization problem that is a variant of the travelling salesman problem. To solve this kind of problem, we can rely on metaheuristics to tend towards a good solution in a reasonable time.

The conceptual framework of the proposed architecture (Fig. 2) consists of three main processes, namely, the profiling process, the filtering process, and the trip planning process.

3.2.1 User profiling process

User data collection or user profiling is a very important step in the proposed framework. This process includes four scenarios, from which the modules constituting the user profile can be extracted.

(1) Inscription. The user explicitly indicates his interests to the system through a registration form; for example, by assigning:

- Comment fields, key-words, or tags to be selected
by the user to indicate his preferences (purpose of the trip, type of items consulted, . . .):

- Ratings on the items consulted indicating the degree of appreciation of an item by this user; and
- Demographic attributes about the user, such as age, gender, socio-professional category, geographical location, personal status, etc. Although these attributes do not provide information on the ratings, they allow us to refine the user profile and adapt the recommendations. Furthermore, demographic data can be used to calculate recommendations for new users; the demographic approach is first used to solve the cold start problem[23, 24].

(2) Social login. Instead of creating a new login account specifically for the system, the user can use his existing login information in a social network such as Facebook, Twitter, or Google+ to log in Refs. [23, 25]. This login then allows the system to extract demographic data as well as some information about the user’s relationships[23–26].

(3) Consultation “even without log-in”. It relies on observation and analysis of user behavior implicitly carried out in the application that embeds the recommender system, the whole thing is done in the “background” (basically without asking the user for anything). We call these behaviors “traces of use”. These traces[27] can include:
- Indicators describing the manipulation, such as “copying/pasting” a text from a page, searching for a text in a page, adding or deleting an item from the shopping cart, or ordering an item (in e-commerce applications), saving or printing a page, etc.
- Navigation indicators, such as frequency and duration of browsing, number of clicks and mouse hovers on a page or links, scrolling, etc.[18]

(4) Context. Context is based on the integration of contextual information (location, time, physical environment, . . .) for the generation of dynamic and personalized visit itineraries.

The data collected about the user are then selected, analyzed, and saved as independent modules. These modules are combined to build the “user profile” (Fig. 3). A profile will contain information that can be used to determine user preferences in terms of items (activities, tourist sites, etc.).

3.2.2 Filtering process
The adaptation of recommendation techniques is entirely based on the result of the profiling process. The process takes as input all the modules that constitute the target user’s profile:

- The content-based module describes the characteristics of tourist sites/activities that the user has consulted in the past in the form of key-word vectors that are generated after an indexing phase. It should be noted that these key words are generally extracted automatically during the consultation or manually assigned during the inscription.
- The collaborative/social module contains the rating data of the consulted items.
- Demographic module contains the user’s demographic attributes. These attributes can either be entered by the user himself by filling in the registration form or extracted from his social login[24].

Once the user profile modules are detected, recommendation approaches and the appropriate hybridization technique are selected. Burke[28] summarized hybridization techniques in seven techniques: weighting, mixing, cascading, switching, features combination, features augmentation, and the meta-level method.

This process returns a list of items with the degrees of appreciation that the target user can give to each item (Fig. 1).

3.2.3 Trip planning process
Once the degrees of appreciation are estimated, the system selects the items considered relevant for the user (exceeding a given threshold, for example), taking into account his context, and uses operational research techniques to correlate these recommendations in the form of a trip.

4 Methodology of Integration of Big Data and AI for Implementing the Proposed Architecture
The integration of big data and AI for the implementation of the proposed recommender system is one of the
main axes of a project which aims to build a big data solution based on hybrid recommendation, sentiments, and opinions analysis using machine and deep learning techniques[13]. This project aims to provide intelligent tools to target and recommend the most suitable tourist offer according to the users’ profile, and to track and analyze their opinions to improve the customer experience and forecast the tourist demand in Morocco. This project will help Moroccan tourism agencies and actors, especially those of the Daraâ-Tafilalet region, and enable them to be more available on the Internet and offer a better service to visitors. To achieve this, a four-layer methodology will be presented which describes the approach of integrating big data and AI in the proposed system.

(1) Tourist data aggregation layer. This layer consists of providing a very wide range of digital tools to increase the visibility and attractiveness of the tourist offers of the Daraâ-Tafilalet destination—portals, mobile/web applications, social media, augmented reality, reconstructions of monuments in 3D, virtual museums, interactive terminals, and maps and E-guides on mobiles. These different tools can provide a considerable amount of informative content, but also images and videos. They offer the visitor a real “immersion” in the destination, thus contributing to the reputation of the destination and the intensity of the tourist experience. In this sense, many projects have already been launched to conserve and enhance the region’s cultural and natural heritage[29] (Fig. 4).

Tourism data are often voluminous and heterogeneous, which leads to the use of big data to store this large volume, and to support and manipulate this wide variety of data. For that, a wide range of innovative technological solutions, such as NoSQL database management systems (Cassandra, MongoDB, . . .) and distributed file systems (Hadoop HDFS), are adopted. Other researches[30,31] also aim at enriching these data semantically by using Semantic Web technologies, especially ontologies, for more intelligent management and access.

(2) Recommendation layer. The work developed in this paper is part of this layer, which aims to identify the user profile in terms of preferences, by determining modules (demographic, contextual, preferences, ratings, etc.) that compose it, and choose the appropriate recommendation approach to use and develop the appropriate algorithm to be executed to take advantage of large datasets. In this sense, big data technologies offer a large-scale implementation of several machine learning and deep learning techniques, including classification, clustering, association rules, regression, collaborative filtering, recurrent neural network, etc. Based on these techniques, we can process and analyze in real-time the different tourist information, exploit the results obtained to predict the next actions of tourists, and recommend appropriate offers and itineraries. For example, by analyzing the routes and the time spent in front of the monuments, the recommender system can better understand and respond to visitors’ expectations.

(3) Results visualization layer. This layer will accompany the tourist during all phases of his trip, from preparation to online sharing, “before/during/after”. Now when preparing a trip, we start by researching sites, portals, and mobile applications to get information and make choices, we prepare our itinerary on the spot with services and personalized content, and we finish by multiplying ourselves on social networks, blogs, and forums. However, to personalize an itinerary, the system will rely on operational research techniques and present the results on interactive maps.

(4) Layer for validating the proposed solution. This layer consists of monitoring and analyzing the opinions and sentiments of visitors presented on blogs and social media, by tourism companies, to understand what changes in their desires and share the information that influences their decisions. Feedback data are usually transformed, through visualization and Business Intelligence (BI) technologies, into graphical representations (graphs, histograms, etc.) that are simple and easy to understand. These graphical representations will allow tourism professionals to have a clear view to make the right decisions, and improve their strategies.
They will thus be able to identify the needs of tourists, predict their future behavior, and propose the resources best suited to each profile.

The proposed recommender system architecture will then be implemented, using new big data and AI technologies. We have chosen the Daraa-Tafilalet region as a destination to design and validate the proposed solution which will then be extended to all regions of the Kingdom.

5 Conclusion

To assist tourists in the selection process and overcome the information overload, recommendation systems were developed in the last decade of the twentieth century. In this paper, we have introduced a literature review of the current tourism recommender systems and then we have presented a new conceptual framework to implement tourism recommender systems. Our hybrid architecture aims to improve the visitor experience by recommending the most relevant items and helping him to personalize his trip.

Once the sets of elements considered relevant to the tourist are selected, our system will plan an appropriate trip by combining these items using operational research techniques.

This architecture will be implemented, through advanced technologies, such as big data tools, machine learning techniques, and the Internet of things.

References


Khalid AL Fararni is a PhD student in computer science at Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco. He received the master degree in imaging and business intelligence from the same university in 2018. His main research is in the areas of big data, recommender systems, and machine learning.

Fouad Nafis is a PhD student at the Sidi Mohamed Ben Abdellah University. He works as a computer science teacher in the preparatory classes. His research focuses on semantic integration based on several technologies (ontologies, RDF/Owl, SPARQL, etc.) of semantic web, cultural heritage data, open data, linked data, and big data technologies.

Badriddine Aghoutane obtained the PhD degree in computer science from Sidi Mohamed Ben Abdellah University, Morocco in 2011. He is a professor at the Department of Computer Sciences, Faculty of Science, Moulay Imail University, Meknes, Morocco. He joined Moulay Ismail University (UMI) in 2011. He was an assistant professor at the Polydisciplinary Faculty of Errachidia from 2011 to 2019. He is the manager of “software platforms for cataloging, management, and dissemination of the cultural heritage” research project (UMI program-2016) and the member of the CUI-UMI/GIRE project in the framework of VLIR UOS Programs. He is the chairman of the Scientific Committees of SRIE’15, CHAT’19, and MITA’2020 conferences as well. He was also a member of the Organizing and the Scientific Committees of several international symposia and conferences dealing with topics related to computer sciences, technologies, and their applications. His research interest is towards Web semantic, recommender systems, IoT security, and big data.

Ali Yahyaouy is a professor at the Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco. He received the PhD degree from Sidi Mohamed Ben Abdellah University in 2010. His research themes are intelligent homes and transports.

Jamal Riffi obtained the PhD degree in computer science from Sidi Mohamed Ben Abdellah University, Morocco in 2015. He is an assistant professor at the Faculty of Science Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco. His main researches focus on machine learning, medical image analysis, and text mining.
Abdelouahed Sabri is a professor at the Faculty of Science Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco. He received the PhD degree from Sidi Mohamed Ben Abdellah University in 2009. His research areas include image processing and information retrieval.