Smartphone-based tourist trip planning system: a context-based approach to offline attraction recommendation

Sergei Mikhailov^{1,2,*} and Alexey Kashevnik²

¹ITMO University, Saint Petersburg, Russian Federation

²SPIIRAS, Saint Petersburg, Russian Federation

Abstract. The paper presents a tourist trip planning system that is aimed at offline attractions recommendation for a tourist based on current situation in the region and user profiles. Authors present the related work analysis in this area and based on this analysis develop the reference model of the proposed system and describe implementation that is based on previously developed Tourist Assistant — TAIS.

1 Introduction

According to the UNWTO World Tourism Organization, the total number of international tourist trips increased from 25 million in 1950 to 674 million in 2000, and to 1.235 million in 2016. Income from international tourism increased from \$ 2 billion in 1950 to \$ 495 billion in 2000 and to \$ 1.220 billion in 2016. It is expected that the number of international travel will increase by 3.3% per year until 2030 and will reach 1.8 billion. The tourism industry began to actively use Internet technologies since the mid-1990s. At the current time, the planning of tourist trip and support for the decisions of the tourist is an object of considerable scientific interest [1].

Over the past ten years there has been an explosive growth in the use of mobile technologies among tourists. E-tourism systems provide a good opportunity for mobile services that help tourists by offering recommendations based on their preferences and context [2]. Presented in the paper Tourist Trip Planning System is based on developed earlier Tourist Assistant — TAIS [3] that is a system that proactively recommends the tourist attractions around, which is currently better to attend personally for him/her. The recommendation is based on tourist’s preferences and current situation in the location region. The collaborative filtering technique is used to determine tourists preferences and generate recommendations based on these preferences. Information about attractions is extracted from different Internet sources on the fly (e.g., Wikipedia, Wikivoyage). The main idea of the approach is that TAIS consists of several services, which are united by the smart space for providing the tourist with information according to personal preferences and current situation in the location.

The Tourist Trip Planning System is a service-oriented system which provide attraction visiting route and information about attraction to a tourist. Authors offers to use

a context-oriented approach for constructing tourist attraction visiting route. The proposed system consist of a several services which are interacting with each other. Also, the system provides a region databases for offline attraction details processing from accessible Internet source on the user smartphone.

The paper is structured as follows. Section 2 describes the related work in the area of attraction recommendation systems. Section 3 offers concept of the tourist trip planning system based on context approach and describers system services. Section 4 provides a detailed description of system services and section 5 gives a conclusion.

2 Related work

The paper [4] describes personalized location-based mobile tourism application. This application is based on the iBike system in Taichung City, Taiwan and uses a hybrid filtering technique to collect tourist information, to gather tourism information and adapts the ant colony optimization algorithm for more precise adjustment of geolocation recommendations to tourists. The authors use the technology acceptance model and the information systems success model for users intentions investigating to further using in the developed system. The technology acceptance model has been used to interpret users acceptance of information technology. It comprises three dimensions: perceived ease of use, perceived usefulness, and usage behavioural intention. The information systems success model contains a six factors: system quality, information quality, system use, user satisfaction, the effect on the individual user, and the effect on the organization.

Authors of the paper [5] develop geolocation services for the multi-agent information collection of Dr.What-Info system [6]. This system uses Google Maps and image recognition technologies as a source of tourist information and a route planner for users. Users requests to system are sent in the format of the canonical user request representa-

*e-mail: saboteurincave@gmail.com

tion language, which represents information as knowledge frames. The user request is processed by several agents: the ontological data mining agent is used to provide current knowledge in the field of data mining techniques and cluster classification; the ontological case-based reasoning agent uses a reasoning mechanism based on use cases to provide a user-centric and information-related solution. If these agents can not complete the task, then the ontology information agent shell searches for solutions from external mechanisms by crawling information with OntoCrawler tool, retrieving data using OntoExtractor tool, classifying data using OntoClassifier tool and the recommended OntoRecommender agent.

The authors of the paper [7] are develop a personalized travel planning system which uses the time framework concept with regulated recommendations to users. The personalized travel planning system assumes that users can setup their personal requirements through the Internet platform. Through the travel planning module a preliminary travel schedule is created, and the configuration functions help the user to obtain the necessary travel plan. The travel planning module uses a schedule determination method to create a time framework-based travel schedule. This method includes distance calculations, popularity calculations, and time calculations. The information of all matching locations and time block categories are the input parameters for the schedule determination method. The travel requirements module cross-referenced the user's requirements and the travel destination to select the destination that best meets the user requirements. Places that meet the user's requirements are connected and formatted in time framework depending on the time for planning a preliminary trip schedule.

In the paper [8] the authors propose a prototype of a tourist system that offers a method based on multi-criteria collaborative filtering. This method improves the accuracy of forecasting tourist guidance systems by using clustering, reducing the size of samples and forecasting methods. The authors used an adaptive network based on the adaptive neuro-fuzzy inference system and support vector regression methods for predictions. The principal component analysis was used as a dimensionality reduction technique. The authors used the self-organizing map and the expectation maximization algorithm, because they are well-known methods of clustering in scientific community. To improve the accuracy of multi-criteria collaborative filtering recommendations the hypergraph partitioning algorithm are applied to the clustering results. The authors evaluated the accuracy of the recommendation method in the TripAdvisor data set and their experiments indicate that cluster ensembles have greater prediction accuracy than single-source clustering techniques.

In the paper [9] the authors represent the PlanTour system, which creates personalized tourist routes using the information collected from the travellers social network MINUBE¹. The system uses an automated planning approach to create a multi-day plan with the most interesting sights of the visited city/region. In particular, the system

collects information about users and points of interest from the MINUBE, groups these points using clustering methods to divide the problem into subtasks. After these steps the system uses an independent from the tourist area automated scheduler, which finds tourist plans. Unlike other tourist recommendation systems, the PlanTour planner can organize the relevant objects of interest, taking into account the expected travel of the user and user estimates from a real social network.

The paper [10] uses geo-tagged photos in social networks to build routes to attractions. Also, the authors use a separate segment of the road as the basic routing unit instead of the GPS path segment and have created a system of recommendations that provides users with the most popular places in the city/area, as well as the best travel routes between attractions. Using Flickr photos with geo-tags, it is possible to determine the best places to travel in the city and the best travel routes between popular tourist destinations. The authors apply the spatial clustering method to determine the main tourist reference points and the subsequent ranking. Using a machine learning method, the system calculates the tourist popularity of the road in terms of the relevant parameters, for example, the number of users and the number of points of interest. The best route recommended to the user should have a minimum distance and maximum tourist popularity.

The authors of the paper [11] solve the widespread issue of information congestion in advisory systems and the problem of incomplete data in areas with low level of tourism and non-standard routes of attractions. To solve these problems, the authors propose a new personalized travel recommendations scheme that uses explicit user interaction and multimodal travel information. At first, the system collects heterogeneous sights information through a multi-user exchange, which can be considered as a collective intelligence used to provide reliable references by other travellers. Secondly, valuable knowledge is extracted from the collective intelligence to filter out noise from the data and to structure travel information. The model of Personalized Attraction Similarity is designed to offer attractions through the merger of heterogeneous information with weighted adaptation and simultaneous study of explicit interaction with the user. A contextual information, such as a user location, is used to generate the recommendations that can affect the user choice at a certain point in time.

The paper [12] offers the «Filter-first, tour-second» approach for creating personalized recommendations on tourist routes, based on information from social networks and other online data sources. The collaborative filtering technique is used to define a subset of additional points of interest that maximize potential user satisfaction from the route. Further, the orienteering problem is solved using the iterative tabu search algorithm, which creates tourist routes that must contain all the mandatory places and maximize the overall satisfaction from the route when using additional daily visited places, taking into account access times and hours of operation, restrictions on tour time, user's budget and other restrictions.

¹<http://www.minube.com/>

The paper [13] presents the eCOMPASS multimodal tourist tour planner. This tourist tour planner provides a navigational aid and incorporates multimodality within its routing logic aiming at deriving near-optimal sequencing of points of interest (POI) along recommended tours to utilize the time available for sightseeing and minimize waiting time at transit stops. eCOMPASS allows users to define arbitrary start/end locations (e.g. the current location of a mobile user) rather than choosing among a fixed set of locations. The core engine of eCOMPASS is based on a novel cluster-based heuristic approach, the SlackRoutes. The main incentive behind approach is to motivate visits to topology (i.e. touristic) areas featuring high density of «good» (i.e. highly profitable) candidate vertices, even if those are located relatively far. SlackRoutes takes into account time dependency (i.e. multimodality) in calculating travel times from one vertex (i.e. POI) to another.

In the paper [14] authors describes a tourist preferences based recommendation system. The system can build a route taking into account the use of various types of vehicles and context information (weather restrictions and attraction schedule). The authors uses an ant colony optimization algorithm with using an attraction weighted categories as a combinatorial optimization method for a global route constructing. The attractions was stored into a light-weighted ontology.

All the solutions considered provide either a set of attractions that a tourist can visit, or a route of attractions, which does not take into account the current situation in the city. The authors of papers [4, 7, 9, 12] does not work with context information during attraction route visiting at all, authors of paper [11] formalizes context only as user location (city and coordinates), [14] context does not contain traffic information, [10] context contains only a topological cafeterias, [5, 6] does not give many details about route planning and more focusing on ontological method for attraction processing and the authors of [8] describes only a recommendation tourism system without route planning. The proposed method in system will take into account the situation on the road, including traffic jams, current weather and season, attractiveness of sights and their schedules. In another words, the constructing route method should use the current contextual information. It is also proposed to use the user profile to build a route with remembering the history of visits and taking into account the assessments and preferences of users on the attractions.

3 Tourist trip planning system approach

The tourist trip planning system is a service-oriented system which provide attraction visiting route and attraction information to a tourist. This system offers to use a context-oriented approach for constructing tourist attraction visiting route in his\her region. Context situation in the region and user profile are formalized accordingly to an ontology. The system can store a region information in sqlite-database and provides it to user for offline attraction processing on a smartphone. For building the attraction

visiting route, methods and algorithms of graph theory are used.

A reference model of tourist trip planning system is presented on Figure 1. The tourist trip planning system provides content to an user and consists of a set of interacting services. The user uses a smartphone for viewing attractions on current region map and information about them and can plan the route in the region. The smartphone stores a user profile, which includes information about the user and his\her preferences. The presented reference model consists of the following services.

Information aggregation service is responsible for gathering information about attractions at the region using the accessible information sources in offline mode. For the information sources at the moment the OpenStreetMaps and Wikipedia are used. OpenStreetMaps is used as geo-data source for fetching information about attractions in the region. This information includes attraction name, position (latitude and longitude) and tag (attraction category such as museums, parks, etc.). After fetching an attraction list, service tries to gather detailed information from open-source databases. Usually, the service is looking for an attraction description and images.

The main purpose of the information aggregation service is gathering attraction information in selected regions, extracting data into database and providing databases to the content delivery service. Process of gathering information can be divided into two parts: bringing geo-data about attractions geolocation services and searching attraction details in accessible information services. As input, the service requires a region list. For each region, the service performs the following operations: district processing, attraction data searching and data writing into the database.

The service uses the developed searching module for information processing and aggregation. The reference model of search module is presented on Figure 2. The input of the attraction information search module is formed by the attraction details, such as title and coordinates given from external geolocation servers. As the search module output, the attraction article with description, keywords and images is given. For improving the quality of searching the following set of measures has been identified. At first, unnecessary words are removed from the attraction title, such as «monument», «tomb», etc. Secondly, module performs geo-search by coordinates, if it existed, and basic search by using attraction title. Main criterion for the attraction article choice correctness is the correspondence of the chosen article with the selected one, computed by using fuzzy search algorithms (with Levenstein distance metric).

Article candidates from geo- and basic- search are compared with each other and the most suitable variant is chosen. After comparison, the selected article is checked for correctness according to the following criteria: whether the article object is in the selected region, whether the content of the article of the original sight is suitable (for example, is the article describes person in case if as input module has monument of that person) and so on. If the check was successful, the module writes information

Fig. 1. Tourist trip planning system reference model

Fig. 2. Reference model of attraction searching module

about the article (title, short description, keywords, links to images) to the list and at the end of the work transfers data to the information aggregation service. At the end, information aggregation service stores information into a local region database.

Content delivery service acts as a storage of information and takes attraction region data from *information aggregation service*. User can download offline region cache from service for the selected region. Also, the content delivery service provides attraction information for the user in case if users does not download region cache for offline mode and Internet connection is available. The *trip planning service* uses a region map in the form of a graph from this service and information about the sights for constructing a route for the user. The *context service* takes information about attraction schedule for its work.

Recommendation service is responsible for ranking attractions, images, and descriptions for providing the tourist the best attractions to see and the best images and description of chosen attraction for acquaintance. To improve the quality of information presented to the tourist the list of attractions found should be ordered with respect

to a predicted degree of interestingness for the tourist as well as reachability (taking into account the current situation in the area). The service takes tourist ratings associated with each attraction by all tourists as an input and performs user-based collaborative filtering.

Trip planning service is interacting with *recommendation service*, *context service* and *content delivery service* for creating an attraction route for the user. The trip planning service is interacting with the following services: constructed region graph are gained from *content delivery service*, user and region context information are provided from *context service* and information about user preferences and attraction ratings are gathered from *recommendation service*.

The purpose of the context-oriented tourist trip planning service development is to increase the tourist attractiveness and total saving of the tourist time when visiting the attraction. The developed method provides attraction visiting plan for the tourist. The context includes the tourist location, travel speed and preferences, temporary tourist restrictions, season and current weather, city traffic and a schedule of attractions. For building the attraction visiting route, methods and algorithms of graph theory are used, since the set of attractions and the roads between them can be represented as a graph. Region graph can be described as a indirection graph (V, E) where every node $v \in V$ represents an attraction and every edge $e \in E$ represents a path between an attractions. Every path has a function f which is representing a cost (e.g. travel time between attractions). The context is described as a set of modifiers that can be applied to the path cost. It assumes the use of algorithms for finding the minimum path using the average path time between the selected sights, tak-

ing into account the current context situation (for example, traffic, preferred attraction types, weather situation, etc.).

The tourist route can be described as a sequence of visited attractions for the developed method. It is possible to apply parallel calculation of the best route for each segment for the route constructing acceleration. The context-oriented constructing route method involves re-evaluating of the region graph using different metrics for applying the new graph to various methods for finding the shortest route, such as ant colony optimization algorithm, tabu search algorithm, brunch and cut algorithm, etc. The method assumes the utilization of knowledge in machine-readable form, as well as the formalization and modelling of the current situation in the region.

Context service is designed for storing and working with a context information. The context has been split into two parts: user context and region context. User context contains information about location and movement of user; user time restrictions (e.g. tourist arrived to a region at the morning and have free-time until evening). Region context contains information about the current traffic situation; attractions schedule of region and information about current weather and season. Traffic and weather information can be obtained via appropriate services — *weather service* and *traffic service*. To create a tourist profile and a contextual information model of the region, it is suggested to use an ontological approach. It is needed to create an tourist profile that will allow describing the tourist himself, his preferences and the situation in the region. The approach used to describe the profile is well established in recent years and allows to precisely formalize the area under investigation.

4 Implementation

The prototype implementation is based on Tourist Assistant — TAIS [15]. The main feature of the implemented application is providing the information to users based on offline source such as sqlite database in contrast the TAIS that searches the information using online resources. The application provides information to user based on the content delivery service region list file developed from the information aggregating service. The user can download desired regions in a background and select one for utilization (Fig. 3).

The region sqlite database contains information about attractions (title, coordinates, keywords, images, etc.) and districts (name and polygon coordinates). The developed application searches the closet to user attractions and provide information about them; track and display the current region district and provided the full map of selected region with attractions. Map viewing (Fig. 4) is organized by using *osmdroid* library and using OpenStreetMaps data for displaying the current region. For improving the perception of a large number of attraction markers on the map, the clustering technique was used.

The information aggregation service is implemented using the Python language with additional packages for working with network, database, images, etc. As geo-data source the OpenStreetMaps is chosen and for attrac-

Fig. 3. Regions download and selection in the application

Fig. 4. Map viewer in the application

tion information extraction the Wikipedia is used. Input configuration file for the service is coded as JSON-file and contains list of regions, which are needed to be processed. Each region in the configuration file contains name (both original and english versions), region identification number in OpenStreetMaps servers, language code and types of attraction for search. The service accesses to OpenStreetMaps server data by using the Overpass API for gathering information about districts and region attractions. The inner Wikipedia API is used for gathering information about attractions.

After collecting the attraction information from the thread pool the information aggregation service writes information about attraction to the database. For each attraction, the service checks possibility and insert into the database information about available attraction (name, description, keywords, language, and images).

The sqlite-database models are presented on Figure 5. Table *Attraction* stores: english name, which is taken from information service (*name* field); name, which is taken from geo-service (*original_name* field); coordinates of place (*latitude* and *longitude* fields), general attraction description, which will be presented to user (*description* field); tag-information related to attraction (*tag* field); source type which is used for content generation (*source* field); translation flag notifier (*translated* field) and *url* field for storing information about url of content page. Image table contains: *attraction_id* for storing attraction identification number, *image* for keeping the image con-

Fig. 5. Database models for information aggregation service

tent, *path* field contains the image url, and *primary* field is a flag for Android application for using particular image as title.

Keyword table works in pairs with AttractionKeyword table to provide keyword support for attractions. *Name* field of Keyword table contains the keyword; *attraction_id* and *keyword_id* from AttractionKeyword table keep information about identification number of attraction and keyword accordingly. District and DistrictPolygon tables stores information about region districts. *name* field of District table contains district title, which is gathered from geo-service and DistrictPolygon table uses the *district_id*, *latitude* and *longitude* fields for storing information about polygon points and district affiliation. All points in the polygon are inserted clockwise.

All gathered information from the search module of the information aggregation service are divided into parts. For each part the following operations are done: attraction entry creation and correctness checking, attraction keywords and image processing.

The content delivery service is written using Python and acts like a web-service. For building web-service *Flask* micro-framework is used, *Gunicorn* is used as a WSGI HTTP server and *nginx* works as proxy-server for a whole application. Web-service for the current moment contains only two available paths: root route for region list file providing (/) and region sqlite database downloading route from service (/get/region/<id>), where *id* acts as region identification number). The region list file is a JSON-file with data about available regions. Every region contains information about its name, identification number, image url for the Android application, current size of sqlite database with region and date of gathering region information, which acts like a version number for the Android application.

5 Conclusion

The paper presents the concept of the smartphone-based tourist trip planning system and describes implementation some parts of it. The architecture of this system is a set of services that interact with each other and provide content to the user. The information aggregation service is responsible for gathering attraction information in regions, such as descriptions, coordinates, keywords, etc. Also, this service save all information into sqlite database, which is used by the Android application for providing offline full-fledged work on finding and displaying tourist information. The trip planning service is interacting with recommendation service, context service and content delivery service for creating an attraction route for user.

The research is funded by the Russian Science Foundation (project # 16-11-10253).

References

- [1] C.D. Huang, J. Goo, K. Nam, C.W. Yoo, Information & Management **54**, 757 (2017), smart Tourism: Traveler, Business, and Organizational Perspectives
- [2] J. Borràs, A. Moreno, A. Valls, Expert Systems with Applications **41**, 7370 (2014)
- [3] A.V. Smirnov, A.M. Kashevnik, A. Ponomarev, Personal Ubiquitous Comput. **21**, 297 (2017)
- [4] C.C. Chen, J.L. Tsai, Future Generation Computer Systems (2017)
- [5] S.Y. Yang, C.L. Hsu, Computers & Electrical Engineering **54**, 87 (2016)
- [6] S.Y. Yang, D.L. Lee, K.Y. Chen, C.L. Hsu, 49–52 (2016)
- [7] H.S. Chiang, T.C. Huang, Information Fusion **21**, 3 (2015)
- [8] M. Nilashi, K. Bagherifard, M. Rahmani, V. Rafe, Computers & Industrial Engineering **109**, 357 (2017)
- [9] I. Cenamor, T. de la Rosa, S. Núñez, D. Borrajo, Expert Systems with Applications **69**, 1 (2017)
- [10] Y. Sun, H. Fan, M. Bakillah, A. Zipf, Computers, Environment and Urban Systems **53**, 110 (2015), special Issue on Volunteered Geographic Information
- [11] J. Shen, C. Deng, X. Gao, Neurocomputing **173**, 789 (2016)
- [12] S. Kotiloglu, T. Lappas, K. Pelechrinis, P. Repoussis, Tourism Management **62**, 76 (2017)
- [13] D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, N. Vathis, C. Zaroliagis, Expert Systems with Applications **42**, 7303 (2015)
- [14] Novye informacionnye tehnologii v avtomatizirovannyh sistemah 75–80 (2016)
- [15] A. Smirnov, A. Kashevnik, A. Ponomarev, N. Teslya, M. Shchekotov, S.I. Balandin, 40–51 (2014)