Recommendation of Short-Term Activity Sequences During Distributed Events

Diana Nurbakova, Léa Laporte, Sylvie Calabretto, Jérôme Gensel

Recommendation of Short-Term Activity Sequences During Distributed Events

Diana Nurbakova[1]*, Léa Laporte[1], Sylvie Calabretto[1], and Jérôme Gensel[2]

[1] LIRIS UMR 5205, INSA Lyon, Villeurbanne, France
{diana.nurbakova,lea.laporte,sylvie.calabretto}@insa-lyon.fr
[2] University Grenoble Alpes, CNRS, LIG, Grenoble, France
jerome.gensel@imag.fr

Abstract
The amount of social events has increased significantly and location-based services have become an integral part of our life. This makes the recommendation of activity sequences an important emerging application. Recently, the notion of a *distributed event* (*e.g.* festival or cruise) that gathers multiple competitive activities has appeared in the literature. An attendee of such events is overwhelmed with numerous possible activities and faces the problem of activity selection with the goal to maximise satisfaction of experience. This selection is subject to various uncertainties. In this paper, we formulate the problem of recommendation of activity sequences as a combination of personalised event recommendation and scheduling problem. We present a new integrated framework to solve it and two computation strategies to analyse the categorical, temporal and textual users' interests. We mine the users' historical traces to extract their behavioural patterns and use them in the construction of the itinerary. The evaluation of our approach on a dataset built over a cruise program shows an average improvement of 9.7% over the state-of-the-art.

Keywords: Spatio-temporal activities, recommendation of activity sequences, uncertainty

1 Introduction

This work focuses on the emerging and challenging problem of the recommendation of activity sequence during distributed events (*e.g.* cruises or festivals) that consist of multiple short-term and highly competitive activities. These events attract millions of attendees and the interest in taking part in them is substantially growing. Let us consider the case of a cruise. According to the Cruise Industry Overview[1], about 23 millions passengers were expected to cruise globally in 2015, and the Cruise Line International Association expected 24 millions of cruisers in 2016[2].

*Funding for this project was provided by a grant from la Région Rhône Alpes
[1]http://www.f-cca.com/downloads/2015-Cruise-Industry-Overview-and-Statistics.pdf
[2]http://www.cruising.org/docs/default-source/research/2016_clia_sotci.pdf?sfvrsn=0

Uncertainty sources	Typical uncertainties	Description
Data	Implicit feedback	The data collected and analysed are not obtained by means of users' direct interaction with the system, there is no precise knowledge about users' preferences or users' ratings of a set of alternatives
	Orientation to the future	The recommendation items (*i.e.* activities) occur in the future implying the lack of information about other users' experience
	Constrained interest	The recommendation items are short-lived and are available in a specific time and a specific place, which makes a user prefer one activity to another for a given time slot
	Attendance bias	A user may join an activity that is not of his/her interest, and not all activities that are missed by a user do not represent personal interest for him/her
Assumption	Maximisation of satisfaction	User's satisfaction is often considered being accumulative, *i.e.* the more activities you join, the more satisfaction you get
Parameters	Parameters estimation	The objective functions that are used to estimate the model parameters and evaluate the quality of the solution
Model	Profiling	The type of representation of both the features of alternatives and the users' preferences, and the way to maintain them up to date impacts the understanding of ongoing processes of users' decision making
	Preference score	The features selection and extraction
	Itinerary construction	The objective function of the scheduling of activity sequences

Table 1: Sources of uncertainty in recommendation of activity sequences

Cruise lines offer various activities in order to provide their customers with the best service. Consequently, while on board, travellers face the problem of activity selection. Let us consider the following example: Scoby is enjoying his holidays on board of a 7-night Caribbean cruise ship. Every day, he has to make up his mind which activities to choose among a hundred offered with the average duration of 45 min. And at every given time, there are about 5 activities going on[3]. Then, what is the best way to plan the day in order to get as much fun as Scoby can?

The decision making process that is associated with the selection of activities to perform, especially in the context of a big event as a cruise journey is not simple. There is a variety of sources and types of uncertainties associated with the recommendation of activity sequences. A number of them could be assigned to Parameter uncertainties group, according to classification of uncertainties given in [8]. These uncertainties arise from the incomplete and inaccurate data that do not fully represent the phenomenon. Table 1 describes some of them.

The problem of activity sequences recommendation rises in various fields, namely, in scheduling during conferences, festivals (*e.g.* ComicCon), and big distributed events [6], holidays/tour planning [2], and mobile crowdsourcing [1]. In this paper, a cruise is considered as an application scenario. We consider indoor and limited outdoor environments, which implies that the travel time is of less importance than in the case of tour and Points of Interest (POI) recommendation.

[3]The statistics about the average duration of activities and the number of simultaneous activities are taken from the dataset described in Section 4.1

The main challenges of the recommendation of activity sequences rely in the fact that activities are unique, happening in future and short-term. Thus, every time a user wants to choose an activity to perform, he/she has to detect an activity he/she may prefer among all the alternatives happening during the same timeslot, taking into account that he/she may not be able to join another activity of interest due to the availability/time constraints. Therefore, the objective is not only to define users' interest in upcoming events, but to provide per day personal program (itinerary) of activities. It implies the necessity to recommend activities with no explicit feedback and any external information (*e.g.* reviews) dealing with highly uncertain users' preferences, and to ensure that a user will be able to attend all the selected activities on time. We assume that we have access to users' past visited locations and the timestamps.

In this paper, we aim at answering the following research question: *How can we predict and maximise users satisfaction with a sequence of activities, given their past experience?* We decompose this problem into three specific research questions:

RQ1: How can we tackle the uncertain users' preferences and evaluate the users' interest for an activity happening in future, given little information about it?

RQ2: How can we retrieve users behavioural patterns from historical data?

RQ3: How can we organise activities into a sequence that maximises the user's satisfaction while taking into account spatio-temporal constraints and the sequential nature of activities?

The present study is related to the following research fields: POI recommendation, Event recommendation, Trip recommendation and Itinerary construction. *POI recommendation* aims at providing users with lists of top-k points of interest according to the users' visiting preferences, considering the geographical component having the biggest impact [3]. Some works also exploit categorical or social [12] influences. The main limitation of the POI recommendation techniques is that they do not consider POI availability constraints, travel time between POIs, visit duration, user's time budget, and visiting order. [4] tackles the problem of top-n *recommendation of events* in Event-Based Social Networks. In addition to textual and collaborative signals, several contextual interests have been exploited (*i.e.* social signal, users' geographical and temporal preferences) that are combined for learning to rank events. Similar to the POI recommendation techniques this method does not consider limited availability of events, and event attendance order. *Trip recommendation* aims at providing a user with a sequence of POIs to visit, by taking into account spatial and temporal constraints. Most of recent studies [11] decompose the problem in two parts: (1) estimation of individual scores for each POI, (2) itinerary construction [2]. Contrary to the problem treated in this paper, trip recommendation problem does not consider the uniqueness of activities. *The itinerary construction* problem is often modelled as an instance of Orienteering Problem (OP) or its variations [10]. OP aims at determining a Hamiltonian path limited by the time budget that maximises the collected score by visiting vertices. OP ignores the way the scores of vertices have been calculated.

In this paper, we propose an integral approach to address the problem of recommendation of activity sequence. The key contributions of our work are: a formal definition of the problem of activity sequence recommendation; an integral method to compute activity scores; an algorithm to retrieve users behavioural patterns based on the past performed activities; an integration of users typical transition patterns on the itinerary construction. Moreover, the user study has been conducted in order to collect the relevant data.

The remainder of the paper is organized as follows. Section 2 defines the problem and introduces the notations. In Section 3, we present our approach that consists of three parts: activities score computation, sequential pattern mining, and itinerary construction. Section 4 describes the dataset used for the evaluation and reports the obtained results. Section 5 concludes the paper and presents our future work.

Activity: *The Comedy & Hypnosis of Ricky Kalmon*
Location: Walt Disney Theatre, $l = (0, 880, 0)$; **Time window**: Day 3, 23:00-23:45, $t = (1435014000, 1435016700)$; **Duration**: $\delta = 2700$; **Categories**: Adults, Variety Show **Description**: Featuring the Comedy & Hypnosis of Ricky Kalmon, as he entertains you in this adult exclusive show.

Table 2: An example of an activity

2 Problem Definition

In this section, we define the notations used throughout the paper and formulate the problem.

An *activity* $a = \langle l, t, \delta, c, d \rangle$ is an event that a user u can attend or take at some geographically located point in a particular time window, and is characterised by its location (latitude, longitude, altitude), $l = (x, y, z)$, the time window (start time t_s and end time t_e) of its availability $t = (t_s, t_e)$, its duration δ, a vector of categories associated with the activity $c = (c_1, ..., c_k)$, and a description d. An example of an activity is given in Tab. 2. $A = \{a_1, a_2, ..., a_N\}$ is the set of all available activities.

An *activity sequence* (or *itinerary*) $\xi^u = \left(a^u_{(s)}, ..., a^u_{(s+k)}\right)$, where $1 \leq s \leq s + k \leq N$, is an ordered series of activities for user u, accounting for spatio-temporal constraints[4] such as the Activity availability constraint and the Time budget constraint. *Activity availability constraint* specifies that an activity a_i can be performed only during its availability time, limited by its start time t_s and end time t_e, *i.e.* $t_s \leq start(a_i) \leq t_e$. Here, $start(a_i)$ denotes the time when a user starts performing the activity a_i, as he/she may join the activity a_i when it becomes available and once he/she has finished to perform the previous activity and moved to the location of the current one, *i.e.* $start(a_{(i)}) = \max\{start(a_{(i-1)}) + \delta(a_{(i-1)}) + time(a_{(i-1)}, a_{(i)}), t_s(a_{(i)})\}$, where $time(a_{(i-1)}, a_{(i)})$ is the travelling time to go from the location of activity $a_{(i-1)}$ to the one of $a_{(i)}$. *Time budget constraint* limits the total time needed to follow all the activities within an itinerary, including activities duration and travelling time with the given time budget T_{max}. It may be defined by a user or the fixed value may be used (*e.g.* day time).

A *satisfaction function* $r\left(a_i, u\right),\ r : A \longrightarrow \Re^+$ characterises the match of the activity a_i with the interest of a user u for this activity. The satisfaction with an itinerary ξ^u for a user u is defined as the sum of scores of activities within the itinerary, $r\left(\xi^u, u\right) = \sum_{a_i \in \xi^u} r\left(a_i, u\right)$.

The problem of activity sequences recommendation consists in finding an itinerary ξ^u that maximises the user's satisfaction $r\left(\xi^u, u\right)$, given a user u and the set of activities $A = \{a_i\}_{i=\overline{1,N}}$.

We make the following *assumptions* throughout this paper: (1) *Non-stop fun*: in the context of users' attendance of a distributed event, their goal is to get the maximum satisfaction from overall experience. (2) *Traceability of users*: the log of users' past experience that consists of geospatial coordinate sets and timestamps is available. (3) *Moving around in space*: the travelling time of the users between locations is a function of distance. We assume that all the users move with the same constant pace.

3 Solution Methods

We propose an integrated framework for activity sequences recommendation that exploits the users' interests, sequential influence, spatial and temporal constraints. It consists of three main

[4]Here we use brackets in subscripts to indicate that the elements within a sequence are ordered, so that they can be differentiated from the elements of the whole set.

parts: 1. Computation of personalised scores for each activity. 2. User's behavioural pattern mining. 3. Itinerary construction using data provided by the previous steps.

3.1 Computation of Personalised Scores

We now investigate our **RQ1**. The activities we consider are unique and can be performed only during their time window. There is no rating of upcoming activities that could be used in order to predict the scores. There are users' traces available (see *Traceability of users* assumption) that in combination with activities program may be used to retrieve the activities joined by a user. Due to the lack of explicit feedback about the degree of user's interest in an attended activity, user's attendance is regarded as positive feedback. In this work, we propose to explore content, categorical and temporal influences in order to estimate users scores of activities.

Content Influence: We consider textual influence by applying a bag-of-words model to the descriptions of activities. We use TF-IDF representation of activities $\vec{e}$ in order to build positive and negative user profiles. The positive user profile U_{pos} consists of summarised TF-IDF vectors of activities performed by the user in the past, while the negative user profile U_{neg} is built over non-performed activities. The content-based score of an upcoming activity is then computed as a linear combination of cosine similarity measures in such a way that the similarity to non-performed activities is used as a penalty, *i.e.* $\hat{r}_{cb}(a,u) = \alpha_u \cdot cos(U_{pos}, \vec{e}) - \beta_u \cdot cos(U_{neg}, \vec{e})$. The parameters α_u and β_u are defined for a given user as optimisation parameters of the loss function with regularisation over the 10-fold cross-validation sets.

Categorical Influence: We suggest to use the categories of activities already performed by the user. Each activity a is associated with a list of categories $\mathcal{C}_a = \{c_j\}$. Thus, for each user and each category, we compute the frequency of a category based on the user's past activities as $freq(c_i, u) = \frac{|A_{u,c_i}| \cdot w_a}{|A_u|}$, where $|A_{u,c_i}|$ is the number of activities performed by user u that belong to the category c_i, $w_a = 1/|\mathcal{C}_a|$ is the weight calculated as a ratio to the number of categories an activity $a \in A_{u,c_i}$ is associated with, and $|A_u|$ is the number of all activities performed by user u. Given a user and an activity, we then estimate an activity categorical score as the sum of corresponding categorical frequencies, *i.e.* $\hat{r}_{cat}(a,u) = \sum_{\mathcal{C}_a} freq(c_j, u)$.

Temporal Influence: Another factor that might have an impact on users' decision on joining an activity is the temporal aspect, *i.e.* when an activity takes place. The intuition behind is that there are several parts of a day when a person is more active. To formalise this intuition, we split a day into 15 minutes long timeslots. We then represent each activity as a binary 1×96-dimensional vector t_a with a vector component set to 1 if the availability time window of an activity includes that timeslot. A user is then represented as the binary vector built over the union of the timeslots of his/her past activities t_u . The temporal score is defined based on the temporal relations between a timeslots vector of an upcoming activity and a user's temporal profile, as: $\hat{r}_{time}(a,u) = \begin{cases} 1, & \text{if } \; t_a \cap t_u \\ 0.5, & \text{if } \; t_a \cap \{t_u - 1 \cup t_u + 1\} \\ 0.1, & \text{otherwise} \end{cases}$. We adapt the scores proposed in [5].

To further improve the effectiveness of recommendation, we propose to make use of the three aforementioned influences by combining them in two ways. First, we propose to define *Hybrid Score* (LinC) as follows: $\hat{r}_{hyb}(u,a) = (\gamma_u \cdot \hat{r}_{cb}(u,a) + \delta_u \cdot \hat{r}_{cat}(a,u)) \cdot \hat{r}_{time}(a,u)$, where γ_u and δ_u are defined for a given user as optimisation parameters of the loss function with regularisation over the 10-fold cross-validation sets. Secondly, we suggest to fit the logistic regression classifier with categorical, textual and temporal scores as parameters, and consider the probability of assigning an activity to the class 1 as *Logistic Regression Score* (LogR).

We suggest to organise scores estimation process under two strategies that we call *Strategy*

Strategy 1: All-at-once ;
Input: User's Attendance Matrix $\mathcal{M}$, New activities $NewEvent$;
Output: Activities scores $\mathcal{R}$;
Calculate $\mathcal{R}(NewEvents, \mathcal{M})$
	Strategy 2: Day-after-Day ;
	Input: User's Attendance Matrix $\mathcal{M}$, New activities $NewEvent$, Number of past days $PastDays$, Total number of days $DayNum$;
	Output: Activities scores $\mathcal{R}$;
	Initialisation $\mathcal{M}^{(0)} \leftarrow \mathcal{M}$;
	for $i \leftarrow PastDays$ **to** $DayNum$ **do**
	  Calculate $\mathcal{R}^{(i)}(NewEvent^{(i)}, \mathcal{M}^{(i)})$;
	  $\mathcal{M}^{(i)} \leftarrow \mathcal{M}^{(i)} \cup \mathcal{R}^{(i)}$;
	  $i \leftarrow i + 1$
	end

Table 3: Strategy 1 and 2 to estimate the interest scores of activities

1 ('All-at-Once') and *Strategy 2* ('Day-after-Day') strategies. Table 3 depicts the pseudocodes of the algorithms. Thus, Strategy 1 considers all upcoming activities at once and estimates their scores with respect to the above models, while Strategy 2 estimates activities scores on a daily basis and enriches user's experience of past events with estimated scores that are further used as user's historical data at the next iteration.

3.2 User's Behavioural Pattern Mining

This section addresses our **RQ2**. Our goal is to retrieve the most typical transitions between consecutive activities, *i.e.* users' activities sub-sequences or behavioural patterns. Two activities are considered to form a sequence if the time interval between the end of the first and the start of the second one is within a fixed threshold [13].

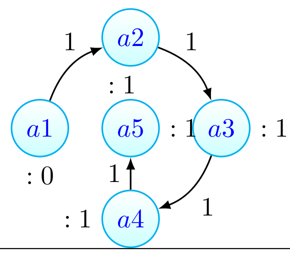
We propose to construct an *activity-activity transition graph* (A^2TG) and a *category-category transition graph* (C^2TG) by extending the concept of location-location transition graph used to model the transitions between POIs [13]. The A^2TG models the transitions between activities determined by their time and location. Due to the uniqueness of activities, the A^2TG cannot be used directly for the estimation of the transition probabilities between new activities. Thus, we use it to construct the C^2TG that models generalised transitions between categories.

Based on users historical traces and the program of a distributed event we retrieve users activity sequences and construct the A^2TG. Its nodes, $V = \{a_1, ..., a_N\}$, correspond to the activities undertaken by a user. In contrast to [13], we assign the number of incoming edges $InCount(a_i)$ since we suppose that the user's satisfaction depends more on the previous experience rather than on the future one. The edges stand for transitions between activities and are associated with the number of transitions, $TransCount(a_i \rightarrow a_j)$. We further propose to move to the category level under the assumption that the categories of activities are known.

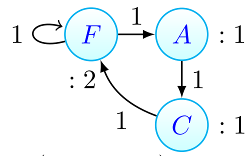
C^2TG is constructed in a similar way. Its nodes represent categories c_i associated with undertaken activities and are characterised by the number of incoming edges, $InCount(c_i)$, calculated as follows: $InCount(c_i) = \sum\limits_{a_j \in c_i} InCount(a_j)$. The edges stand for transitions between categories and are associated with the number of transitions, $TransCount(c_i \rightarrow c_j)$, which is calculated using the $TransCount$ of corresponding activities as follows: $TransCount(c_i \rightarrow c_j) = \sum_{\substack{a_k \in c_j, \\ a_g \in c_i}} TransCount(a_k \rightarrow a_g)$. Given C^2TG, we estimate the probability of transition from

Activity sequence: {*Crafts: Door Hangers* → *Pictionary Challenge* → *Singles' Lunch* → *Goofy* → *The Comedy and Hypnosis of Ricky Kalmon*}
Categories sequence: {*Fun for All Ages* (F) → *Fun for All Ages* (F) → *Adults* (A) → *Characters* (C) → *Fun for All Ages* (F)}

A^2TG:



C^2TG:



$P_T(F \rightarrow A) = 1$; $P_T(F \rightarrow F) = 1/2$; $P_T(A \rightarrow C) = 1$; $P_T(C \rightarrow F) = 1/2$

Table 4: Example of A^2TG, C^2TG and P_T. Node labels stand for $InCount$.

one category to another as follows: $$P_T(c_i \rightarrow c_j) = \begin{cases} \frac{TransCount(c_i \rightarrow c_j)}{InCount(c_j)}, & \text{if } InCount(c_j) \neq 0 \\ 0, & \text{if } InCount(c_j) = 0 \text{ and } c_i \neq c_j \\ 1, & \text{if } InCount(c_j) = 0 \text{ and } c_i = c_j \end{cases}$$

To estimate the transition probability between two activities, the reverse process has to be undertaken. An example of A^2TG, C^2TG and transition probabilities P_T is shown in Tab. 4.

3.3 Itinerary construction

In this section, we focus on **RQ3**. Given a user, a set of activities A defined by their locations given by coordinates, the time windows of their availability, the duration, personalised interest scores, the travel time between a pair of locations, the fixed starting and ending point, we want to find a sequence of activities that maximises the overall collected score, *i.e.* user's satisfaction from undertaken activities. Therefore, the itinerary construction problem can be formulated as the Orienteering Problem with Time Windows (OPTW) [9]. Vansteenwegen *et al.* [9] proposed the Iterated Local Search (ILS) algorithm to solve OPTW. ILS is a heuristic algorithm that iteratively searches a node to be included in the current path that will maximise the total score of the itinerary. A feasibility check is performed to ensure that the insertion of a new node would not make any already included visit violate its time window constraint. At each iteration, ILS adds a feasible node a_k with the highest $Ratio_k = \frac{\hat{r}_k^2}{Shift_k}$, *i.e.* the ratio of squared node score $\hat{r}_k$ to the time shift $Shift_k$ needed for insertion of the activity a_k into the path caused by the insertion of the node. Due to the limited space, we do not include the mathematical formulation of the OPTW problem here (for more details, refer to [9]). We propose an adaptation of the ILS algorithm in adjusting the value of $Ratio_k$ with the transition probability from the previous activity a_{k-1} to the current one a_k: $Ratio_k = \frac{\hat{r}_k * P_T(a_{k-1} \rightarrow a_k)}{Shift_k}$, where $\hat{r}_k$ is the score of the activity a_k, $P_T(a_{k-1} \rightarrow a_k)$ is the transition probability from the activity a_{k-1} to a_k. We denote our transition probability enhanced ILS as ILS_TP. The intuition behind is that incorporating the sequential characteristics of already undertaken activities could enhance the prediction power of itinerary recommendation.

Statistic	# Activities	# Days	# Users	# Locations	# Categories
Value	595	7	23	47	52

Table 5: Dataset Statistics

4 Results and Discussion

4.1 Dataset

To the best of our knowledge, there is no available dataset for personalised recommendation of itineraries during distributed events. Therefore, we have created a dataset and have performed the evaluation offline. The dataset simulates users' attendance of a cruise and was created as follows. We have collected the plannings of activities from a Disney's 7-nights cruise[5]. We have created an online survey that consisted of 3 parts: (I) Personal Profile (*i.e.* completion of personal information, *e.g.* gender, cruise experience, type of group a user is travelling with) (II) Interest in Activities (*i.e.* given a list of activities, respondents have indicated their interest in each activity on a 5-point scale ranging from 1-Never to 5-Won't Miss) (III) Daily Itineraries (*i.e.* given a list of activities with their time slots, users have indicated their intent to join or not a particular activity for each day out of 7, *i.e.* they have organised the activities into a day-wise itineraries). Thus, 23 contributions have been collected. Dataset statistics are given in Tab. 5. If not provided, we have enriched activity descriptions of movies and characters using IMDb (http://www.imdb.com/) and Wikia (http://disney.wikia.com/) respectively.

4.2 Evaluation Protocol

The itineraries provided by users have been split into two parts: the first was used as historical data in order to extract the sequential patterns and users' interests, and the second was used as the test set to evaluate the results obtained with our approach. Partitions of various sizes (on the day level ranging from 1 to 6) have been explored. The evaluation process was performed in two steps. First, we have evaluated the accuracy of estimated activities scores using four metrics: MAE, RMSE, Precision at rank k [7] and AUC (area under the ROC-curve). The rank k of Precision was defined for each user and set to the average number of daily activities performed by a given user in the past days. The motivation behind such a setting lies in the fact that different users have different density of activities. The lower values of MAE, RMSE and the higher values of Precision and AUC are the better. Ten cases were considered w.r.t. the method to compute activity score: Category-based, Content-based, Time-based, Hybrid and Logistic Regression under two strategies (see Section 3.1). The results have been compared with the ground truth composed of the binary attendance of activities. Second, we have evaluated the itinerary construction by comparing ILS_TP and the original ILS with the sequences that were annotated by users (Ground Truth). We consider the ratio of recommended activities matching Ground Truth to the length of the Ground Truth as the evaluation metric ("Similarity").

4.3 Results

We implemented the estimation of the activities scores and the transition probabilities using Python 3.5.2, and itinerary construction using GNU Octave. Firstly, we evaluated the accuracy of obtained activities scores (see Fig. 1). It can be seen that the exploitation of multiple factors

[5]http://disneycruiselineblog.com/2015/07/personal-navigators-7-night-eastern-caribbean-cruise-on-disney-fantasy-itinerary-a-june-20-2015/

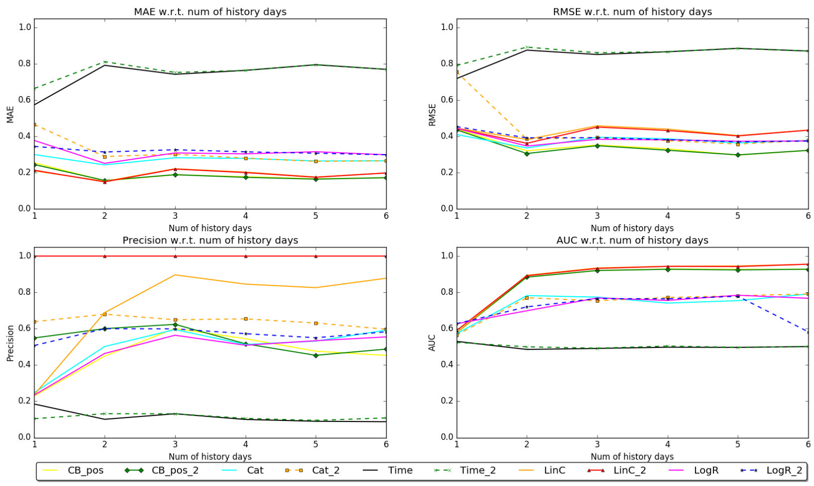


Figure 1: Values of evaluation metrics of the quality of activities score estimation.

History Days	1	2	3	4	5	6	Average
Strategy 1	6.1	3.4	6.5	6.2	9.9	11.3	7.4
Strategy 2	25.0	10.3	4.4	13.5	14.3	11.3	13.5

Table 6: Improvement of ILS_TP over ILS in terms of similarity to Ground Truth, %

increases the predictive power of the model. Thus, the Hybrid scores outperform the others in terms of precision and AUC, while having the second result in terms of MAE. It can be noticed that the performance is enhanced when using the Strategy 2 (the dash lines in the plots). These results were expected since Strategy 2 applies the enrichment of historical data. We can see that considering 2 history days and more dos not affect much the performance of the algorithms. We find that the results reflect well the users' intentions, in particular the Hybrid score-Strategy 2. Thus, we selected Hybrid scores under Strategy 1 and 2 for the next evaluation step. We used them as input data for ILS and ILS_TP algorithms. We varied the number of history days from 1 to 6. The obtained results show that incorporating the transition probability into the itinerary construction improves the performance. Thus, the average improvement of ILS_TP over ILS is 7.3% for Hybrid score-Strategy 1 and 14.1% for Hybrid score-Strategy 2 (see Tab. 6). The variations of performance w.r.t. the number of history days could be explained by the introduction of transition probabilities that are computed over the users' past experience.

5 Conclusion

Here we studied the problem of the recommendation of activity sequences during distributed events which is highly challenging as it combines the problem of item recommendation and the scheduling problem. We proposed an integrated framework for the recommendation of unique activity sequences that exploits categorical, textual and temporal influences to estimate users' interest scores in activities, and makes use of sequential influence in order to recommend a better itinerary. Two strategies of calculation of activity score were presented. We evaluated our approach on a dataset created by conducting a user study that simulated the users' attendance of a cruise. The experiments show that our solution outperforms state-of-the-art algorithm. The described framework allows mitigating the impact of uncertainties stemming from uncertain preferences, imprecise and unclear relationship between the characteristics of activities and interests of an attendee. It is to note that users' interest in an activity does not always result in joining it and vice versa. Thus, according to our user study, 15.73% of activities were selected by users as performed (*i.e.* marked as *'going'* by respondents), but marked as non representing interest; while 58.12% of activities where marked as interesting but were not attended. In future work, we plan to explore different research directions. First, other types of influence on the interest score of activities could be investigated, including users demographics, or the group that a user performs activities with. Second, we plan to incorporate multiple time windows and multiple locations into the itinerary construction part of our approach. Finally, we would like to use the crowdsourcing as an evaluation tool: 1. to rate the activities according to interests. 2. to make a planning of activities to perform. 3. to evaluate recommendation results.

References

[1] A. S. Fonteles, S. Bouveret, and J. Gensel. *Web and Wireless Geographical Information Systems: 14th International Symposium, W2GIS 2015, Proceedings*, chapter Opportunistic Trajectory Recommendation for Task Accomplishment in Crowdsourcing Systems, pages 178–190. 2015.

[2] D. Gavalas, Ch. Konstantopoulos, K. Mastakas, and G. Pantziou. Mobile recommender systems in tourism. *J. Netw. Comput. Appl.*, 39:319 – 333, 2014.

[3] X. Li, G. Cong, X.-L. Li, T.-A. Nguyen Pham, and Sh. Krishnaswamy. Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 38th ACM SIGIR*, pages 433–442, 2015.

[4] A. Macedo, L. Marinho, and R. Santos. Context-aware event recommendation in event-based social networks. In *Proceedings of the 9th ACM RecSys*, pages 123–130, 2015.

[5] J. Sang, T. Mei, and Ch. Xu. Activity sensor: Check-in usage mining for local recommendation. *ACM Trans. Intell. Syst. Technol.*, 6(3):41:1–41:24, April 2015.

[6] R. Schaller, M. Harvey, and D. Elsweiler. Recsys for distributed events: Investigating the influence of recommendations on visitor plans. In *Proceedings of the 36th ACM SIGIR*, pages 953–956, 2013.

[7] G. Shani and A. Gunawardana. *Recommender Systems Handbook*, chapter Evaluating Recommendation Systems, pages 257–297. 2011.

[8] A. Tchernykh, U. Schwiegelsohn, V. Alexandrov, and El ghazali Talbi. Towards understanding uncertainty in cloud computing resource provisioning. *Procedia Computer Science*, 51:1772 – 1781, 2015. International Conference On Computational Science, {ICCS} 2015 Computational Science at the Gates of Nature.

[9] P. Vansteenwegen, W. Souffriau, G. Vanden B., and D. Van Oudheusden. Iterated local search for the team orienteering problem with time windows. *Comput. Oper. Res.*, 36(12):3281–3290, December 2009.

[10] Pieter Vansteenwegen, Wouter Souffriau, and Dirk Van Oudheusden. The orienteering problem: A survey. *Eur. J. Oper. Res.*, 209(1):1 – 10, 2011.

[11] Ch. Zhang, H. Liang, K. Wang, and J. Sun. Personalized trip recommendation with poi availability and uncertain traveling time. In *Proceedings of the 24th ACM CIKM*, pages 911–920, 2015.

[12] J.-D. Zhang and Ch.-Y. Chow. Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th ACM SIGIR*, pages 443–452, 2015.

[13] J.-D. Zhang, Ch.-Y. Chow, and Y. Li. Lore: Exploiting sequential influence for location recommendations. In *Proceedings of the 22Nd ACM SIGSPATIAL*, pages 103–112, 2014.