Recommendation of Activity Sequences during Distributed Events

Diana Nurbakova

Recommendation of Activity Sequences during Distributed Events

Diana Nurbakova*
LIRIS, INSA Lyon
Villeurbanne, France
diana.nurbakova@insa-lyon.fr

ABSTRACT

Leisure activities constitute an important part of our life. Nowadays, the offer of activities to undertake is constantly growing. This can be easily seen not only by the increasing number of social events created and promoted on Facebook, Couchsurfing, etc., but also by the appearance of specialised online services and event-based social networks, such as Meetup, Eventbrite, etc. Moreover, multi-day events (*e.g.* conventions, festivals, cruise trips, exhibitions), to which we refer to as distributed events, attract thousands of participants. Their attendees are often overwhelmed with the amount of available options. Recommender systems appear as a common solution in such a context. In this Ph.D. project, we formulate the problem of recommendation of activity sequences and aim at providing an integrated support for users to create a personalised itinerary of activities in order to facilitate their decision making process which events to join. Such assistance is expected to bring a positive impact on well-being and satisfaction with life of individuals.

CCS CONCEPTS

• **Information systems** → **Personalization**;

KEYWORDS

Recommender systems, event recommendation, itinerary recommendation, personalisation, users psychological profiles

1 INTRODUCTION

The selection of leisure activities is not always a trivial task. The process becomes even more complicated when it comes to organising one's time and creating a custom schedule of events at a festival, cruise, or any other big event that gathers multiple smaller events. We refer to the latter as *distributed events* [13]. The problem of selection of the 'best' items faced by the attendees is of interest of Recommender systems. Let us consider two following scenarios.

Scenario 1. Didi is planning her weekend of 2-4 December 2016 (see Fig. 1). She has preselected 8 activities that she fancies to do but she struggles with the selection. When considered and rated individually, all these activities represent high interest to her (see the five-star rating given next to activity names on Fig. 1). Suppose, Didi's final choice is to go to Forró Wochenende 2016. The independent interest judgement of this selection is $\hat{s}(a_1) = 5$. An alternative option could be to chain other activities, *e.g.* $\xi = a_3 \rightarrow a_6 \rightarrow a_7 \rightarrow a_8$. If we consider that the total interest score of such a chaining equals the sum of its parts, then $\hat{s}(\xi) = 20$. Even though $\hat{s}(\xi) > \hat{s}(a_1)$, the final selection is made in favour to a_1, which indicates that not only independent judgements are considered. One can also note that the activities are overlapping in terms of time (*e.g.* $< a_3\ vs.\ a_4 >$, or $< a_1\ vs.\ others >$). This implies that they cannot be performed all, which amplifies the need for selection. Thus, while recommending a sequence of activities, their time availability should be considered.

Scenario 2. Didi is going to Comic-Con International: San Diego. It is one of the biggest multi-day conventions primarily focused on comic books and related culture. Each year, it offers about 1,900 events of the average duration of 108 minutes. Ti estimate the density of Comic-Con events (see Fig. 2), we divide a day into 15-minute timeslots and calculate the number of events occurring at each of them. We can see that the number of competitive events is 37 in average, with the maximum of 112. This makes it hard for attendees to select events and organise their time to be sure to perform the maximum of the activities they would enjoy. Recommender systems are powerful assistance tools in such a decision-making process.

We identified the following challenging research questions:

RQ1 ► What constitutes the problem of recommendation of **s**patio-**t**emporal **a**ctivity **s**equences (STAS)?
- How to define the STAS problem?
- What makes it different from other recommendation problems, in particular, Event Recommendation and Trip recommendation? And what are the similarities?

RQ2 ► What drives the selection of activities by individuals?
- What makes an individual to select one activity in a given time interval and not the other?
- What are the types of influence on user's interest one can consider when dealing with leisure activities?

RQ3 ► Do psychological profiles of an individual define his/her preferences for leisure activities?
- How to implicitly acquire the user's psychological profiles from their past engagement in leisure activities?
- To which extent do psychological profiles of an individual define his/her preferences for leisure activities?

RQ4 ► How to provide a user with the best support during a distributed event via personalised recommendation?
- How to take into account a set of constraints (time availability, travel time, etc.) in order to return to a user a sequence of activities he/she will be able to perform?
- How to acquire the relevant preference of an activity?

The main goal of this Ph.D. project is to provide *an integrated support* for users to create *a personalised itinerary of activities* and *a list of events to join*, in order to facilitate their selection process.

On our way to achieve this goal, the following **contributions** are expected. In the list below, we use the marker symbols to indicate the current state of an achievement: a checked box ☑ for

*D. Nurbakova was mainly supported by Région Auvergne-Rhône-Alpes. Additional support was received from the Franco-German University as part of PhD-Track Program entitled "Multimedia Distributed and Pervasive Secure systems".

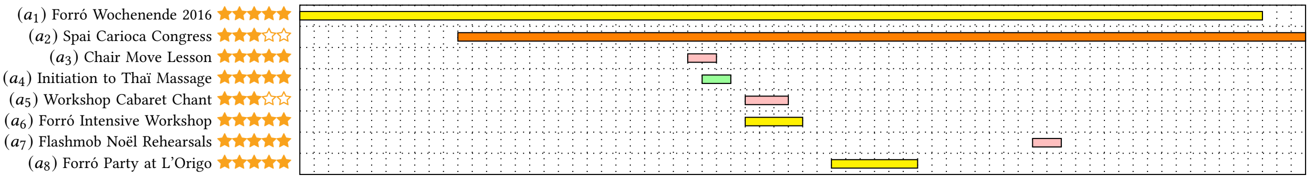

Figure 1: A choice of activities for a weekend during the time frame 1 December 21:00 - 5 December 01:00 (time windows of their availability). The bar colours indicate categories of activities: yellow - <Dance →Forró>, orange - <Dance →Samba de Gafieira>, pink - <Dance →Chair Dance>, green - <Well Being →Massage →Thaï>.

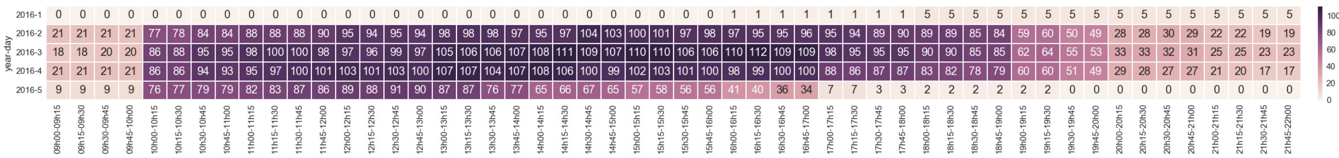

Figure 2: Heatmap of the overlapping events at Comic-Con 2016 with respect to 15 min long timeslots from 6am to 10pm.

accomplished work, an empty box □ for work in progress. The clock symbol ◷ indicates a work which is currently under review.

- ☑ **C1** Definition of **STAS**, overview of related recommendation problems and their solutions, classification of types of influence used for estimation of user's interest scores. ◷
- ☑ **C2** Design and implementation of **a n**ew **a**pproach for **s**hort-**t**erm **a**ctivity **s**equences and **i**tineraries recommend**a**tion during distributed events, **ANASTASIA** [7].
- □ **C3** Design and implementation of a psychologically-driven approach for recommendation of events.
- □ **C4** Design a new model for **p**ersonalised **r**ecommendation of **i**tineraries during **d**istributed **e**vents that takes into account users' psychological profiles, **PRIDE**: **C2 + C3**.
- □ **C5** Design and building of **test collections**, namely:
  - ☑ **C5-1 ANASTASIA-D**: a dataset for itinerary recommendation on board of a cruise [6]
  - ☑ **C5-2 REvIt**: a dataset for **r**ecommendation of **ev**ents and **it**ineraries at Comic-Con International: San Diego ◷
  - ☑ **C5-3** a dataset of psychological profiles and selection of leisure activities [8]
  - □ **C5-4 REvIt++**: a dataset of psychological profiles and selection of activities at Comic-Con International: San Diego for recommendation of events and itineraries

2 PROBLEM STATEMENT (C1)

In this Ph.D. project, we define the problem of recommendation of **s**patio-**t**emporal **a**ctivity **s**equences (STAS) and related notions. Here, we present the definitions briefly.

STAS problem: Let $\mathtt{U}$ be the set of users, $\mathtt{A} = \{\mathtt{a}_1, \mathtt{a}_2, ..., \mathtt{a}_\mathtt{N}\}$ be the set of all available ***activities***, $\mathtt{N}$ be the total number of activities. For a given user $\mathtt{u} \in \mathtt{U}$, we aim to provide a feasible ***itinerary*** ξ that maximises the user's overall ***satisfaction*** with all performed activities: $\mathtt{Max}\ \sigma(\xi, \mathtt{u})$, subject to the set of ***constraints***. The concepts of **activities**, **itineraries** and **satisfaction** are defined as follows.

An **activity** $\mathtt{a}$ is a unique action taking place at some geographical location during a particular time window, and, therefore, can be represented as a tuple $\mathtt{a} = (\mathtt{id}, \mathtt{n}, \mathtt{l}, \mathtt{t_s}, \mathtt{t_e}, \delta, \mathtt{c}, \mathtt{d})$, where $\mathtt{id}$ is its identifier, $\mathtt{n}$ is its name, $\mathtt{l}$ is the location where it takes place, $\mathtt{t_s}$ and $\mathtt{t_e}$ are the start and the end time (time window) of its availability respectively, δ is its duration, $\mathtt{c} = (\mathtt{c}_1, ..., \mathtt{c}_\mathtt{m})$ is a vector of categories associated with the activity, and $\mathtt{d}$ is its description.

An **itinerary** (or **activity sequence**) $\xi(\mathtt{u}) = \mathtt{a_i} \to \mathtt{a_j} \to ... \to \mathtt{a_k}$ is a chronologically ordered series of activities of the user $\mathtt{u} \in \mathtt{U}$.

The *feasibility* of a sequence is defined based on the satisfaction of the set of constraints, among which: (1) *Activity availability*: an activity can be performed only within the time window of its availability. (2) *Activity completion*: a user may join an activity if there is enough time to perform it. (3) *Time budget*: the total time needed to perform all the activities within a sequence should not exceed the time budget.

The **satisfaction** $\hat{s}(a, u)$ **of a user** $\mathtt{u}$ **w.r.t. an activity** $\mathtt{a}$ is a quantitative measure of the matching between an activity $\mathtt{a}$ and a user $\mathtt{u}$, subject to several influences. In the most simple case, **the satisfaction w.r.t. an activity sequence** can be defined as the sum of satisfaction scores of individual activities forming the sequence, *i.e.* $\sigma(\xi, \mathtt{u}) = \sum_{\mathtt{a} \in \xi} \hat{\mathtt{s}}(\mathtt{a}, \mathtt{u})$.

3 BACKGROUND (C1)

This Ph.D. project is mainly dealing with the problem of recommendation of activity sequences (STAS). The latter brings our project close to two related recommendation problems, namely: Event Recommendation and Trip Recommendation.

Event Recommendation. The event recommendation problem is usually defined as a top-k recommendation problem that seeks to provide a user with a ranked list of events that might be interesting for him/her. Thus, it is always formulated as a list-wise [1, 11] or a pairwise [5] ranking problem. Various influences can be taken into account in order to estimate the preference of an event for a given user, *e.g.* users' temporal and geographical preferences, set of groups to join, event content [1]. It has to be noted that event recommendation is known to be intrinsically cold-start and it differs from the 'traditional' recommendation domains (movies,

Table 1: Comparison of related recommendation problems

Characteristic	Event Rec	Trip Rec	STAS
Limited Availability	✓	✓	✓
Travel Time	X	✓	✓
Unique Unit	✓	X	✓
Unique Visit	✓	✓	✓
Future oriented	✓	X	✓

Table 2: Interconnection between contributions and the data used for evaluation. Colours correspond to project phases.

	C5-1 [6]	**C5-2**	**C5-3** [8]	**C5-4**	External
C1	NA	NA	NA	NA	NA
C2 [7]					
C3					Meetup [1]
C4					

books, etc.) in terms of the peculiarities of items, as events have limited life time and happen in future, which results in the lack of collaborative data such as ratings provided by other users [1].

STAS has many points in common with event recommendation. In our Ph.D. project, we get inspired by the works in this field, but the problem we treat goes beyond the state-of-the-art of event recommendation by searching for the most desirable sequence of activities for a user, particularly in the context of a distributed event.

Trip Recommendation. The trip recommendation problem can be perceived as en extension of the Point-of-Interest (POI) recommendation problem which has become very popular in tourism domain. Given a user, his/her starting and ending points, a starting time, and a time budget, the trip recommendation aims at finding an optimal trip route limited by a time budget, that maximises the user's happiness and satisfies the POI availability constraints [15]. Trip recommendation is usually treated in two steps, namely (1) the estimation of user's interest in POIs and (2) the itinerary construction [13, 15]. Another strategy consists in estimating the transition probabilities between POIs. Thus, Sang *et al.* [12] formulate the problem as a ranking problem w.r.t. the visiting probability of a sequence of POIs, estimated using POI transition probability under the given user context. The problem is addressed on two levels, namely POI category sequences and POI sequences, which constitutes the hierarchical structure of the proposed solution.

If we place STAS within the context of the two aforementioned problems (see Tab. 1), one can note that (1) what distinguishes STAS from Event Recommendation is its sequential nature, while (2) the uniqueness of activities and their occurrence in future differ STAS from Trip Recommendation, adding more challenges.

4 METHODOLOGY AND PROPOSALS

This Ph.D. project undergoes four phases. Table 2 depicts the data used at each of the phases and visualises the interconnection between contributions that this Ph.D. projects aims to achieve.

4.1 STAS: the problem of recommendation of Spatio-Temporal Activity Sequences (C1)

The **first phase** of this research was to explore the field of recommendation of sequences of spatial items in order to *identify* the peculiarities of leisure activities as recommendation items, *define* the STAS problem aiming for a *better understanding* and *providing more insight* into the process of selection of leisure activities by individuals. We have investigated different types of influence from the state-of-the-art solutions that may impact the user's interest in an item, and proposed their classification. Moreover, we have provided an overview of available datasets and have discussed their use for STAS. We summarised this work in the form of a survey paper that is currently under review. **C1**

4.2 ANASTASIA: A Novel Approach for Short-Term Activity Sequence and Itinerary recommendAtion (C2, C5-1)

In the **second phase**, we have proposed a novel approach for the recommendation of planning of activities that we call ANASTASIA (**A** **N**ovel **A**pproach for **S**patio-**T**emporal **A**ctivity **S**equence and **I**tinerary recommend**A**tion) presented in [7]. It makes use of users behavioural patterns to construct the planning of events that best suit users constraints, preferences, and expectations. It is a hybrid approach that integrates categorical, temporal and textual scores of user's interest in an activity. Based on the estimated scores of activities and the extracted behavioural patterns (sequential influence) a personalised itinerary is constructed. Two computation strategy have been suggested. For more details, please, refer to [7]. **C2**

The evaluation has been performed on a dataset of a cruise attendance, issued from a user study based on a 7-night Disney Fantasy cruise. The main aim of the study was to simulate cruise attendance and create a dataset that could be used for personalised itinerary construction. Participants were recruited via a link to the online questionnaire sent by email to several research mailing lists. Some of the results of this study were reported in [6]. **C5-1**

4.3 Introducing Users' Psychological Profiles to Event Recommendation (C3, C5-2)

The selection of leisure activities is a complex process subject to various features of an activity and multiple layers of user's individuality. The **third phase** of this Ph.D. project is focused on investigating the impact of psychological profiles of individuals on the selection of leisure activities. We decompose this aim into three parts: (1) implicit acquisition of users' psychological profiles from their selection of leisure activities, (2) estimation of user's interest scores based on his/her psychological profile, (3) incorporation of these scores into a hybrid event recommendation model. In terms of psychological profiles, we are focusing on the following dimensions: (1) Orientations to Happiness (OTH) [9]; (2) Big5 Personality traits (Big5) [4]; and (3) Fear of Missing Out (FoMO) [10].

As the initial step of this work, we have conducted a user study, some of the results of which we have reported in [8]. **C5-3**

This study will allow us to estimate the desirability of events for a user on a category level. We further incorporate this psychological profile based scores into a learning to rank model for personalised recommendation of events together with social, textual, geographical, temporal scores [1]. The evaluation will be performed on a large-scale dataset crawled from Meetup.com [1]. **C3**

4.4 PRIDE: Personalised Recommendation of Itineraries during Distributed Events (C4), REvIt (C5-2) and REvIt++ (C5-4)

The **fourth phase** aims at finding a new solution for STAS that exploits various types of influence, *i.e.* textual, categorical, temporal, social, psychological, sequential, in order to estimate the desirability of events and event sequence for a user (**PRIDE**). **C4**

In [6], we have identified the characteristics of the data treated by activity sequence recommendation. Among the list of 14 characteristics, we distinguish 5 core ones: (1) time windows (start and end time of activity), (2) coordinates (geographical location of an activity), (3) service time (duration of an activity), (4) categories, and (5) users historical data. Using them as criteria, we have performed a comparative analysis of existing datasets, which has revealed that none of the available datasets contains all the core characteristics.

To overcome this issue, we have crawled the website of Comic-Con International: San Diego to create a new dataset that we call **REvIt** (**R**ecommendation of **Ev**ents and **It**ineraries). It contains Comic-Con event programs 2013-2017 and the pre-selection of the events made by participants, more precisely the following entities: events, locations, categories, tags, event-user, user-user, event-categories, event-tags. We have described the data collection process and the conducted case study on the use of REvIt for the recommendation purpose in a paper submitted to ACM UMAP conference. **C5-2**

We are further planning to extend REvIt by enriching it in terms of user profiles. Thus, we will launch a crowdsourcing campaign that will allow us to align the selection of Comic-Con events and psychological profiles of users. The study will consist of four parts. The first three parts are similar to [8], *i.e.*: demographic profile, psychological profiles (Big5, OTH, FoMO), and selection of categories of leisure activities. In the fourth part, the participants will be provided with the event program of Comic-Con 2017 and their task will be to simulate their attendance of Comic-Con by creating their schedules for each day of the convention. This user study aims at providing more insight into the process of event selection during a distributed event and users' time management. Moreover, it will allow us to investigate if the impact of psychological profiles of users on their selection and scheduling behaviour. The study will result in a new publicly available dataset **REvIt++**. **C5-4**

REvIt and REvIt++ datasets will be used for evaluation of PRIDE.

5 FURTHER RESEARCH

We foresee the following perspectives of this Ph.D. project.

FR1 – Recommendation of activity sequences meets aggregation operators [2]: A collaboration project between LIRIS / INSA Lyon and the University of Milano-Bicocca, supported by grant of IDEXLYON as part of the program "Doctoral Students' International Mobility" has been set up.

FR2 – Satisfaction w.r.t. a sequence of activities: An interesting direction of future work consists in exploring the interdependence of activities within a sequence in terms of satisfaction a user gets (*e.g.* [3]).

FR3 – Leisure activities and Well-being: Another direction to explore is to conduct a long-term study of the affect of recommended activities on the individual well-being.

FR4 – STAS and recurrent events: A more practical direction of the further research will be considering multiple time windows of activity availability [14], multiple locations, etc. in order to better deal with the recurrent events.

REFERENCES

[1] Augusto Q. Macedo, Leandro B. Marinho, and Rodrygo L.T. Santos. 2015. Context-Aware Event Recommendation in Event-based Social Networks. In *RecSys '15*. 123–130.

[2] Stefania Marrara, Gabriella Pasi, and Marco Viviani. 2017. Aggregation operators in Information Retrieval. *Fuzzy Sets and Systems* 324 (2017), 3–19.

[3] Judith Masthoff and Albert Gatt. 2006. In Pursuit of Satisfaction and the Prevention of Embarrassment: Affective State in Group Recommender Systems. *User Modeling and User-Adapted Interaction* 16, 3-4 (Sept. 2006), 281–319.

[4] Robert R. McCrae and Oliver P. John. 1992. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality* 60, 2 (1992).

[5] Einat Minkov, Ben Charrow, Jonathan Ledlie, Seth Teller, and Tommi Jaakkola. 2010. Collaborative Future Event Recommendation. In *CIKM '10*. 819–828.

[6] Diana Nurbakova, Léa Laporte, Sylvie Calabretto, and Jérôme Gensel. 2017. Itinerary Recommendation for Cruises: User Study. In *Proceedings of the 2nd Workshop on Recommenders in Tourism co-located with 11th ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, August 27, 2017*. 31–34.

[7] Diana Nurbakova, Léa Laporte, Sylvie Calabretto, and Jérôme Gensel. 2017. Recommendation of Short-Term Activity Sequences During Distributed Events. *Procedia Computer Science* 108, ICCS'17, Supplement C (2017), 2069 – 2078.

[8] Diana Nurbakova, Léa Laporte, Sylvie Calabretto, and Jérôme Gensel. 2017. Users psychological profiles for leisure activity recommendation: user study. In *CitRec@RecSys*. ACM, 3:1–3:4.

[9] Christopher Peterson, Nansook Park, and Martin E. P. Seligman. 2005. Orientations to happiness and life satisfaction: the full life versus the empty life. *Journal of Happiness Studies* 6, 1 (2005), 25–41.

[10] Andrew K. Przybylski, Kou Murayama, Cody R. DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior* 29, 4 (2013), 1841 – 1848.

[11] Zhi Qiao, Peng Zhang, Yanan Cao, Chuan Zhou, Li Guo, and Binxing Fang. 2014. Combining Heterogenous Social and Geographical Information for Event Recommendation. In *Proc. of the 28th AAAI Conference on Artificial Intelligence*. 145–151.

[12] Jitao Sang, Tao Mei, and Changsheng Xu. 2015. Activity Sensor: Check-In Usage Mining for Local Recommendation. *ACM Trans. Intell. Syst. Technol.* 6, 3, Article 41 (April 2015), 24 pages.

[13] Richard Schaller, Morgan Harvey, and David Elsweiler. 2013. RecSys for Distributed Events: Investigating the Influence of Recommendations on Visitor Plans. In *SIGIR'13*. 953–956.

[14] Wouter Souffriau, Pieter Vansteenwegen, Greet Vanden Berghe, and Dirk Van Oudheusden. 2013. The Multiconstraint Team Orienteering Problem with Multiple Time Windows. *Transportation Science* 47, 1 (Feb. 2013), 53–63.

[15] Chenyi Zhang, Hongwei Liang, Ke Wang, and Jianling Sun. 2015. Personalized Trip Recommendation with POI Availability and Uncertain Traveling Time. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM '15)*. 911–920.