

STAS : Recommendation of Spatio-Temporal Activities Sequences

Diana Nurbakova, Sylvie Calabretto, Léa Laporte, Jérôme Gensel

2^{ème} année de thèse, Financement ARC6, Équipe DRIM

diana.nurbakova@liris.cnrs.fr – <https://liris.cnrs.fr/membres?idn=dnurbako>

Résumé de la thèse

In this work, we explore a combination of users' sequential patterns and textual and categorical information in order to provide a user with a personalised itinerary of spatio-temporal activities during a distributed event (e.g. a cruise or a festival). Firstly, we model users' interest in activities by combining textual and categorical information. Secondly, we retrieve users' sequential patterns to model their typical transitions between activities and categories. Finally, we model the itinerary construction problem as an instance of Orienteering Problem with Time Windows (OPTW) using the obtained interest scores. We incorporate sequential patterns in the iterative algorithm to solve it. We evaluate our approach on a dataset built over a cruise program.

According to Billboard, 32 million people in the U.S. attend music festivals at least once per year¹, while Cruise Line International Association expects 24 million passengers to cruise globally in 2016². The common concept in festivals and cruises is that both provide a wide variety of experience by sampling hundreds of competitive activities in one place in a few days. Attendees of such distributed events are looking for personalised experience. In such a context, activities are short-lived, unique and available in a specific location only during particular time interval. Being overwhelmed with the amount of activities, it is difficult for attendees to choose activities of their interest and to plan their time. Therefore, a personalised itinerary that consists of a sequence of activities accounting for spatio-temporal constraints would be helpful for participants. In this work, we investigate the problem of recommendation of spatio-temporal activities sequences (STAS) and propose an integral approach to solve it based on users past experience.

Our approach consists of three parts. Firstly, we estimate user's interest score in activities by exploiting content and categorical influences. We propose to consider 3 scores : category-based, content-based and hybrid. We assume that an activity is assigned to a list of categories. Thus, for a given user and an activity, the category-based score is calculated as the frequency of user's attendance of activities of the same category than a given activity. The content-based score determines the textual similarity of an activity with user's past activities penalised by the similarity with non-performed activities. We represent activities with TD-IDF vectors of their descriptions. The hybrid score is calculated as a linear combination of category-based and content-based scores. Secondly, we extend the work of [1] in order to retrieve users transitional patterns based on their past activities. We construct a category-category transition graph,

in which the nodes represent categories of past activities and the edges stand for transitions between them and are associated with the transition frequency. We then calculate the transition probability between categories that is further used to estimate the transition probability between activities. Finally, we model the STAS problem as an instance of the OPTW [2], given activities scores, their time windows and duration, the travel time between the locations of activities. We adapt the Iterated Local Search (ILS) algorithm [2] by integrating the sequential patterns extracted on the previous step. Thus, we adjust the score that is maximised at each iteration of ILS with the transition probability from the previous activity to the current one.

We evaluated our approach on a dataset created via on-line survey. We simulated the attendance of a cruise taking the lists of activities on board of the shore. The respondents were asked to evaluate the activities and to create their daily itineraries. The evaluation process was hold in two steps : (I) evaluation of the accuracy of estimated activities scores ; (II) evaluation of created sequences compared with the ones obtained using ILS. The comparison was performed w.r.t. the similarity of the itineraries to the plannings created by users while completing the survey. The results of scores estimation show that the hybrid score outperformed pure category-based method and content-based method. As for the itinerary construction, our approach has shown 9.34% improvement compared to the state-of-the art approach.

Références

- [1] J.-D. Zhang, C.-Y. Chow, Y. Li. LORE : Exploiting Sequential Influence for Location Recommendations. In *Proceedings of the 22Nd ACM SIGSPATIAL, 2014*
- [2] P. Vansteenwegen, W. Souffriau, B. Vanden, D. Van Oudheusden. Iterated Local Search for the Team Orienteering Problem with Time Windows. *Comput. Oper. Res.*, 36(12), pp. 3281-3290, Dec. 2009.

1. <http://www.billboard.com/articles/columns/music-festivals/6539009/music-festival-statistics-graphic>

2. http://www.cruising.org/docs/default-source/research/2016_clia_sotci.pdf?sfvrsn=0