Itinerary Recommendation for Cruises: User Study

Diana Nurbakova, Léa Laporte, Sylvie Calabretto, Jérôme Gensel
Motivation

Cruising Statistics
2016 – 24.2 mln passengers globally
2017 - ≈25.3 mln expected
1980-2017 - average annual passenger growth rate – 7%/annum
2005-2015 – increase in demand for cruising – 62%
2016 Deployed Capacity Share – 33.7% Caribbean/Bahamas (FCCA, 2017)
Motivation

Who Cruises?
- Preferred vacation choice for families, esp. with kids <18
- Kids are involved in decision process
- Millennials & Generation X: cruises > land-based vacations
- The best for relaxing and getting away from it all, for see and do new things (FCCA, 2017)
Motivation
Key Questions

- What are the data characteristics?
- Is there any existing dataset?
- What is it?
- What makes it challenging?
Itinerary Recommendation: Problem Statement

What We Have

- **Set of Activities**, $\mathcal{A} = \{a_i\}_{i=1}^{1,N}$
  - $a = \langle l, t, \delta, c, d \rangle$
    - $l = (x, y, z)$ – location
    - $t = (t_s, t_e)$ – time window (start & end)
    - $\delta = (t_e - t_s)$ – duration
    - $c = (c_1, c_2, \ldots, c_k)$ – vector of categories
    - $d$ – textual description

- **Set of Users**, $U = \{u_j\}_{j=1,M}$

- **Users’ History**, $\mathcal{M}$:
  - $\mathcal{M}_{i,j} = \begin{cases} 1, & j^{th} \text{ user joined } i^{th} \text{ activity} \\ 0, & \text{otherwise} \end{cases}$

What We Want

- **Activity Sequence (itinerary)**,
  - $\xi(u) = (a_{(1)} \rightarrow \ldots \rightarrow a_{(s)} \rightarrow \ldots \rightarrow a_{(s+k)})$, $1 \leq s \leq s + k \leq N$
  - Activity availability constraint:
    - $t_s(a_{(i)}) \leq \text{start}(a_{(i)}) \leq t_e(a_{(i)})$
    - $\text{start}(a_{(i)}) = \max\{\text{start}(a_{(i-1)}) + \delta(a_{(i-1)}) + \text{time}(a_{(i-1)}, a_{(i)}), t_s(a_{(i)})\}$
  - Time budget constraint:
    - $\sum_{a_{(i)} \in \xi(u)} \text{time}(a_{(i-1)}, a_{(i)}) + \delta(a_{(i)}) \leq T_{\text{max}}$

- **User’s Satisfaction**: 
  - w.r.t. activity $r(a, u)$, $r: \mathcal{A} \rightarrow \mathbb{R}^+$
  - w.r.t. itinerary $\rho(\xi, u)$, $\rho: \Xi \rightarrow \mathbb{R}^+$

- **Find**: $\forall u \in U$, $\xi(u) = \max \rho(\xi, u)$

(Nurbakova et al., 2017)
Challenges

- List vs. Itinerary
- Implicit Feedback
- Interest vs. Attendance
Data Characteristics

**ITEM**
- Time Windows
- Coordinates
- Service Time
- Categories
- Description
- Price
- Item Additional Attributes

**SEQUENCE**
- Time Budget
- Starting/Ending Point
- Tour Additional Attributes

**USER**
- User’s personal data

**USER-ITEM**
- Historical data
- Score

**USER-USER**
- Social links
## Existing Datasets

### Need for Dataset

<table>
<thead>
<tr>
<th>Entity</th>
<th>Characteristic</th>
<th>Single Item</th>
<th>Schedule</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td><strong>Time windows</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td><strong>Coordinates</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td><strong>Service Time</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td><strong>Categories</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Price</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Item Additional Attributes</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Sequence</td>
<td>Time budget</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Starting/Ending Point</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Tour Additional Attributes</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>User</td>
<td><strong>User’s personal data</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>User-Item</td>
<td><strong>Historical Data</strong></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>User-User</td>
<td>Social links</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

2. [https://github.com/jalbertbowden/foursquare-user-dataset](https://github.com/jalbertbowden/foursquare-user-dataset)
# User Study: Stats

## Dataset Statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users</td>
<td>23</td>
</tr>
<tr>
<td>Number of activities in overall program</td>
<td>595</td>
</tr>
<tr>
<td>Number of days</td>
<td>7</td>
</tr>
<tr>
<td>Number of DCL categories</td>
<td>10</td>
</tr>
<tr>
<td>Number of No DCL categories</td>
<td>42</td>
</tr>
<tr>
<td>Number of locations</td>
<td>47</td>
</tr>
<tr>
<td>Average time of completion</td>
<td>50min-1h</td>
</tr>
</tbody>
</table>

Link: [https://goo.gl/forms/ZEX4LPhcg0qDAzIr1](https://goo.gl/forms/ZEX4LPhcg0qDAzIr1)
User Study: Part I

USER PROFILE

Number of questions – 10

Basic users’ features and their cruising experience

Examples:
1. Your gender
2. Have you already experienced DCL (Disney Cruise Line)
3. Have you tried any other cruise?
4. The type of group you were/are travelling with. Please, choose, the option that best describes you.
5. If you were travelling with a group, have you split to attend different activities or you mostly preferred to stay together?
USERS PREFERENCES
Number of questions – 311

Evaluation of a list of proposed activities on 5-point scale: 1 – Never (not interested at all and won’t recommend to anyone to attend it), 5 – Won’t miss

Examples:
A Pirate's Life For Me. Don't Miss Event.
Description: Calling all Pirates, we be! If ye have an adventurous spirit or pirate savvy, come spin the "Wheel of Destiny" fer a treasure trove of fun be ripe for the takin' in this action packed pirate game show.
Available: Day 4, 18:30-19:00, Location: D Lounge & Day 4, 21:30-22:00, Location: D Lounge

Never ○ ○ ○ ○ ● Won’t Miss
User Study: Part III

ITINERARY PLANNER

Number of questions – 593

Organisation of daily planner of activities by indicating the intention ‘Going’ or ‘Not Going’

Examples:

<table>
<thead>
<tr>
<th>Event</th>
<th>Going</th>
<th>Not going</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 - 14:30. Walking Ship Tour.</td>
<td>✅</td>
<td>❌</td>
</tr>
<tr>
<td>Category: Fun for all ages.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location: Preludes. Don't Miss Event</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
USER STUDY: PART IV

AFTERWARDS
Number of questions – 5

Concluding questions

Examples:
1. Could you, please, select the categories of activities that represented the most interest for you.
2. When you were having a choice among different activities of your interest, did you consider the distance to the venue while making your choice? If you prefer a nearby activity rather than an activity on the opposite part of the ship, please select yes. If the distance doesn't matter for you, please select no.
3. How do you usually manage the list of activities to perform during your vacations?
Interest vs. Attendance
List vs. Itinerary

Top-\(k\) vs. Itinerary construction

- **Content-based (CB):** TF-IDF representation of activity (title + description) and user’s past activities
- **Category-based (Cat):** weighted frequency of categories (Nurbakova et al., 2017)
- **Logistic Regression (LogR):** CB + Cat
- **ILS + Scores:** hybrid scores + transition probabilities between activities + ILS algorithm (Nurbakova et al., 2017)
What We’ve Learnt?

- The shorter questionnaire – the more respondents
- The most desirable characteristics of a dataset for the itinerary recommendation: time windows of an item, coordinates, service time, categories and users historical data
- There’s a gap between users’ interest in an activity and their engagement to it
- In the context of a distributed event (e.g. a cruise), a personalised itinerary fits better users’ behaviour than a list of top-$k$ activities
The End

WARNER BROS. FIRST NATIONAL PICTURE
References


