Travel Itinerary Mining

Sihem Amer-Yahia
CNRS/LIG
Nov 16th, 2015
Extracting travel itineraries from Flickr

**Goal:** extract the itinerary of each traveler by mapping photos into Points Of Interest (POIs) and aggregate actions of many travelers into coherent queryable itineraries.

- Feedback on various aspects of the itineraries constructed by our system from a large number of anonymous users

Problem definition

• **Definitions**
  - Each itinerary is a timed path
  - The set of timed paths implies a *weighted graph* G over POIs
  - An *itinerary* is a path in the graph G
  - The *value* of an itinerary is the sum of popularities of its POIs
  - The *time* of an itinerary is the sum of POI visit and transit times

• **Problem Instance ("Orienteering")**
  - Find an itinerary in G from a *source* POI to a *target* POI of budget (=time) at most B maximizing total value
  - The time budget B is typically whole days
  - *source* and *target POIs* provided by user (e.g. hotel)
• Identify photos of a given city
• Filter out residents of a city
• Validate photo timestamps

• Extract Candidate POIs
  o Lonely Planet/Y! Travel to extract landmarks
  o Yahoo! Maps API to retrieve their geo-locations

• Tag & geo-based POI association

• Photo Streams Segmentation
  o Split the stream whenever the time difference between two successive photos is “large”

• Distillation of Timed Visits
  • <POI, start time, end time>

• Construction of Timed Paths
  o A sequence of Timed Visits
# Data preparation

- Five popular and geographically distributed cities were chosen: Barcelona, London, New York City (NYC), Paris, and San Francisco.
- For each city, we generate four itineraries using our system.

<table>
<thead>
<tr>
<th>City</th>
<th>#POIs</th>
<th>#Timed Paths</th>
<th>Sample POIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcelona</td>
<td>74</td>
<td>6,087</td>
<td>Museu Picasso, Plaza Reial</td>
</tr>
<tr>
<td>London</td>
<td>163</td>
<td>19,052</td>
<td>Buckingham Palace, Churchill Museum, Tower Bridge</td>
</tr>
<tr>
<td>New York City</td>
<td>100</td>
<td>3,991</td>
<td>Brooklyn Bridge, Ellis Island</td>
</tr>
<tr>
<td>Paris</td>
<td>114</td>
<td>10,651</td>
<td>Tour Eiffel, Musee du Louvre</td>
</tr>
<tr>
<td>San Francisco</td>
<td>80</td>
<td>12,308</td>
<td>Aquarium of the Bay, Golden Gate Bridge, Lombard Street</td>
</tr>
</tbody>
</table>
Itinerary generation

• For each city, we generate four itineraries using our system.

• We first select the city’s four most popular POIs and designate them as \( \ell_1 \) (most popular) through \( \ell_4 \).
  – The popularity of a POI is determined by the number of distinct users who have provided a photo associated with the POI.

• The four itineraries for each city are then constructed by setting the starting point and ending point as \((\ell_1, \ell_3), (\ell_1, \ell_4), (\ell_2, \ell_3), (\ell_2, \ell_4)\), with a time budget of 12 hours.
Example itinerary for NYC (single-day)

Time 09:00 : Start from ground zero
Time 09:00 : Spend 27 minutes at ground zero.
Time 09:27 : Transit to empire state building (estimated travel time: 52 minutes)
Time 10:19 : Spend 1 hour and 13 minutes at empire state building.
Time 11:32 : Transit to new york public library (estimated travel time: 15 minutes)
Time 11:47 : Spend 29 minutes at new york public library.
Time 12:16 : Transit to radio city music hall (estimated travel time: 24 minutes)
Time 12:43 : Spend 51 minutes at radio city music hall.
Time 13:34 : Transit to central park (estimated travel time: 23 minutes)
Time 13:57 : Spend 40 minutes at central park.
Time 14:37 : Transit to rockefeller center (estimated travel time: 33 minutes)
Time 15:10 : Spend 37 minutes at rockefeller center.
Time 15:47 : Transit to grand central terminal (estimated travel time: 22 minutes)
Time 16:09 : Spend 27 minutes at grand central terminal.
Time 16:36 : Transit to chrysler building (estimated travel time: 6 minutes)
Time 16:42 : Spend 31 minutes at chrysler building.
Time 17:13 : Transit to brooklyn bridge (estimated travel time: 32 minutes)
Time 17:45 : Spend 36 minutes at brooklyn bridge.
Time 18:21 : Transit to statue of liberty (estimated travel time: 21 minutes)
Time 18:42 : Spend 42 minutes at statue of liberty.
Time 19:24 : Transit to little korea (estimated travel time: 26 minutes)
Time 19:50 : Spend 31 minutes at little korea.
Time 20:21 : Transit to ground zero (estimated travel time: 38 minutes)
Goal of user study

• Estimate the usefulness of the itineraries from two aspects:
  – overall utility of the itineraries
  – appropriateness of POIs

• Challenge
  – design a set of questions to AMT users and collect and interpret feedback
  – what is our ground truth?
## Ground truth

<table>
<thead>
<tr>
<th>City</th>
<th>Ground Truth Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcelona</td>
<td><a href="http://www.barcelona-tourist-guide.com">www.barcelona-tourist-guide.com</a></td>
</tr>
<tr>
<td>London</td>
<td><a href="http://www.theoriginaltour.com">www.theoriginaltour.com</a></td>
</tr>
<tr>
<td>New York City</td>
<td><a href="http://www.newyorksightseeing.com">www.newyorksightseeing.com</a></td>
</tr>
<tr>
<td>Paris</td>
<td><a href="http://www.carsrouges.com">www.carsrouges.com</a></td>
</tr>
<tr>
<td>San Francisco</td>
<td><a href="http://www.allsanfranciscotours.com">www.allsanfranciscotours.com</a></td>
</tr>
</tbody>
</table>
User study design summary

- Side-by-side evaluation comparing our itineraries to ground-truths
- Independent evaluation examining our itineraries in detail

Questions?
- Which itinerary is better?
  - POIs
  - Transit times
  - Visit times

Questions?
- Is the itinerary reasonable?
  - POIs
  - Transit times
  - Visit times
Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it’s convenient.

509,249 HITs available. View them now.

Make Money by working on HITs

HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:
- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task Work Earn money

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you’re satisfied with the results

Find your account Load your tasks Get results

Find HITs Now

or learn more about being a Worker
<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Time Allotted</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jon Breig</td>
<td>Dec 16, 2014 (6 days 23 hours)</td>
<td>2 hours</td>
<td>$0.09</td>
<td>19011</td>
</tr>
<tr>
<td>Amazon Requester Inc.</td>
<td>Dec 30, 2014 (3 weeks)</td>
<td>60 minutes</td>
<td>$0.00</td>
<td>18999</td>
</tr>
</tbody>
</table>
Comparative evaluation

Evaluation Questions:

I. Overall, which one of the above two proposed itineraries you would rate higher?
   - Itinerary 1 is significantly more useful than Itinerary 2.
   - Itinerary 1 is somewhat more useful than Itinerary 2.
   - Both are similar.
   - Itinerary 2 is somewhat more useful than Itinerary 1.
   - Itinerary 2 is significantly more useful than Itinerary 1.

II. How would you rate the set of points of interest included in the two itineraries?
   - Itinerary 1 has significantly more appropriate points of interest than Itinerary 2.
   - Itinerary 1 has somewhat more appropriate points of interest than Itinerary 2.
   - Both are comparatively similar.
   - Itinerary 2 has somewhat more appropriate points of interest than Itinerary 1.
   - Itinerary 2 has significantly more appropriate points of interest than Itinerary 1.

III. How would you rate the transit times at the points of interest in the two itineraries (from a tourist perspective)?
   - Itinerary 1 has significantly more accurate transit times than Itinerary 2.
   - Itinerary 1 has somewhat more accurate transit times than Itinerary 2.
   - Both are comparatively similar.
   - Itinerary 2 has somewhat more accurate transit times than Itinerary 1.
   - Itinerary 2 has significantly more accurate transit times than Itinerary 1.

IV. Any additional comments?
Independent evaluation

Q1: Overall, would you rate the proposed itinerary as:
—Not at all useful to a tourist
—Not so useful to a tourist
—Somewhat useful to a tourist
—Very useful to a tourist

Q2: How would you rate the set of points of interest included in the itinerary?
—Make no sense
—Mostly inappropriate
—Somewhat appropriate
—Mostly appropriate

Q3: How would you rate the visit times at the landmarks, as proposed by the itinerary (from a tourist perspective)?
—Not accurate at all
—Somewhat accurate
—Mostly accurate
—Completely accurate
If you picked choices 3 or 4, did you find the visit times too short or too long?

Q4: How would you rate the transit times between the landmarks, as proposed by the itinerary (from a tourist perspective)?
—Not accurate at all
—Somewhat accurate
—Mostly accurate
—Completely accurate
If you picked choices 3 or 4, did you find the transit times too short or too long?
Evaluation measures

– **Mean Weighted Response (MWR)** – aggregate the responses to each question from the workers in the same group, into a single number. Take mean across different itineraries generated by our method.

– **Mean Average Error Fraction (MAEF)** – compute the percentage of the number of POIs, visit times, or transit times, that are considered bad or inaccurate by a particular worker, out of the total number of POIs.
Results for side-by-side comparison

\[ MRV_{\text{opt}, q} = \frac{1}{n_{q}(\text{opt})} \frac{1}{|C|} \sum_{C \in C} \sum_{I} n_{q}^{I,C}(\text{opt}), \]  

where \( n_{q}^{I,C}(\text{opt}) \) is the number of workers who chose the option \( \text{opt} \) in question \( q \) for the HIT involving our system-generated itinerary \( I \) and city \( C \); and \( n_{q}(\text{opt}) \) is the total number of workers who responded to option \( \text{opt} \) for question \( q \) across all HITs.

### Q1: Itinerary Usefulness

- Significantly better
- Somewhat better
- Similar
- Somewhat worse
- Significantly worse

### Q2: POI Appropriateness

Mean Response Volume

- 0
- 0.1
- 0.2
- 0.3
- 0.4

- 0
- 0.1
- 0.2
- 0.3
- 0.4
The mean error fraction of (a) POIs, (b) Visit Times, and (c) Transit Times:

<table>
<thead>
<tr>
<th>London Itineraries</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP It. 1</td>
<td>3.1</td>
<td>2.9</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>IMP It. 2</td>
<td>3.5</td>
<td>2.1</td>
<td>2.7</td>
<td>2.1</td>
</tr>
<tr>
<td>IMP It. 3</td>
<td>3.4</td>
<td>2.5</td>
<td>2.8</td>
<td>2.7</td>
</tr>
<tr>
<td>IMP It. 4</td>
<td>3.5</td>
<td>2.7</td>
<td>2.9</td>
<td>3.1</td>
</tr>
<tr>
<td>Ground Truth</td>
<td>3.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Summary and challenges

- AMT enables scaling up user studies to hundreds, thousands of users
- AMT is just a hiring platform
- Experiment designer must “track” users and enforce consistency
  - in group recommendations, have users really seen the movies they are asked to rate to build their profile?
  - in itinerary planning, do hired users really know about a city?
Filtering expert AMT workers

- Multiple-choice questions on “less-known” POIs

**QUALIFICATION EVALUATION**

Please choose the most suitable name of the point of interest based on your experience. This would judge your fitness to take the travel itinerary evaluation task in the next section.

- Empire State Building
- Rockefeller Center
- Flatiron Building
- Saint Patrick’s Cathedral
- Chrysler Building
- Trinity Church
- Herald Square
- Washington Sq Park
- Lincoln Center