Evaluating routes in tourism recommenders

Priit Järv, TUT, 2016

Outline

- What's a tourism recommender
- Usual approaches to recommendation
- Item collections and sequences
- Existing metrics and methods:
  - Evaluating the collections (tours)
  - Evaluating the approach itself
- Combining RS techniques with discrete optimization (OP)

Tourism recommender

- Also called “tourist recommender”
- suggests trips, individual POI-s or packages

In this presentation, we talk about trips, such as:

“See statue X, visit church Y, dine at restaurant Z”

(a sequence of individual POI-s)

Relevant fields

Two lines of research:

- Recommender systems (information retrieval)
- Tourist trip planning (optimization, operations research)

Not much overlap in publications

Recommender system

- We have a set of items I
- Each item has a set of features F_i
- The user has a set of preferences P (a “profile”)
- Find utility $u_i = f(F_i, P)$ for each item
- Present top-n items sorted by u_i

This is called “content-based filtering”

TTDP solver

TTDP defined by Vansteenwegen et al. 2007 as an optimization problem

- each POI gives a reward r_i
- moving between POIs incurs cost c_{ij}
- maximise $\sum_{i \in S} r_i$ so that total cost is below some limit

This is the “orienteering problem” from operations research.

Problem

A tour consists of multiple items, consumed in a SEQUENCE

- Top-n is great for picking a single item to buy (online shops)
- OP is great for logistics, where we care about a distinct metric ($ value of goods delivered) vs a cost (time or operational costs)

Neither approach allows evaluating a trip as a single entity. A bunch of top-n items on a shortest possible route is not necessarily a fun trip.

Existing workarounds

- recommender systems try to measure and factor in: diversity, novelty, serendipity (“oh look I found this cool thing I totally wasn’t even looking for”)
- TTDP solvers attempt to add constraints, such as max-n of certain type (promotes diversity)

These approaches do not consider interaction between items.

Recommending a sequence

Hansen and Golbeck (2009): evaluating a collection should include:

1. individual ratings
2. co-occurrence interaction effects
3. order interaction effects

Their example is mixtapes. Trips are similar, but add the dimension of location/travel.

Evaluating sequences

Measuring diversity:

intra-list similarity (Ziegler 2005)

c(i,j) similarity between items i, j

ILS = ½ * $\sum_{i \in S} \sum_{j \in S;\, i \neq j} c(i,j)$

Evaluating sequences

Other, niche metrics have been suggested to complement individual item rankings.

TODO: some examples?

Evaluating sequences

OR approaches:

Just maximize the reward (evaluation mostly swept under the rug)

- Secondary criteria are encoded as constraints
- or, use weighted linear aggregation of multiple criteria

RS evaluation

Accuracy metrics

N - total recommended; n - number of relevant items; R - relevant items in recommendation; T - tail items in recommendation

precision - R / N

recall - R / n

fallout - T / N

How to create meaningful sequences?

The plan:

Combine RS methods (list/sequence recommendation) with discrete optimization (construction of routes)

UNDER CONSTRUCTION

Counterpoint

Cremonesi (2013) showed that recall and fallout metrics correlated with user study; cited by Jannach (2015) as evidence that in tourism domain, accuracy (picking items in top-n) is sufficient.

Ideas

Automated playlist generation (AGP) is a similar problem, try to learn from their methods.

Similarity comes from:

- consumed as a sequence
- interaction is relevant (for example, coherence between songs)
- very subjective evaluation

Dissimilar: all songs constantly available, selection of POI-s (sometimes severely) limited by cost of travel.

Ideas

Playlist generation (Bonnin and Jannach 2014):

- choose songs by similarity
- collaborative filtering (i.e. nearest neighbor)
- pattern mining (n-gram)
- statistical (Markov chain)
- discrete optimization (CSP)

Ideas

Methods can be hybridized. Some examples (Burke 2002):

- weighted linear combination of scores
- mixed (present at the same time)
- combination (mix techniques into one algorithm)
- feature augmentation (one method input of another)
- cascade (refine results with different method)
- meta-level (learn model to drive another method)

Recommending a sequence

Some discussions about recommending a sequence:

Ziegler et al. (2005) - introduces the idea of topic diversification, discusses radio station programming

Masthoff, in “RS handbook” (2011) - TV programming (but in the context of group recommendation)

Adamopoulos et al. (2013) - use sets, not individual items. Consider interaction, prerequisites.

Suggested user study

Evaluate the suggested approach

UNDER CONSTRUCTION