Automatically Planning Itineraries Using Business Review Data

Reynis Vazquez, Deeksha Goyal, Robert Fearon

Code Repository: https://github.com/rafearon/DayPlanner

Introduction

While business review aggregators, such as Yelp, make finding singular businesses easier than ever, it is still time consuming to create an itinerary by researching locations separately, calculating the time cost, and finding routes to each location. Using Yelp data from the San Francisco Bay Area and the Google Maps and Google Places APIs, we want to automatically chain activities together that a person would reasonably enjoy while staying within constraints guided by user preferences. The final itinerary will include a schedule of activities along with the route to each activity and the estimated duration of each activity.

Parameters

Data: We scraped Yelp records for 13,838 Bay Area businesses

1. **Inputs**
    a. Time range for itinerary
    b. General locations to visit (ex: Palo Alto, San Francisco)
    c. Genre of itinerary to create
    d. Public transportation versus car
    e. Include meals versus no meals
    f. Budget

    Example: Time = 9:00-13:00, Location = Berkeley, Genre = Indoors, Transportation = Car, Food = Yes, Budget = $20

2. **Output**
    a. List of activities for the time range
    b. Travel time and route between locations

    We also will have the output itineraries be malleable where the user can specify that they want a different activity and we replace it with something better for them.

Example (Using baseline model):
1) 9:00, Lawrence Hall of Science, Planetarium, Berkeley, CA, 4.0/5.0
2) 12:00, Mei Lan Aquarium, Aquarium, Oakland, CA, 4.5/5.0

Itinerary Scoring Model

We plan on scoring our algorithm's itineraries by trying to minimize distance, maximize yelp rating, keep food between 11-2, 5-8, and keep cost within the range provided.

Baseline

Our naive model takes user-provided input constraints and outputs a list of the first activities it finds that fall within the constraints. These activities may not be co-located at

all. This baseline succeeds in staying within budget, choosing activities within the user-provided genre, and staying within the designated time range. However it does poorly with the rest of the scoring model as it does not take into consideration the distance between activities, the ratings and reviews of businesses, or the time that a meal is scheduled.

Oracle

Our oracle data are person-created itineraries that represent what a normal person would like to do within our standard time parameters. We would want to compare the oracle data with our itineraries and have them be indistinguishable.

Proposed AI Algorithms

1. Constraint Satisfaction Problem
2. Search Problem
3. Business Review Classification to identify good matches

Challenges

This requires handling a lot of data and many different user constraints. We will have to think of clever ways of making itineraries that are are reasonable and enjoyable. We will also need to make the algorithm fast so that it doesn't bore the user while they are searching for itineraries. Lastly, we would want this to provide more value than trip-planning products that are already out there (see literature review below).

Relevant Work

There are some companies that have tried to tackle this problem, such as Utrip and JetBlue. We have, however, noticed several pain points in their systems such as unrealistic itineraries and bad input/output design. We think that our new design will tackle those issues head-on.

- Being accurate is not enough: how accuracy metrics have hurt recommender systems
  https://dl.acm.org/citation.cfm?id=1125659
- The City Trip Planner: An expert system for tourists
  http://www.sciencedirect.com/science/article/pii/S0957417410013230
- Artificial Intelligence Trip Planning Technology
  https://thepointsguy.com/2017/06/jetblue-vacations-ai-trip-planning/
  https://www.utrippro.com/