User Model Enrichment for Venue Recommendation

AIRS 2016

Mohammad Aliannejadi, Ida Mele, and Fabio Crestani

Università della Svizzera italiana (USI)
Lugano, Switzerland

December 2^{nd} 2016
Motivation

Challenges

- To model a user based on her history of preferences
- Different ratings for similar venues
- No reviews from the users, only ratings

Our Goal

- To model the user based on venue content
- To mine the reasons a user gave a specific rating to a venue
Approach

- A combination of multimodal scores from multiple sources
- Sources: Yelp, Foursquare, and TripAdvisor
- Types of information: categories, venue taste keywords, reviews
- Two types of scores:
  - Content based
  - Review based
To have a better idea of the user’s taste and interest we need to take into account their liked/disliked categories.

It is not clear exactly which category or subcategory a user likes/dislikes.

In this example, we see the corresponding categories to three attractions a user likes:

- Pizzeria - Italian - Takeaway - **Pizza**
- Restaurant - Pasta - **Pizza** - Sandwich
- Restaurant - American - **Pizza** - Burger

The user likes **Pizza**, since it is the only category in common.

We introduce a score to model user interest.
for all $v_i \in V$ do
  for all $c_j \in C(v_i)$ do
    if $c_j \notin CM_{pos}$ then
      $CM_{pos} \leftarrow CM_{pos} \cup c_j$
      $\text{count}(c_j) = \sum_{v_s \in V} \sum_{c_k \in C(v_s)} \delta(c_j, c_k)$
      $N = \sum_{v_s \in V} \sum_{c_k \in C(v_s)} 1$
      $cf_{pos}(c_j) = \text{count}(c_j) / N$
    end if
  end for
end for
Given a user $u$ and a venue $v$, the category-based similarity score $S_{CM}(u, v)$ is:

$$S_{CM}(u, v) = \sum_{c_i \in C(v)} cf_{pos}(c_i) - cf_{neg}(c_i)$$

where $cf_{pos}$ and $cf_{neg}$ are respectively the positive and negative categories’ frequencies.
We calculate three frequency-based scores using different types and sources of information:

- Categories from Yelp: $S_{CM}^{Yelp}$
- Categories from TripAdvisor: $S_{CM}^{TAvisor}$
- Venue taste keywords from Foursquare: $S_{TM}$
Review-based Score

Parco Ciani

Positive Profile
- Great view and nature
- Well kept – super clean
- Very serene. A must-see

Reviews with similar ratings

Ristorante Tango

Negative Profile
- The prices are very high
- Very crowded
- I did not like the pasta
We assume that user likes what others like about a place and vice versa.

Find reviews with similar rating:

- **Positive Profile**: Reviews with rating 3 or 4 corresponding to places that user gave a similar rating.
- **Negative Profile**: Reviews with rating 0 or 1 corresponding to places that user gave a similar rating.

Train a classifier for each user: SVM and Naïve Bayes.

Features: TF-IDF score of each term.

Score: decision function → $S_{BM}$.
We rank the venues based on their similarity with the user.

Given user $u$ and venue $v$, we calculate the similarity score as follows:

$$SIM(u, v) = \alpha \times S_{CM}^{el}(u, v) + \beta \times S_{CM}^{Advisor}(u, v) + \eta \times S_{TM}(u, v) + \gamma \times S_{BM}(u, v)$$
Introduction

TREC CS Track

TREC 2015

- Contextual Suggestion Track deals with complex information needs which are highly dependent on context and user interests.

What do we have?

- 211 users
- User context
- User history: 60 rated venues in two cities

What should we do?

- Rank the candidate list: 30 venues in a new city

Evaluation: P@5 and MRR
Context

- A city the user is located in, which consists of:
  - An ID
  - A city - The name of the city
  - A state - The name of the US state the city is in
  - A latitude and longitude - These are available for convenience and do not represent the exact user location but are analogous to the city name.

- A trip type (optionally), which is one of:
  - Business
  - Holiday
  - Other
A trip duration (optionally), which is one of:
- Night out
- Day trip
- Weekend trip
- Longer

The type of group the person is traveling with (optionally), which is one of:
- Traveling alone (Alone)
- Traveling with a group of friends (Friends)
- Traveling with family (Family)
- Traveling with an other group (Other)

The season the trip will occur in (optionally)
User History

- Profiles consist of a list of attractions the user has previously rated. For each attraction the profile will include a rating as follows:
  - 4: Strongly interested
  - 3: Interested
  - 2: Neither interested or uninterested
  - 1: Uninterested
  - 0: Strongly uninterested
  - -1: No rating given

- Additionally the user may annotate the attraction with tags that indicate why the user likes the particular attraction:
  - Art Galleries, Family Friendly, Fine Art Museums, etc.

- The user’s age and gender (optionally).
Dataset

What was provided by the organizers?
- An attraction ID
- A city ID which indicates which city this attraction is in
- A URL with more information about the attraction
- A title

What did we collect?
- Crawl venues from Location-based Social Networks (LBSNs):
  - Foursquare
  - Yelp
  - TripAdvisor
## Dataset (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td># of crawled venues</td>
<td>6290</td>
<td>4633</td>
<td>5534</td>
</tr>
<tr>
<td>Distribution of categories over venues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>2.80</td>
<td>1.94</td>
<td>1.63</td>
</tr>
<tr>
<td>Variance</td>
<td>1.98</td>
<td>1.23</td>
<td>0.63</td>
</tr>
<tr>
<td>Distribution of reviews over venues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>17</td>
<td>89</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td>117.34</td>
<td>446.42</td>
<td>-</td>
</tr>
<tr>
<td>Maximum</td>
<td>6060</td>
<td>57365</td>
<td>-</td>
</tr>
<tr>
<td>Distribution of taste tags over venues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>-</td>
<td>8.73</td>
</tr>
<tr>
<td>Variance</td>
<td>-</td>
<td>-</td>
<td>7.22</td>
</tr>
<tr>
<td>Approach</td>
<td>P@5</td>
<td>Rank</td>
<td>P@5</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>CatRev-SVM</td>
<td>0.5858</td>
<td>1</td>
<td>0.7404</td>
</tr>
<tr>
<td>CatRev-NB</td>
<td>0.5450</td>
<td>7</td>
<td>0.6991</td>
</tr>
<tr>
<td>BASE1</td>
<td>0.5706</td>
<td>2</td>
<td>0.7190</td>
</tr>
<tr>
<td>BASE2</td>
<td>0.5583</td>
<td>3</td>
<td>0.6815</td>
</tr>
<tr>
<td>TREC Median</td>
<td>0.5090</td>
<td></td>
<td>0.6716</td>
</tr>
</tbody>
</table>

17 teams - 30 runs
Analysis

![Bar Chart]

- P@k values for different k values (1 to 30).
- The chart shows a decreasing trend in P@k as k increases.

Legend:
- P@k: Precision at k
- k: Number of recommendations

Table:

<table>
<thead>
<tr>
<th>k</th>
<th>P@k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td>2</td>
<td>0.64</td>
</tr>
<tr>
<td>3</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>0.58</td>
</tr>
<tr>
<td>10</td>
<td>0.55</td>
</tr>
<tr>
<td>15</td>
<td>0.53</td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
</tr>
<tr>
<td>30</td>
<td>0.45</td>
</tr>
</tbody>
</table>
We proposed content-based and review-based scores
We combined multimodal scores from multiple LBSNs
Official results of TREC 2015 proves the effectiveness of our approach
Context-aware venue recommendation
Mapping user tags into venue content to have a more precise user model
Thanks for your attention
Thanks to ACM SIGIR for supporting my travel

Mohammad Aliannejadimohammad.alian.nejadi@usi.ch
@maliannejadi