User Model Enrichment for Venue Recommendation

AIRS 2016

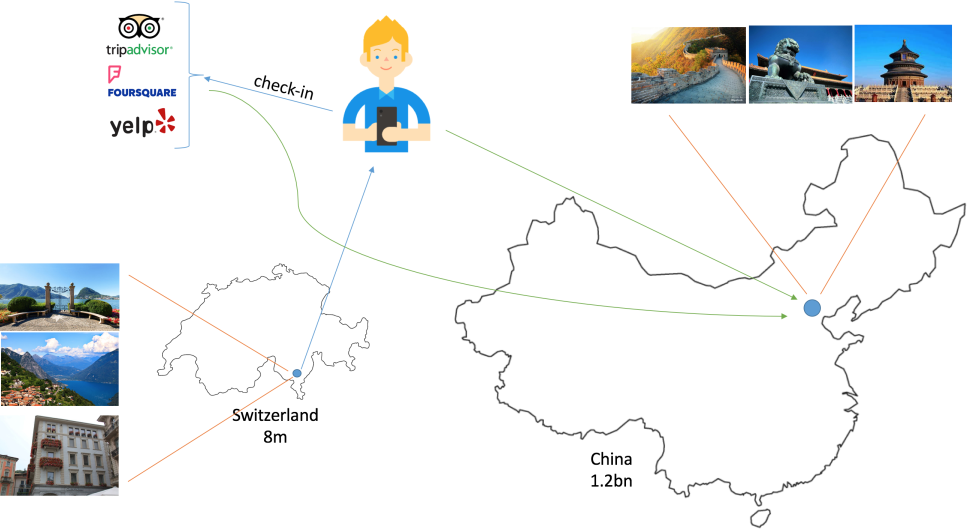
Mohammad Aliannejadi, Ida Mele, and Fabio Crestani

Università della Svizzera italiana (USI)
Lugano, Switzerland

December 2nd 2016

Venue Recommendation

Asia Information Retrieval Societies



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

Content-based Scores

- 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

Content-based Scores (cont.)

for all $v_i \in V$ **do**
  for all $c_j \in C(v_i)$ **do**
    if $c_j \notin \mathrm{CM}_{pos}$ **then**
      $\mathrm{CM}_{pos} \leftarrow \mathrm{CM}_{pos} \cup c_j$
      $\mathrm{count}(c_j) = \sum_{v_s \in V} \sum_{c_k \in C(v_s)} \delta(c_j, c_k)$
      $N = \sum_{v_s \in V} \sum_{c_k \in C(v_s)} 1$
      $\mathrm{cf}_{pos}(c_j) = \mathrm{count}(c_j)/N$
    end if
  end for
end for

Content-based Scores (cont.)

Given a user u and a venue v, the category-based similarity score $S_{CM}(u, v)$ is:

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

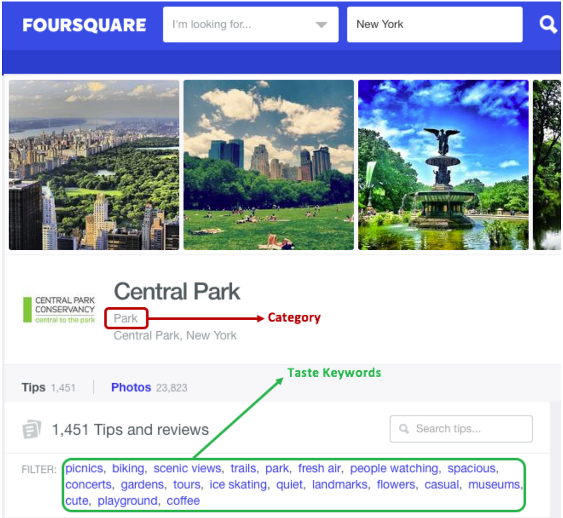
where cf_{pos} and cf_{neg} are respectively the positive and negative categories' frequencies.

Content-based Scores (cont.)

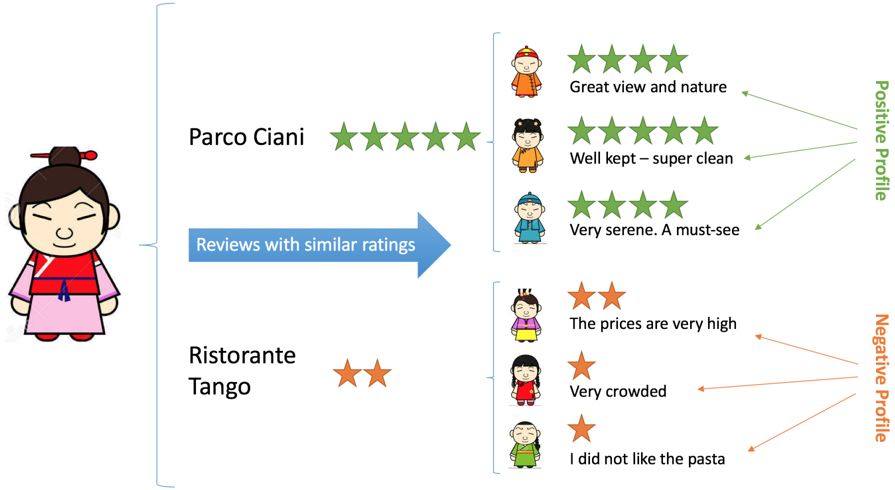
We calculate three frequency-based scores using different types and sources of information:

- Categories from Yelp: S_{CM}^{Yelp}
- Categories from TripAdvisor: $S_{CM}^{TAdvisor}$
- Venue taste keywords from Foursquare: S_{TM}

Venue Taste Keywords



Review-based Score



Review-based Score (cont.)

- 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 $\rightarrow S_{BM}$

Suggestion Ranking

- We rank the venues based on their similarity with the user
- Given user *u* and venue *v*, we calculate the similarity score as follows:

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

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

Context (cont.)

- 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.)

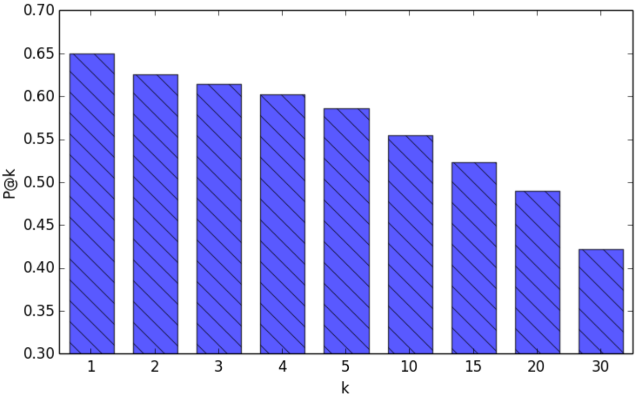
	Y	T	F
# of crawled venues	6290	4633	5534
Distribution of categories over venues			
Median	2	2	1
Mean	2.80	1.94	1.63
Variance	1.98	1.23	0.63
Distribution of reviews over venues			
Median	17	89	-
Mean	117.34	446.42	-
Maximum	6060	57365	-
Distribution of taste tags over venues			
Median	-	-	7
Mean	-	-	8.73
Variance	-	-	7.22

Results

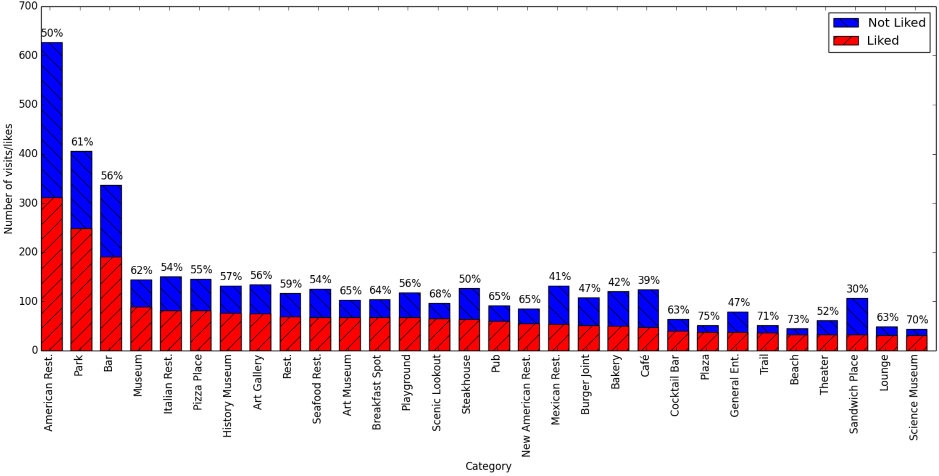
Approach	P@5 Rank	P@5	MRR
CatRev-SVM	**1**	**0.5858**	**0.7404**
CatRev-NB	7	0.5450	0.6991
BASE1	2	0.5706	0.7190
BASE2	3	0.5583	0.6815
TREC Median		0.5090	0.6716

- 17 teams - 30 runs

Analysis



Analysis (cont.)



Conclusion

- 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

Asia Information Retrieval Societies

Thanks for your attention
Thanks to ACM SIGIR for supporting my travel

Mohammad Aliannejadi
mohammad.alian.nejadi@usi.ch
@maliannejadi