Recommender Systems for Orienteering Problems

Stratis Tsirtsis* & Dimitris Fotakis

National Technical University of Athens, Greece

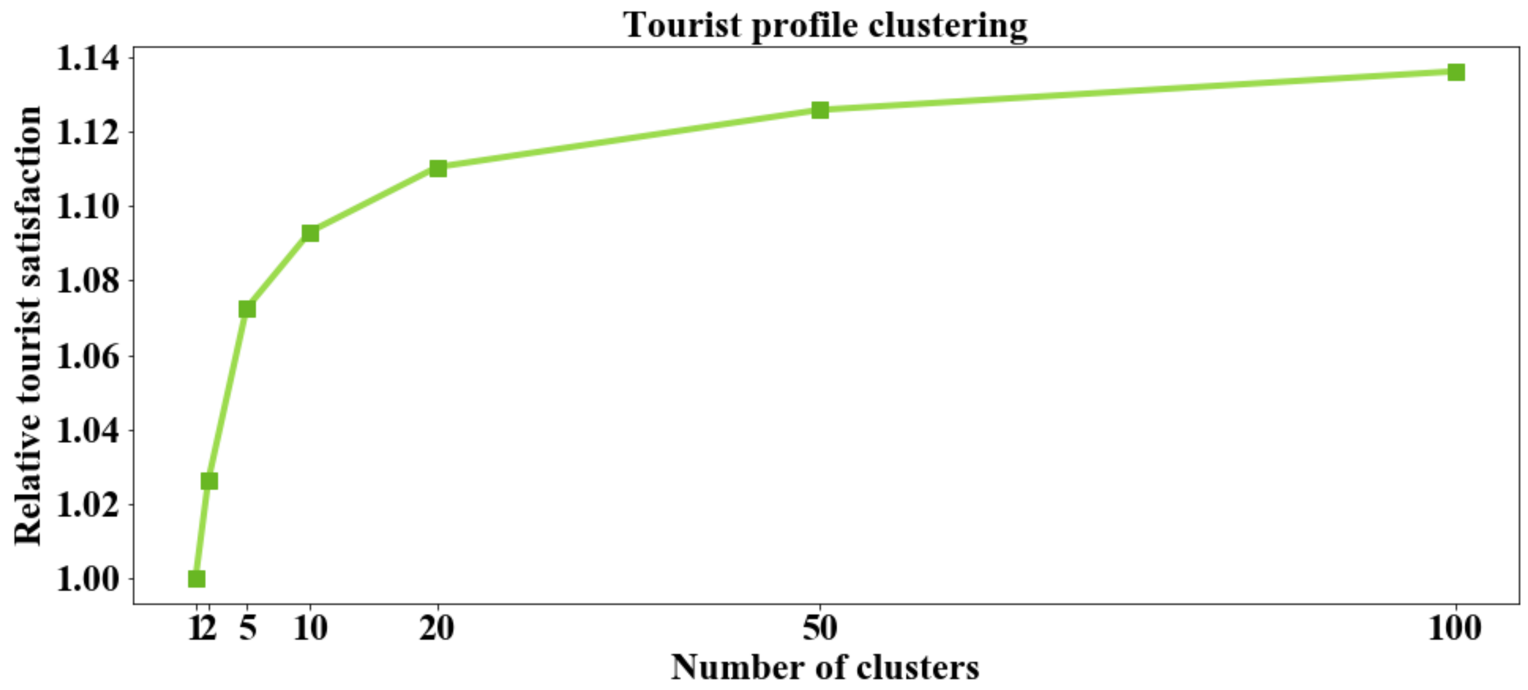
Introduction

How can we automatically design a city tour for a tourist considering the following?
- Excess of touristic attractions
- Limited time
- Personal preferences

Formalization as a Travelling Salesman Problem extension, the Orienteering Problem:
- Complete graph with nodes representing the city's Points of Interest (POIs) and weighted edges representing the roads and the respective travelling times
- Nodes accompanied by profits representing the tourist's preference to the respective POIs
- Objective: Out of all paths that satisfy a given constraint on the total weight (time budget), find the one with maximum total profit

Regarding the solution:
- NP-Hard even on its simplest variants
- Huge literature on heuristics, approximation algorithms & evolutionary methods
- **Common assumption:** Preference profile known based on previous behavior

But if a tourist visits museums in Paris, should we also recommend museums in New York? In other words, is recommendation based solely on previous personal activity optimal?



Observations:
- Fully personalized paths offer ~14% more satisfaction than a general one
- A few representative profiles (e.g. 10) are sufficient to design paths offering close to optimal individual satisfaction

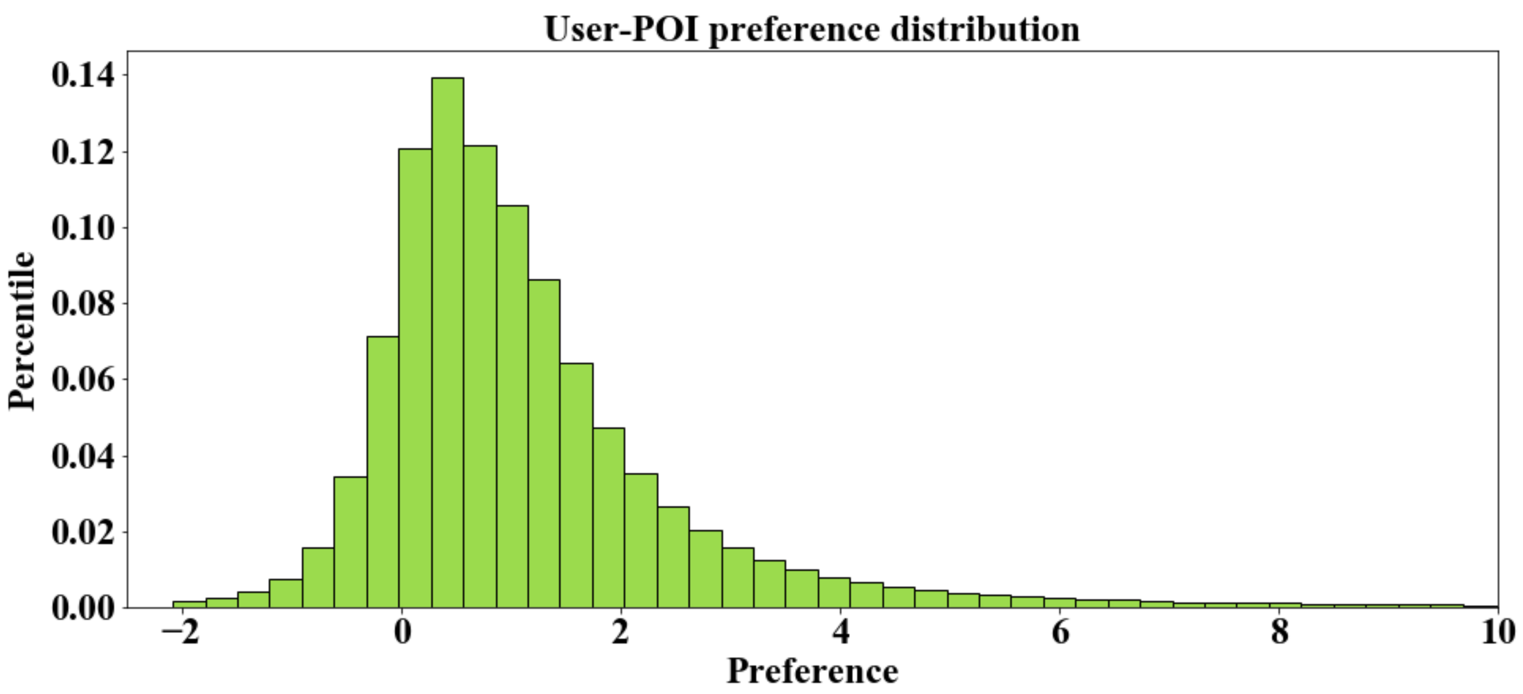
Data

Flickr + Wikipedia + **Doc2Vec Embedding**

Itinerary extraction
1)Flickr API call for geotagged photos in a given radius around city center
2)Matching of each photo with the closest geotagged Wikipedia page

Tourist & POI profiling
1)Vector embedding of all extracted Wikipedia pages → POI profile
2)Weighted (by number of photos) average of the POIs that the user visited → User profile

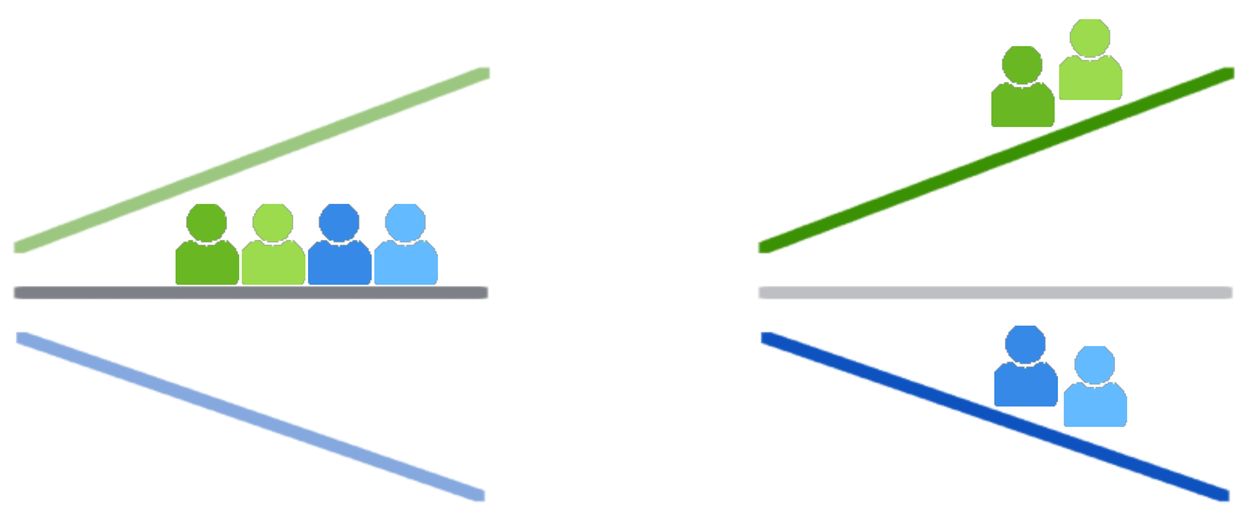
Definition: The preference (or expected satisfaction) of a user for a POI is defined as the inner product of their respective vectors



Observation: The distribution has a tail → Each tourist's profile is the weighted average of POIs being close to each other on the embedding space → Each tourist has higher preference towards specific types of POIs

Profile diversity

Clustering tourists' profiles
- If a tourist's profile is unknown but an approximate one is used to design a path, how satisfactory is that path to the target tourist compared to a fully personalized one?
- Can we group tourists by their interests?
- If a small number of paths are designed based on a representative profile for each group how does the number of different groups affect individual satisfaction?



Experiment
- Random selection of 100 tourists
- K-Means clustering of their profiles
- Computation of k paths based on each cluster's centroid
- Tourist assignment to their closest cluster's path
- Varying number of clusters from 1 (no personalization) to 100 (full personalization)

Intercity Recommendation

The Problem
- When a tourist visits a new city their profile vector is unknown and depends on the city's available attractions
- How should be the nodes' profits determined to compute a satisfactory path?

Experiment
- Tourists who have visited Cities A, B are picked as a test set and their profiles in City B are hidden
- Test set tourists visit City A, their profiles are known and personal paths are computed accordingly
- Tourists who have visited only City B are used as existing knowledge about a city, their profiles are grouped in k clusters and k representative paths are computed
- Test set tourists visit City B, their profiles are considered unknown and recommended paths are computed by the following algorithms
- Satisfaction by the recommended paths is evaluated based on their real (hidden) profile and presented as a fraction of the satisfaction by a path computed using the real profile

Baseline Algorithms
- **Popular:** Attractions' profits are equal to the number of visits by non test set tourists
- **Old Profile:** Profits are equal to the preference based on previous activity in another city
- **Random Cluster:** Test set tourist picks randomly one of the k pre-computed paths

Recommendation Algorithms
- **Most Similar Cluster:** Test set tourist picks the pre-computed path most similar to the one they followed in City A
- **Weighted Cluster:** Profile vector is picked as the average of the k centroids weighted by their respective path similarity to the test set tourist's path in City A
- **Global Average:** Profile vector is picked as the mean of all non test set tourists

#Clusters	Popular	Old Profile	Random Cluster	Most Similar Cluster	Weighted Cluster	Global Average	Optimal Cluster
5			0.8386	0.8772	0.8755		0.90979
10	0.5495	0.7687	0.8152	0.876	0.8723	0.8683	0.9251
20			0.7985	0.8657	0.8676		0.9398

Observations:
- Recommendation strategies based on popularity or solely on previous activity are far from optimal
- Our recommendation algorithms outperform all classic techniques
- Most Similar Cluster fails to predict the **Optimal Cluster** (the cluster in which the tourist's real profile belongs) but returns satisfactory results
- Personalized algorithms (Most Similar Cluster, Weighted Cluster) present a slight advantage
- Splitting tourists in many clusters gives a higher upper bound on Optimal Cluster score (following from previous section results) but recommendation algorithms' performance drops

Discussion

Open questions:
- Is there an algorithm to efficiently predict the Optimal Cluster?
- Does the profile's transformation from one city to another present some pattern? Knowing the POIs of City A and City B and the profile of a tourist at City A, can we predict the profile in City B?
- Are there extra parameters in the problem to consider?

Future work
- Application of machine learning algorithms on intercity tourist profile prediction

References

[1] T. Tsiligirides. Heuristic Methods Applied to Orienteering. Journal of the Operational Research Society, 35(9):797–809, September 1984.
[2] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. A Survey on Algorithmic Approaches for Solving Tourist Trip Design Problems. Journal of Heuristics, 20(3):291–328, June 2014.
[3] Aris Anagnostopoulos, Reem Atassi, Luca Becchetti, Adriano Fazzone, and Fabrizio Silvestri. Tour Recommendation for Groups. Data Min. Knowl. Discov., 31(5):1157–1188, September 2017.
[4] Cristina Ioana Muntean, Franco Maria Nardini, Fabrizio Silvestri, and Ranieri Baraglia. On Learning Prediction Models for Tourists Paths. ACM Trans. Intell. Syst. Technol., 7(1):8:1–8:34, October 2015.

*stsirtsis@gmail.com