Social Network Analysis of Public Lists of POIs

Ioannis Karagiannis Avi Arampatzis Pavlos S. Efraimidis Giorgos Stamatelatos
Dept. of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi 67 100, Greece
{ikaragi,avi,pefraimi,gstamat}@ee.duth.gr

ABSTRACT

In this work, we show how social network analysis can be applied to lists of points of interest (POIs) in order to extract important information about the POIs and the relations between them. More precisely, we use public lists of POIs to build the PoiGraph, a social graph of POIs, and then apply the Hyperlink-Induced Topic Search (HITS) algorithm and the Normalized Pointwise Mutual Information (NPMI) measure to estimate the user rating of each POI and the pairwise similarity between POIs, respectively. We evaluate our approach on POIs from the cities of Athens, Thessaloniki, and Rhodes. As a data source we use the corresponding publicly accessible user-specified lists of POIs of Foursquare. Our results show that for each POI the authority score obtained with the HITS algorithm is firmly correlated with the actual rating of Foursquare. Moreover, preliminary evidence shows that the NPMI-based measure gives valuable information about the pairwise similarity between POIs.

Categories and Subject Descriptors

G.2.2 [**Graph Theory**]: Graph algorithms; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

HITS, Points of Interest, Social network, Foursquare

1. INTRODUCTION

Point Of Interest (POI) ratings and recommendations are valuable to tourists. There is a lot of active research in the field of recommendations for tourism and recently in mobile systems for tourism recommendations. An up-to-date survey for mobile recommendation systems for tourism is given in [3]. A privacy-enhanced non-invasive contextual suggestion system for tourists is the Pythia system presented in [2]. Some other recent works are the myVisitPlanner^{GR} system [6], and the iGuide system [7].

PCI 2015 October 01-03, 2015, Athens, Greece

DOI: http://dx.doi.org/10.1145/2801948.2802031

In our current work, we propose a way to offer recommendations based on two different attributes of the POIs: the popularity and the pairwise similarity, using the HITS algorithm and the NPMI measure, respectively. Our methods could possibly be applied to any context where user created lists of items are available. Moreover, our methods take advantage of large amounts of user created data to extract useful conclusions about the relationships between different POIs. Finally, the dataset used in this work is publicly available online for any future reference[1].

2. EXTRACTING POI RATINGS

The HITS algorithm [4] generalizes eigenvalue centrality to allow the nodes of a graph to have two scores: a ‘hub’ and an ‘authority’ score. Each node is considered a good ‘hub’ when it has outgoing edges to important ‘authorities’, and a good ‘authority’ when it has incoming edges from important ‘hubs’. HITS is an iterative algorithm where each iteration can be summarized as:

$$\vec{\alpha}^{(k)} = L^T \vec{h}^{(k-1)} \quad \text{and} \quad \vec{h}^{(k)} = L\vec{\alpha}^{(k)} \tag{1}$$

where α and h are vectors comprising the authority and hub scores respectively, L is the adjacency matrix, and k is the step of the algorithm. Before each iteration the hub and authority scores are normalized. The above process is repeated until the scores converge for some k.

In the context of lists and POIs, it is possible to construct a biparte graph where each list is linked to the POIs it contains. By considering the lists of our graph like ‘hubs’ and the POIs like ‘authorities’, the application of HITS is straightforward. HITS provides an evaluation for both lists and POIs, which helps us degrade the effect that unimportant lists have. Additionally, each POI has only incoming links from lists, which means that in our graph each list is a source and each POI is a sink. This fact overcomes a limitation that HITS has: when some authorities are densely linked, they tend to get a very high authority score [5]. We use the JUNG framework[2] to calculate hubs and authority scores for the PoiGraph.

Authority Scores Evaluation. After applying HITS on our graph, each POI receives an authority score which we can use to rank the POIs. Similarily, we can create a second ranking by making use of the POI rating provided by Foursquare’s Web API. Then we can calculate Spearman’s

[1]https://euclid.ee.duth.gr/research/PoiGraph/
[2]http://jung.sourceforge.net

rank correlation coefficient[3] to evaluate our results. The authority scores seem to be firmly correlated with Foursquare's rating, as we achieve a coefficient of 0.59, 0.61, and 0.68, for the cities of Athens, Thessaloniki, and Rhodes, respectively.

3. ESTIMATING POI SIMILARITIES

Often, a tourist in a city would like to visit places similar to a certain POI. The recommendation of similar POIs can be accomplished for example with collaborative filtering approaches. However, this would require access to personal data of the users or to corporate databases. We propose an approach for similarity estimation between POIs based on the PoiGraph, which relies solely on publicly accessible data.

More precisely, we begin with the observation that the users of Foursquare tend to create lists with POIs that they consider similar or somehow related. This information can be used to define a user-perceived similarity between POIs. Given a pair and a list of POIs we can define two random variables where the two of them together have four possible outcomes: both POIs can be in the list, only one of them is in the list, or none of them is. We can associate each pair of outcomes by applying normalized pointwise mutual information (NPMI), an association measure which can be negative if the association is inverse [1]. However, if we imagine a list as a way to define two distinct sets on the POIs, then each NPMI has a different interpretation regarding similarity. We can argue that when two POIs belong to the same set, then it is more probable to be similar. Likewise, the outcome of belonging to different sets will have a negative effect on the similarity. More formally, if $X = 1$ and $Y = 1$ are the outcomes of the two POIs occurring in a list, we can define our similarity measure $\mathrm{sim}(x, y)$ between the two POIs x and y as:

$$\begin{aligned}\mathrm{sim}(x,y) = \mathrm{NPMI}(X=1;Y=1) - \mathrm{NPMI}(X=1;Y=0)\\ -\mathrm{NPMI}(X=0;Y=1) + \mathrm{NPMI}(X=0;Y=0)\end{aligned} \tag{2}$$

NPMI Evaluation via a User Study. A preliminary examination of our similarity score shows promising results for our method. For all the well-known POIs we achieve to relate them with POIs that belong to the same or to a closely related category, according to the categories provided by Foursquare. Moreover, our claim is further supported by a prelimilary user study, which evaluates the NPMI metric also for POIs within the same category.

The survey was carried out by selecting a sample of 10 POIs for each of the cities of Athens and Thessaloniki. Then, for each one of these POIs, a list of 5 additional POIs was generated so that they have a high NPMI-based similarity score with the main. The questionees were presented with a grid-based interface for each of the 10 "questions" and was asked to sort the 5 additional POIs from the most similar one to the least similar one. The study was performed with the help of 28 volunteers, 14 residents of Athens and 14 residents of Thessaloniki and each group completed the survey referring to their local 10 POIs. Admittedly, the sample size is very small. Thus, the statistical power of the survey results is low.

Spearman's rank correlation coefficient was utilized in order to compare the user's answer with the sorting that the MI metric provided, leading to K_i which is the average rank correlation of all answers. Furthermore, we computed the maximum rank correlation which could be achieved using the set of answers given by the volunteers K_m using a brute-force method to simulate all possible rankings for each main POI. The values of (K_i, K_m) were $(0.21, 0.40)$ and $(0.14, 0.41)$ for the cities of Athens and Thessaloniki, respectively.

4. DISCUSSION & CONCLUSION

In this paper, we leverage user created lists of POIs in order to provide suggestions. By using the HITS algorithm we were able to define a rating for the POIs, and through the utilization of normalized pointwise mutual information we were able to quantify the similarity between each pair of POIs. Our methods are deployed with respect to user privacy, as the lists are publicly available information.

Our evaluation has shown interesting, and to some extent, surprising results. Our rating of the POIs is firmly correlated with Foursquare's, which means that we have been successful in uncovering the popular ones. Moreover, by interpreting NPMI as a similarity measure, we were able to identify POIs which belong to the same category or similar ones. However, in the absence of ground truth, we have mixed results in identifying similar POIs which belong in the same categories.

In our immediate plans, we intend to apply our analysis to additional types of public user-specified lists concerning other types of goods or attractions. The option for users to specify lists of items is supported in different application domains beyond tourist attractions. For example, Amazon users can specify list of products, and imdb users can define their own lists of entities. Finally, another interesting direction for future work is to enhance our PoiGraph method with further centrality criteria and analysis approaches from the field of affiliation networks.

5. REFERENCES

[1] G. Bouma. Normalized (pointwise) mutual information in collocation extraction. *Proc. GSCL, pp 31-40*, 2009.

[2] G. Drosatos, P. S. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. N. Athanasiadis. Pythia: A privacy-enhanced personalized contextual suggestion system for tourisms. In *IEEE COMPSAC 2015*.

[3] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou. Mobile recommender systems in tourism. *J. of Netw. and Comp. Applic.*, 39(0):319–333, 2014.

[4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, Sept. 1999.

[5] S. Nomura, S. Oyama, T. Hayamizu, and T. Ishida. Analysis and improvement of hits algorithm for detecting web communities. *Systems and Computers in Japan*, 35(13):32–42, 2004.

[6] I. Refanidis et al. myVisitPlanner^{GR}: Personalized itinerary planning system for tourism. In *Artificial Intelligence: Methods and Applications*, LNCS 8445, pp. 615-629, 2014.

[7] S. Tsekeridou, V. Tsetsos, A. Chalamandaris, C. Chamzas, T. Filippou, and C. Pantzoglou. iGuide: Socially-enriched mobile tourist guide for unexplored sites. In *Artificial Intelligence: Methods and Applications*, volume 8445 of *LNCS*, pages 603–614, 2014.

[3] http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient