Tourismo:A User-Preference Tourist Trip Search Engine

Gregor Jossé[1], Klaus Arthur Schmid[1], Andreas Züfle[1], Georgios Skoumas[2], Matthias Schubert[1], Dieter Pfoser[3]

[1] Ludwig-Maximilians-Universität München, Germany
{josse,schmid,schubert,zuefle}@dbs.ifi.lmu.de,
[2] National Technical University of Athens, Greece
gskoumas@dblab.ece.ntua.gr,
[3] George Mason University, USA
dpfoser@gmu.edu

Abstract. In this demonstration we re-visit the problem of finding an optimal route from location A to B. Currently, navigation systems compute shortest, fastest, most economic routes or any combination thereof. More often than not users want to consider "soft" qualitative metrics such as popularity, scenic value, and general appeal of a route. Routing algorithms have not (yet) been able to appreciate, measure, and evaluate such qualitative measures. Given the emergence of user-generated content, data exists that records user preference. This work exploits user-generated data, including image data, text data and trajectory data, to estimate the attractiveness of parts of the spatial network in relation to a particular user. We enrich the spatial network dataset by quantitative scores reflecting qualitative attractiveness. These scores are derived from a user-specific self-assessment ("On vacation I am interested in: family entertainment, cultural activities, exotic food") and the selection of a respective subset of existing POIs. Using the enriched network, our demonstrator allows to perform a bicriterion optimal path search, which optimizes both travel time as well as the attractiveness of the route. Users will be able to choose from a whole skyline of alternative routes based on their preference. A chosen route will also be illustrated using user-generated data, such as images, textual narrative, and trajectories, i.e., data that showcase attractiveness and hopefully lead to a perfect trip.

1 Introduction

Nowadays, social networks are a great source of rich geo-spatial data. Almost every social network allows users to incorporate geo-social features into their data stream. The different features include, amongst others, geo-tagged pictures (e.g., Flickr), geo-descriptive text (e.g., travel blogs), and tracked movement (e.g., runners' trajectories). For this demo, we rely on all these kinds of user-generated data to define attractiveness on a real world road network. Our aim is to reflect human fondness according to the crowd by using qualitative information and making it measurable. We present Tourismo, a tourist search engine, which computes attractive paths along points of interest (POIs), tailored to the interest of the user issuing the query. Based on this enriched spatial network, which has information about the attractiveness of locations, we aim at answering *attractive path queries*. Currently, navigation systems, i.e., machines, perform this task for us, computing routes such as the shortest route, the fastest route, the most economic route [1], or some combination of such quantitative measure on a spatial network [2]. In all of these cases, the employed algorithms optimize cost measures inherent in the underlying road network. What is rarely reflected, however, is user preference on subjective measures, such as attractiveness and interestingness of a route. Often users are willing to take a suboptimal detour, a deviation from quantitative optimality (shortest, fastest, etc.), in order to improve the quality of their route. In order to see more attractions, for instance, a tourist may be willing to take a moderate detour from a fast, but not very attractive, highway.

How can we measure a subjective concept of "quality"? How to measure attractive, scenic, recreative routes? As machines are not (yet) capable to reflect this concept, we rely on the crowd to answer this question, i.e., we propose to use crowdsourced data to estimate the attractiveness of an area. Relying on different datasets, image data (from Flickr[1]), textual narratives (from travel blogs), and trajectory data (from Endomondo[1]), we investigate the applicability of different data sources as cost measures for the underlying road network. More precisely, we enrich the road network by quantitative scores of qualitative statements as follows:

- areas having a large density of Flickr images indicate a particularly attractive area, increasing the attractiveness score;
- locations mentioned in the positive context of travel blogs increase attractiveness scores;
- routes commonly used by other users are also considered more attractive.

Furthermore, we incorporate meta-information from OpenStreetMap[1] (OSM), in order to categorize POIs and, using the aforementioned popoularity score, propose routes according to the user's preferences and the fondness of the crowd. Tourismo presents solutions to enrich the underlying road network using the data sources. We show an initial approach to map these *attractiveness scores* to cost measures correlated with travel time, allowing to apply existing routing algorithms which aim at minimizing edge-labeled cost metrics. We apply algorithms for pareto-optimal route search similar to [3] and [4], to find paths which are optimal w.r.t. the popularity scores. Our framework allows to specify origin and destination, computes and displays the skyline of pareto-optimal paths. Furthermore, the reasons for attractiveness of each path are illustrated: Flickr images along the way, travel blog entries mentioning locations on the way, and historical trajectories which share the same route. Our demonstrator is an extension of [5] it has three major features: First, we incorporate new, route search algorithms which enable higher dimensional cost spaces at the same reducing computation time. Second, the demonstration allows to specify the interest of a user, thus returning routes that contain POIs which are of particular interest to the user. Third, this version considers a third type of data to enrich the underlying road network with attractiveness information: In addition to geotagged images, and texts containing geospatial references, we also learn attractiveness from an existing base of historic trajectory data.

2 State of the Art

Recently, a lot of interesting research has been done in the context of finding scenic, interesting or popular routes. The first set of related work focuses on providing paths which are easier to memorize, describe, and follow. For example, the authors of [6] and [7] try to tackle the problem by introducing cost criteria that allow for a trade-off between minimizing the length of a path while also minimizing the complexity in

[1] www.flickr.com, www.endomondo.com, www.openstreetmap.org

terms of instructions or turns along the path. Furthermore, an existing research direction covers the problem of defining tourist routes, which maximize the subset of a set of predefined POIs which can be visited in a tourist tour that has a time-constraint [8,9]. In these works, the set of interesting POIs is given, and the main conceptual contribution of is to automatically extract interesting locations, as well as a quantitative estimate of the popularity of this location from a variety of data sources. Another research direction, which is not necessarily restricted to touristic routing but lacks the aspect of qualitative measures, are the Trip Planning Query and variants thereof [10,11].

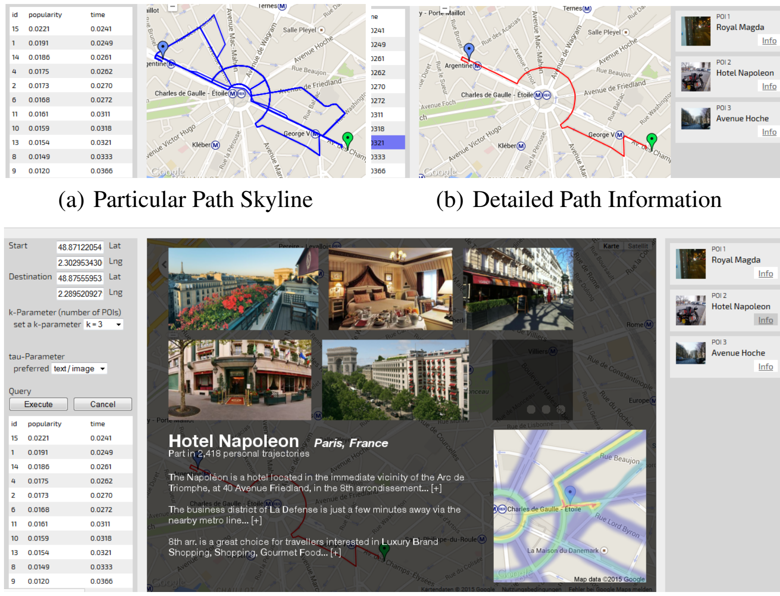
The approach most similar to the one presented in this work is [12], which proposes a method for computing beautiful paths, as the authors phrase it. However, in order to quantify quality, the authors rely on explicit statements about the beauty of specific locations, obtained from a crowd-sourcing platform which collects user opinions on photos of specific locations. In contrast, we propose to mine this kind of information from existing crowd-sourced data, which does not require any monetary investment to aquire. Thus our approach has the crucial advantage that it is scalable as the used data is already available globally available, while having local expert users rate photos one by one can hardly be extended to a global scale.

Another important research direction is the *stitching* existing trajectories in order to obtain new trajectories which guarantee that each sub-trajectory is used by other users, and is thus, "popular" following the definition [13] of Chen et al. This, however, only reflects a notion common usage, not taking into account, why a specific sub-trajectory has been favored. For instance, when mining trajectories of commuters, the fastest path is most likely to be chosen by most users. Hence, we propose mining trajectories specific to recreational use and merging this information with the attractiveness scores we derive from other user-generated data sources.

3 Features

The main feature of this demonstrator is the estimation of attractiveness from text, image, and trajectory data. Details covering text and image data can be found in [5]. In this section, we briefly describe how we enrich the underlying road network using historical trajectory data. For our demonstrator, we use trajectories of walkers, runners and bikers that have uploaded their workouts to Endomondo. Our dataset contains eight million trajectories, which are located all around the world, but have a strong regional focus in Northern Europe. To match each of the GPS trajectories, we apply state-of-the-art map matching techniques, similar to those presented in [14]. In a first step, we perform a basic enrichment: For each edge e of the spatial network, we count the number tra(e) of historical trajectories that contain this edge. This count can be used as an indication of attractiveness of the nodes delimiting the edge, following the assumption that runners are, in average, more likely to choose a particularly nice running trail. We assign vertices in a road network attractiveness scores derived from different datasets. We refer to the score of a vertex v derived from Flickr image data as im(v), to the score derived from travel blog text data as txt(v), and to the score derived from trajectory data as tra(v). In contrast to [5], where the different data source scores were combined into one cost measure, we now propose to diversified measures. For each edge $e = (u, v)$ and each of the scores $f \in \{$im,txt,tra$\}$ we define:

$$p_f(e) = tt(e) \cdot \phi^{C_f(f(u)+f(v))}$$



(a) Particular Path Skyline (b) Detailed Path Information

(c) Detailed information about selected PoI

Fig. 1. Functionality of the presented framework.

where $tt(e)$ denotes the travel time alonge e, C_f denotes a scaling parameter dependent only on the data source, and $\phi \in (0, 1)$ is a scaling factor for the influence of the respective attractiveness score. Hence, we obtain three travel time correlated cost measures reflecting notions of attractiveness according to the different data sources. Consequently, we may query the enriched road network, computing pareto-optimal paths as presented in [3,4] w.r.t. to the introduced cost measures.

Additionally, Tourismo features category-specific path queries. If the user chooses to specify his personal touristic interests, they can choose one or more options from a list containing outdoor activities, cultural sightseeing, culinary interest, and more. In order to provide paths which fulfill these requirements, we mine the OSM meta-information. Thanks to a very active community, the data contains well-tended information about POIs, that is named, categorized, and subcategorized. For instance, the categories "food" and "tourist" contain subcategories "bar", "restaurant", "fastfood" and "monument", "museum", "archeological", respectively. Mapping these categories onto the preferences, we filter POIs which correspond to the particular interest of the user. When querying a route with a specific set of interests, the user is provided a number of pareto-optimal paths, guiding him along POIs tailored to his preference.

4 Framework Description

The demonstrated framework allows users to validate that the notions of attractiveness defined in this paper indeed coincide with the general intuition. The result paths returned to the user yield competitive solutions in terms of travel time while passing POIs perceived as significant, appealing, and/or recognizable. Using OSM as a road network, our demonstrator visualizes a map relying on Google Maps. Upon selecting an origin and a destination location on the map, the user is presented with the skyline view as shown in Figure 1(a). In this view, the path skyline is presented to the user. For each such route, the corresponding cost values are shown in a table in the lower left corner of Figure 1(a). Using this table, the user can browse the choise and select a desired routeyielding the route view shown in Figure 1(b). For the selected route A, this view shows the most "popular" points of interest on A. Once a point of interest is selected, the sources of popularity of this POI are displayed, as shown in Figure 1(c). For this purpose, Figure 1(c) shows all the pictures relevant for the selected POI, i.e., the set of images having a sufficiently low distance. The bottom-left corner shows all travel blog entries where this entry was mentioned in a positive context. Finally, the lower left corner shows a heatmap derived from all trajectories that share the same trajectory. During the demonstration, users will be able to specify start and target locations (also, if desired, specific categories of interest) and compute different sets of skyline paths.

Acknowledgements

This research has received funding from the Shared E-Fleet project (in the IKTII program) by the BMWi (grant no. 01ME12107), from the DFG (grant no. RE 266/5-1), from the DAAD supported by the BMBF (grant no. 57052426). Dieter Pfoser has been partially supported by NGA NURI (grant no. HM02101410004).

References

1. Andersen, O., Jensen, C.S., Torp, K., Yang, B.: Ecotour: Reducing the environmental footprint of vehicles using eco-routes. In: MDM. (2013) 338–340
2. Graf, F., Kriegel, H.P., Renz, M., Schubert, M.: Mario: Multi-attribute routing in open street map. In: SSTD. (2011) 486–490
3. Shekelyan, M., Jossé, G., Schubert, M.: Linear path skylines in multicriteria networks. In: ICDE15. (2015) 459–470
4. Shekelyan, M., Jossé, G., Schubert, M.: Paretoprep: Efficient lower bounds for path skylines and fast path computation. In: SSTD15. (2015)
5. Jossé, G., Franzke, M., Skoumas, G., Züfle, A., Nascimento, M.A., Renz, M.: A framework for computation of popular paths from crowdsourced data. In: ICDE. (2015) 1428–1431
6. Sacharidis, D., Bouros, P.: Routing directions: keeping it fast and simple. In: ACM SIGSPATIAL GIS. (2013) 164–173
7. Westphal, M., Renz, J.: Evaluating and minimizing ambiguities in qualitative route instructions. In: ACM SIGSPATIAL GIS. (2011) 171–180
8. Garcia, A., Arbelaitz, O., Linaza, M.T., Vansteenwegen, P., Souffriau, W.: Personalized tourist route generation. Springer (2010)
9. Gavalas, Konstantopoulos, C., Mastakas, K., Pantziou, G.: A survey on algorithmic approaches for solving tourist trip design problems. Journal of Heuristics 20 (2014) 291–328
10. Kanza, Y., Safra, E., Sagiv, Y., Doytsher, Y.: Heuristic algorithms for route-search queries over geographical data. In: ACM SIGSPATIAL GIS. (2008) 11
11. Chen, H., Ku, W.S., Sun, M.T., Zimmermann, R.: The partial sequenced route query with traveling rules in road networks. Geoinformatica 15 (2011) 541–569
12. Quercia, D., Schifanella, R., Aiello, L.M.: The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. CoRR (abs/1407.1031) (2014)
13. Chen, Z., Shen, H.T., Zhou, X.: Discovering popular routes from trajectories. In: ICDE. (2011) 900–911
14. Newson, P., Krumm, J.: Hidden markov map matching through noise and sparseness. In: ACM SIGSPATIAL GIS. (2009) 336–343