Interactive Route Personalization Using Regions of Interest

Harm Delva[1], Annelien Smets[2], Pieter Colpaert[1], Pieter Ballon[2], and Ruben Verborgh[1]

[1] IDLab, Department of Electronics and Information Systems, Ghent University – imec
[2] imec-SMIT, Vrije Universiteit Brussel

Abstract. There is an abundance of services and applications that find the most efficient route between two places, people are not always interested in efficiency; sometimes we just want a pleasant route. Such routes are subjective though, and may depend on contextual factors that route planners are oblivious to. One possible solution is to automatically learn what a user wants, but this requires behavioral data, leading to a cold start problem. An alternative approach is to let the user express their desires explicitly, effectively helping them create the most pleasant route themselves. In this paper we provide a proof of concept of a client-side route planner that does exactly that. We aggregated the Point of Interest information from OpenStreetMap into Regions of Interest, and published the results on the Web. These regions are described semantically, enabling the route planner to align the user's input to what is known about their environment. Planning a 3 km long pedestrian route through a city center takes 5 seconds, but subsequent adjustments to the route require less than a second to compute. These execution times imply that our approach is feasible, although further optimizations are needed to bring this to the general public.

Keywords: Linked Open Data · Point of Interest · Route Planning · OpenStreetMap

1 Introduction

Route planning applications have become so common that many, such as Google Maps, have almost become household names. The majority of these applications focus on generating the most efficient route, even though people are not always interested in efficiency. Some people may want to enjoy the first days of good weather by making a small detour through a local park, others may want to avoid the same park because of allergies. There are also pragmatic reasons for wanting to avoid certain places: shopping streets can get overcrowded on Saturday afternoons, and parks may also get overcrowded – which can be a problem during pandemics for instance.

The list of examples goes on, which is exactly the problem we address in this paper: the ideal route depends on personal preference and contextual factors, many of which are subjective and change over time. Instead of trying to predict what a user wants, we propose to let them specify their preferences explicitly through an interactive application.

2 Related Work

The subjective value of routes is not a new scientific subject; philosophers in 1955 even coined the term *psychogeography* for the study of how people experience their environment [5], and by extension, how they experience navigating through it. Research in this field addressed questions such as how people perceive their commute to work [1] or how they value their time on the bus [7]. Today, the planning of personalized routes remains an open research question, with some researchers claiming that "insufficient criteria modeling for a personalized system" is one of the major difficulties [9], while others note that the innate flexibility of such route planners forms a computational problem [6].

One notable work in this field is aptly named "The Shortest Route to Happiness" [11], where the authors used quantitative measures of how people perceive different locations to recommend emotionally pleasing routes to their destination. Their user study revealed two important lessons for further research: on one hand, it confirms that users can indeed appreciate pleasant routes over efficient routes. However, they nuanced this by adding that context plays an important role. For instance, the pleasantness of a location changes throughout the day, and not everyone appreciates crowded shopping streets. The authors suggested that future work could focus on incorporating these preferences through personalizaton.

The latter is, however, a delicate operation. Several scholars have raised concerns about algorithmic personalization based on users' historic preferences. They argue that this kind of personalization might result in a decrease of exposure diversity, a problem that is known as the so-called filter bubble [3, 8, 10]. In an urban environment, this results in people being exposed to different parts of the city [13], which deteriorates the opportunities that arise in diverse urban contexts and could in turn reinforce societal stereotypes. To overcome this, scholars have suggested to increase the incompleteness of the digital information environment [2]. They argue that this invites users to explore more actively and thus provides the opportunity to escape their bubble.

3 Method

We propose to drop the assumption that user preferences have to be predicted, or learned, and instead develop a proof of concept application that asks the user for their preferences explicitly. Users have the ability to personalize their own route, which is not only a solution to the data availability problem, it also puts the user back in control of the results they see.

At the core of our approach lie *Regions of Interest*, which contain several *Points of Interest* (POIs) such as shops, public parks, restaurants, etc. These

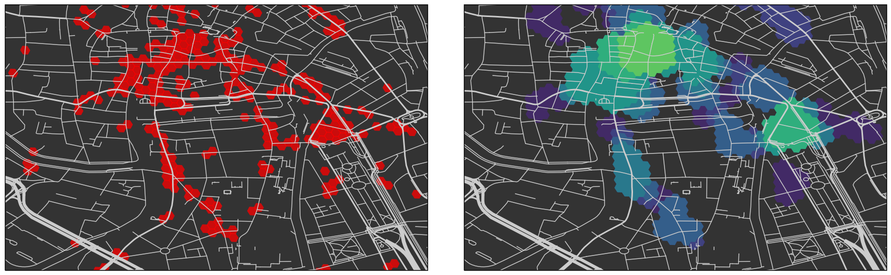


Fig. 1. Visualized on the right are the clusters that are found by running the clustering algorithm on the cells of entities that have a `shop` tag, visualized on the left. If a sufficient amount of cells are sufficiently close to each other, they are placed in the same cluster. This is a variation of the well-known DBSCAN algorithm, and choosing appropriate values for these parameters can be challenging. However, it is made easier because we are working on a discrete input space. If we only consider directly adjacent cells as neighboring, there can be at most 6 cells in a cell's neighborhood, and by running the algorithm for each of these values we obtain a set of clusters of increasing strictness. We have found good results by setting the maximum distance to 2, in which case there are at 18 possible neighboring cells, and we add a padding layer of 1 cell around every found region.

regions are obtained by placing the publicly available POI data from the OpenStreetMap project[3] of a given type (e.g., all shops) onto a discrete grid[4], and clustering the grid cells using a variation of DBSCAN. The discrete grid forms a unified representation for all locations, regardless of whether they are described as simple points or as multi-polygons, while a notion of density is retained by performing the clustering with increasingly strict parameters, as illustrated by Figure 1. Every region is then semantically described, as in Listing 1, and the results are published through a Linked Data Fragments interface, similar to our existing Routable Tiles dataset [4].

```
[
    {
      "@type": "owl:Restriction",
      "owl:onClass": "osm:Entity",
      "owl:onProperty": "osm:hasTag",
      "owl:someValuesFrom": [
        "taginfo:shop=bicycle",
        "taginfo:shop=hardware",
        ...
      ],
      "@id": "_:CommercialEntity"
    },
```

[3] Available as Linked Data Fragments, e.g. at https://opoi.org/14/8411/5485/
[4] Using the H3 spatial index, see https://h3geo.org/

```
    {
      "@id": "_:fuzzy_subject",
      "rdfs:subPropertyOf": "dct:subject",
      "truth:degree": 0.2222222
    },
    {
      "geo:asWKT": "POLYGON ((3.0016827 51.2498880, 3.0018188
       51.2496799, ...",
      "_:fuzzy_subject": "_:CommercialEntity"
    }
]
```

Listing 1. JSON-LD representation of a single Region of Interest. Each region is described using terms from the GeoSPARQL vocabulary, and a `subject` property from dcterms refers to a description of the region's contents. These descriptions use OWL restriction classes to group similar kinds of POIs together. To model the specificity of different regions, we approximate the semantics of fuzzy sets by adding a self-defined `truth:degree` value to the `subject` properties, using the clustering strictness as a proxy for the membership value.

4 Demonstrator

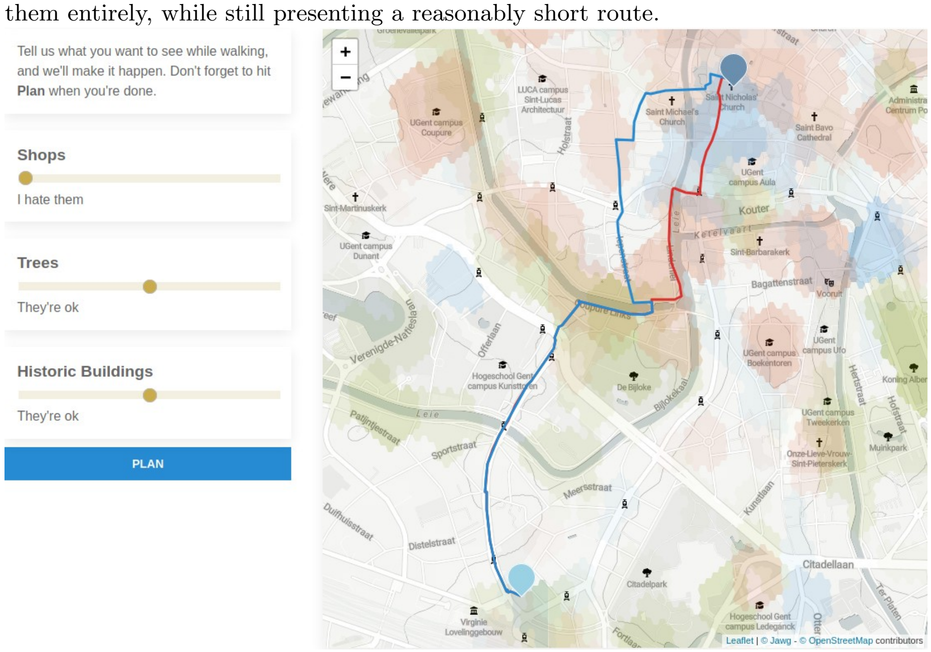
A live version of the demonstrator is available at https://hdelva.be/sem4tra2020/demo.html, or through the HTML version of this paper at https://hdelva.be/articles/regions-of-interest-demo/. The demo works on all devices with a modern web browser, including mobile devices (in landscape mode), and computing a route for the default scenario takes roughly 5 seconds with subsequent adjustments taking less than a second.

Figure 2 contains a screenshot of the application, as well as a description of how it works. We let users provide their current preferences explicitly, which are then cross-referenced with a set of Regions of Interest to add additional weights to a road network, and Dijkstra's algorithm is used to perform the route planning itself. Although we only support pedestrian routing for the moment, the same principles can be applied to other modes of transport by constructing the base road network graph accordingly.

5 Discussion

As mentioned in Listing 1, our route planner only has a limited view of reality. It has a vague idea of where the shops are, but not which shops exactly, or how many. This information could be added through an additional data source, however the imprecise aggregated data has some interesting implications. The obvious one is that using aggregated data is less resource intensive; there is less data to process, and the processing itself is easier. One could start from the raw data, but this would add more workload to the client, and as noted in section 2, personalized route planners are already computationally intensive.

Fig. 2. A screenshot of the demonstrator calculating a route from the historic center of the city of Ghent to the train station, roughly 3 km apart. The different kinds of Regions of Interest are visualized on the map: green indicates nature, blue indicates shops, and orange-red indicates historic buildings. The red route shows the most efficient route, the blue route incorporates the user's preferences. Note that the main shopping streets are just to the south of the historic center, and that the personalized route circumvents them entirely, while still presenting a reasonably short route.



Working entirely with intentionally imprecise data can also be a pragmatic solution to working around data quality issues. For example, it does not matter if the base data contains 20 out of 50 restaurants on a single street – it is a street with many restaurants. Clients are also shielded from the data model of the raw data, which means that even non-semantic data sources can be used, assuming some mapping happens during the aggregation.

The incompleteness of the results should also cause users to more actively explore their environment by presenting less complete information. Ultimately, this apparent loss of user experience should result in unexpected discoveries, users might be presented with serendipitous discoveries that are known to increase overall user satisfaction [12].

6 Conclusion

We have built a publicly available application that plans a pedestrian route from the historic center of Ghent to the train station, using aggregated Point of Interest data from the OpenStreetMap project. This application is just one potential use-case of the underlying principles though, and it hopefully it will serve as a basis for future work. For instance, the same data can be reused to build a different application that builds itineraries for tourists, or to visualize characteristics of a city. So far we have only used data from a single source, but the benefits of using Linked Data will really come to fruition if data from other sources, such as official databases of cultural heritage sites, is added as well. In fact, the Regions of Interest are defined in terms of externally defined concepts, so that anyone can generate them from any sort of geospatial data, as long as the source data can be described semantically as well.

References

1. Algers, S., Hansen, S., Tegner, G.: Role of waiting time, comfort, and convenience in modal choice for work trip. Transportation research record **534**, 38–51 (1975)
2. Björneborn, L.: Three key affordances for serendipity. Journal of Documentation (2017)
3. Bozdag, E.: Bias in algorithmic filtering and personalization. Ethics and information technology **15**(3), 209–227 (2013)
4. Colpaert, P., Abelshausen, B., Meléndez, J.A.R., Delva, H., Verborgh, R.: Republishing openstreetmap's roads as linked routable tiles. In: European Semantic Web Conference. pp. 13–17. Springer (2019)
5. Debord, G.: Introduction to a critique of urban geography. Praxis (e) press (2008)
6. Funke, S., Storandt, S.: Personalized route planning in road networks. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 1–10 (2015)
7. Horowitz, A.J.: Subjective value of time in bus transit travel. Transportation **10**(2), 149–164 (1981)
8. Jiang, R., Chiappa, S., Lattimore, T., György, A., Kohli, P.: Degenerate feedback loops in recommender systems. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. pp. 383–390 (2019)
9. Niaraki, A.S., Kim, K.: Ontology based personalized route planning system using a multi-criteria decision making approach. Expert Systems with Applications **36**(2), 2250–2259 (2009)
10. Pariser, E.: The filter bubble: How the new personalized web is changing what we read and how we think. Penguin (2011)
11. Quercia, D., Schifanella, R., Aiello, L.M.: The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In: Proceedings of the 25th ACM conference on Hypertext and social media. pp. 116–125 (2014)
12. Reviglio, U.: Serendipity as an emerging design principle of the infosphere: challenges and opportunities. Ethics and Information Technology **21**(2), 151–166 (2019)
13. Smets, A., Montero, E., Ballon, P.: Does the bubble go beyond? In: Proceedings of The 1st Workshop on the Impact of Recommender Systems with ACM RecSys 2019 (2019)