Recommending Alternative Cycling Routes via Predicted Usage Patterns

Dakai Men[1], Jannis Becktepe[1], Mahdi Esmailoghli[1], David Bermbach[2], Ziawasch Abedjan[1]

1 Introduction

Assume you have just moved to Berlin and you are looking for a healthy, fun, and eco-friendly way to get around. You know that biking is a reasonable option, but where do you even start? One can start with the most obvious option: Google Maps. Although Google Maps is the most popular choice, is it also the best way to discover the hidden gems of the city? Not always.

As it turns out, Google Maps may not always lead the cyclist down the best path. When one new traveler decided to bike from *Kunsthalle Berlin* to *Technische Universität Berlin*, Google Maps recommends a loud, traffic-packed, and bumpy route shown in the Fig. 1.

However, the power of local knowledge can assist bike riders. With the help of a colleague who knew all the best biking routes in the city, they discovered a much better, safer, and more scenic path as shown in Fig. 2.

While technology has made our lives easier in many ways, it pays off to trust the locals, i.e., domain experts. After all, local cyclists know their city better than any machine-generated general-purpose navigation tool. Therefore, the locals can recommend routes that a new-to-town traveler would never discover on their own.

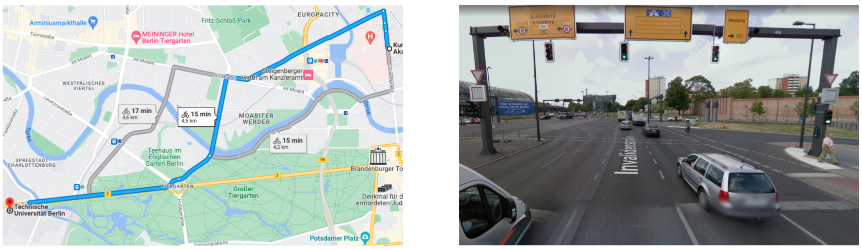

Fig. 1: Route proposed by Google Maps.

In this blog post, we will share with you the importance of routing services designed specifically for cyclists and how they can make a significant difference in the safety of bike riders.

[1] Leibniz Universität Hannover, L3S, Germany {men,becktepe,esmailoghli,abedjan}@dbs.uni-hannover.de
[2] TU Berlin, Germany david.bermbach@tu-berlin.de

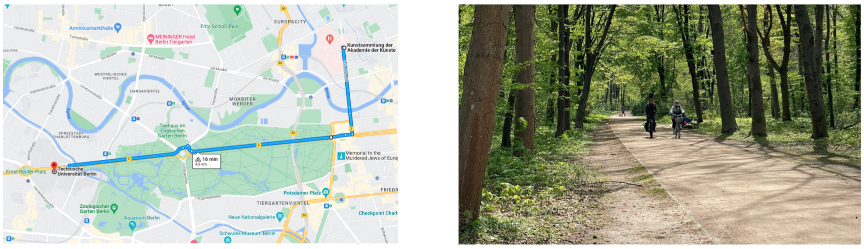

Fig. 2: Route proposed by a local.

As shown in our example above, many routing services mainly recommend large streets with a huge amount of traffic rather than small and neat paths that you would prefer to enjoy during your cycling trip. While traditional yet common navigation systems may work perfectly well for drivers, they often fail to provide cyclists with the safest and most efficient routes.

Several of these differences between drivers and cyclists include: (i) Cyclists are not as safe as car drivers because bikes are not as equipped with safety features as modern cars. Therefore, cyclists tend to bike on quieter and safer streets. (ii) The navigation for bike riders is a multi-goal optimization problem. The most important objective of car navigation systems is driving time, as car drivers prefer to arrive at their destination as quickly as possible. However, this objective is more complex for bike riders, who not only would like a faster route, but also prefer to cycle on flat surfaces, quiet areas, and scenic streets. As a result, most of the current bike navigation systems, such as Google Maps, are not used by users to find the best paths. (iii) Even if the goal of the biker is to arrive quickly at the destination, there are always shortcuts that not only navigation systems cannot recommend, but also require local knowledge.

Our goal is to provide cyclists with safety, environmental features, and local knowledge in recommending a path from a starting point to a destination. We participated in the BTW Data Science Challenge 2023 to introduce a comprehensive solution that recommends bike paths similar to those taken by local cyclists.

To discover safer and more suitable routes for cycling we leverage the data generated by many volunteers using tracking apps to log their bike riding behavior.

To achieve our goal, we analyze how local cyclists in certain cities select their routes. Based on this knowledge, we transfer this information to new cities where bike-riding data is not available. We obtain the driving factors of cyclists when selecting routes and understand how to support them in the future, both when recommending routes and building cycling infrastructure.

We leverage usage data collected by the SimRa app in Berlin to understand which roads were taken by cyclists, including tracked statistics such as speed, location, and possible incidents during several bike rides. But this information is not enough to understand why certain routes were taken. Therefore, we combine the usage data with additional environmental data, such as weather and road information, and train a machine learning model to analyze the behavior of local users based on these environmental features. By learning the behavior of local riders, we can identify patterns that influence the selection of certain routes. For instance, if our model learns that local bike riders avoid sandy roads when it rains, we can apply this rule to recommend biking paths in other cities where usage data is not available but weather and road information is publicly accessible.

This enables us to only reason about routes taken by locals. However, the paths that are not taken by the locals might also contain information about environmental properties that prevent users from selecting certain roads. To also incorporate the information about certain routes that were not taken chosen by locals, we add alternative routes to our usage data. If we consider our running example, we would also add the route proposed by Google Maps to understand why the local does not select it.

Our model is ultimately trained to rank the possible routes between a given starting point and a destination based on their similarity to the potential locals. We call this similarity the safety score since it is directly related to the safety of cyclists[1].

We build our system and train the model on the Berlin SimRa dataset and utlimately apply the results in Hannover city. In this blog post, we will discuss several of the results achieved during our participation in the BTW Data Science Challenge.

2 System

Figure 3 illustrates the workflow of our system, which consists of three phases: the pre-processing phase, the analysis phase, and the inference phase.

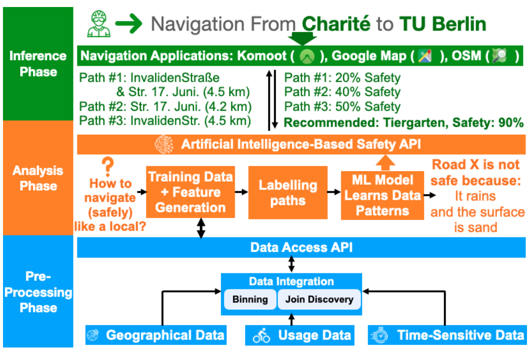

Fig. 3: The data flow of our proposed system.

[1] https://ecf.com/news-and-events/news/new-survey-people-cycle-most-where-they-feel-its-safe

During the pre-processing phase, we integrate data from various sources and organize it into geographical or time-based bins. We also identify join keys [EQA21; EQA22] between sources to gather environmental data, such as weather and street information, as well as usage data derived from local cyclists that implicitly encode their preferences and behavioral patterns.

During the analysis phase, the system fetches route information and labels it based on usage data. When a local cyclist selects a route, it is considered a preferred path. The system then augments the training data with additional environmental data and uses it to train a machine learning model. The model learns the relationship between preferred paths and environmental features, serving two purposes: explaining why a particular path is selected [Es19; Me] or ignored by a local cyclist and reusing learned patterns in areas where usage data is scarce.

Finally, in the inference phase, our model can contribute to current navigation apps by evaluating possible routes based on the likelihood of expert cyclists selecting them and predicting usage patterns when data is missing.

3 Results

In this section, we would like to share with you the key insights we gained from our participation in the data science challenge.

Firstly, let's talk about **incident levels**. The SimRa dataset not only provides us with route information, but also includes incidents that took place during the rides. These incidents are reported by local cyclists, who categorized them into groups labeled between 0 and 8 based on the incident type. We call this label the incident level. According to our observation, the higher the incident level, the more severe it is.

Our evaluation focused on comparing the average incident level for both local and alternative routes. The routes taken by locals have an average incident level of 0.6 in comparison to the incident level of 1.2 for alternative paths. This experiment shows that local cyclists tend to choose safer routes compared to those recommended by state-of-the-art bike navigation systems. The average incident level for the routes they select is lower.

Now we discuss the **most important features**. In order to gain a better understanding of the routes preferred by locals, we utilized a machine learning model to predict the likelihood of a route being selected. Our analysis focused on identifying the key characteristics that distinguish safe and unsafe routes. Figure 4 presents the three most important features of route selection.

Dedicated bicycle lanes were found to be the most important feature in route selection, followed by the presence of protective strips and bike occurrences on the route. We can assume that dedicated bicycle lanes provide adequate safety for cyclists by isolating them from vehicle traffic, which is highly preferred by local riders. Additionally, our analysis showed that local cyclists tend to favor areas with lower bicycle traffic.

Unfortunately, modern navigators for cyclists often do not prioritize bicycle lanes or bike traffic information, which results in less safe and desirable routes being recommended to riders.

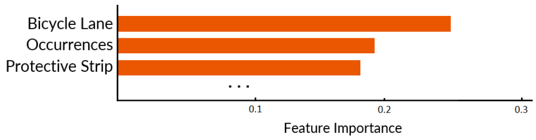

Fig. 4: Top 3 most important features.

In addition to analyzing feature importance, we also examined the statistical differences among various features. Our analysis revealed that alternative routes, i.e., routes recommended by navigation tools, have nearly four times as many tram tracks as the selected routes and are situated in areas with a population that is approximately double.

As railways can pose a hazard to bicyclists by trapping their wheels and high-population areas often result in heavy traffic, avoiding railways can significantly reduce the likelihood of incidents or injuries. However, these factors are often not given enough consideration when determining a safe route, despite the significant impact they can have on safety.

Let's now turn our attention to the concept of **transferability**. In the final phase of our experiment, we sought to test whether our system could be applied to a new city with limited usage data available. To do this, we trained our model using usage and environmental data from Berlin and then evaluated its performance in predicting route preferences in Hanover, where only environmental data was available.

Figure 5 displays the recommended routes from *Bahnhof Leinhausen* to *Leibniz Universität Hannover*, with one route suggested by Google Maps and the other selected by a local rider. The model scored the local's route higher, indicating a preference for it over the Google-recommended route. This is likely due to the fact that the local's route is flatter and features a protective strip for the bicycle lane, providing a safer and more comfortable riding experience.

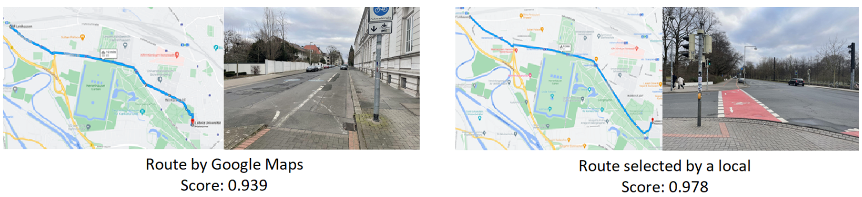

Fig. 5: Route overview, pictures, and scores from Train Station Leinhausen to the Leibniz University of Hanover. Left: route recommended by Google Maps. Right: route selected by a local.

In conclusion, a routing service specifically designed for cyclists is essential to ensure their safety, protect the environment, and provide local knowledge. Our prototype biking route recommendation system is just the beginning, and we hope to develop it further to revolutionize the cycling experience. So why not hop on your bike and let our system guide you to the hidden gems of any new city?

References

[EQA21] Esmailoghli, M.; Quiané-Ruiz, J.-A.; Abedjan, Z.: COCOA: COrrelation COefficient-Aware Data Augmentation. In. EDBT, 2021.

[EQA22] Esmailoghli, M.; Quiané-Ruiz, J.; Abedjan, Z.: MATE: Multi-Attribute Table Extraction. Proc. VLDB Endow. 15/8, pp. 1684–1696, 2022, URL: `https://www.vldb.org/pvldb/vol15/p1684-esmailoghli.pdf`.

[Es19] Esmailoghli, M.; Redyuk, S.; Martinez, R.; Abedjan, Z.; Rabl, T.; Markl, V.: Explanation of Air Pollution Using External Data Sources. In: BTW. 2019.

[Me] Meyer, H. J.; Grunert, H.; Waizenegger, T.; Woltmann, L.; Hartmann, C.; Lehner, W.; Esmailoghli, M.; Redyuk, S.; Martinez, R.; Abedjan, Z., et al.: Particulate Matter Matters-The Data Science Challenge@ BTW 2019. Datenbank-Spektrum/, pp. 1–18.