November 2021

Naturalistic Routing Using Inverse Reinforcement Learning

Renaud Hartert

Matt Barnes

Matthew Abueg

Naturalistic Routing Using Inverse Reinforcement Learning

ABSTRACT

This disclosure describes techniques, referred to as naturalistic routing (NR), that improve the quality of routes found by map applications by learning from users' real-world navigation actions, accessed with user permission. The techniques leverage the principle that users, in the aggregate, tend to travel on optimal routes to reach their destinations. A machine learning model is trained using inverse reinforcement learning and provides routes that are optimal by the users' definition of optimality, as determined from a dataset of navigation actions.

KEYWORDS

- Naturalistic routing
- Navigation
- Inverse reinforcement learning
- Inverse shortest-path learning
- Route-finding
- Digital map
- Origin-destination pair
- Global positioning system (GPS)

BACKGROUND

Computation of routes in digital map applications is performed by assigning costs to the roads interconnecting an origin and a destination and finding a route that optimizes costs. Some examples of costs include transit time, bumpiness of road, tolls, distance, etc.

However, when looking for directions (without route-finding software), human users (cyclists, drivers, walkers, truckers, motorcyclists, etc.) trade off many different criteria, e.g., avoiding traffic, the availability of bicycle lanes (for bicyclists), avoiding hills (for bicyclists, pedestrians, and non-motorized road users), etc. It is a priori difficult to build a routing model that matches users' expectations, since the designers of a digital map application may not only be unaware of all the criteria but also of the relative importance assigned to the various criteria by different users. Manually building a routing model (as is done in route-finding software) not only requires a designer to properly account for all observable factors but also to account for factors that are not necessarily represented in a data model of the world (e.g., quality of cycling lanes, abundance of drivers, scenery, etc.). The problem is made harder by the fact that users themselves aren't necessarily fully aware of the factors that influence their preference for a route over another.

Inverse reinforcement learning (IRL), or inverse shortest-path learning, is a technique that enables a machine to learn a behavior by studying user demonstrations. Rather than minimize a given cost function, IRL searches for a cost (or reward) function whose minimum (maximum) is the observed human behavior. IRL is described in greater detail, for example, in [3], while [1] and [2] apply it to route-finding.

Existing IRL-based route-finding doesn't scale beyond the size of a typical city, whereas the route-finding customer can have an interest in finding a continent-spanning route. The behavior that existing IRL applications aim to reproduce tends to be simple, e.g., predict the behavior of a taxi driver based on about fifteen simple factors such as speed limit and number of lanes. There is strong dependency on the accuracy of the dataset of observed routes, e.g., sensitivity to noise.

Existing IRL-based route-finding maximizes the likelihood of routes over all possible alternatives or over a static subset of alternatives. Maximizing likelihood over all possible routes is difficult to scale to continent-wide road networks, and historically, has only been implemented on networks the size of a small city. Maximizing likelihood over a static subset of alternatives can improve scalability but introduces unbounded suboptimality in the learning process. For example, the quality of learning depends on the quality of the subset of generated alternatives to learn from. For instance, if the alternatives do not contain routes with tolls, then the algorithm is not able to learn whether tolls are to be avoided or not. This introduced suboptimality is a limitation when routing behavior depends on a large set of factors that are difficult to model.

Obtaining a dataset that includes location data and can be used for learning is difficult. Most current route-finding techniques do not have the capacity to learn routing models in a fully privacy-preserving manner.

DESCRIPTION

This disclosure describes techniques, referred to as naturalistic routing (NR), that improve the quality of routes found by digital map applications based on user-permitted data on how users navigate in the real world. The techniques leverage the principle that users, in the aggregate, tend to travel on optimal routes to reach their destinations; indeed, the very selection of a certain route by a human user over competing routes is an indication of its optimality. A machine learning model that reproduces the aggregate behavior of users can thus provide routes that are optimal by the users' definition of optimality.

NR is a technique that determines a cost function that is such that its optimum corresponds to the actual tracks selected by users that have provided permission to access such data. It is based on the following components:

- A dataset of tracks (e.g., sequences of locations obtained by GPS, WiFi triangulation, etc.), obtained with user permission, that demonstrate real-world paths selected by users.

- A routing engine which generates routes that NR optimizes over.
- The IRL module which tunes the routing engine using the dataset of tracks.

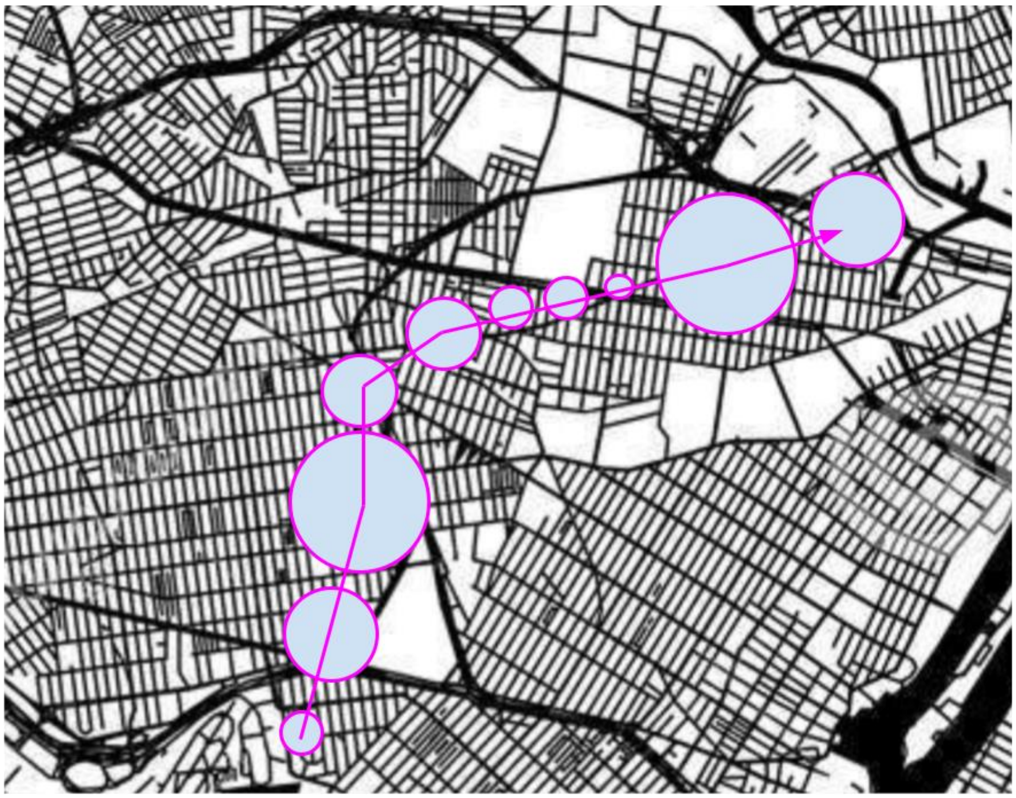

Fig. 1: An example user track from an origin (bottom left) to a destination (top right). The blue circles indicate ambiguity in path information: larger diameter circles are more ambiguous and vice-versa

Fig. 1 illustrates an example of a user track from an origin (bottom left) to a destination (top right). As indicated by the blue circles, the precision of the location records can change with time, and there is some uncertainty in the knowledge of the route the user actually traveled on.

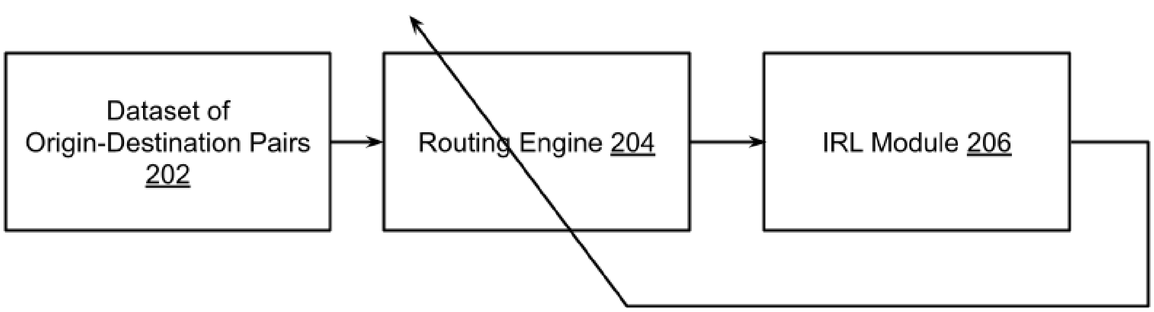

Fig. 2: Naturalistic routing

Fig. 2 illustrates naturalistic routing. For each user track in a dataset of origin-destination pairs (202), a routing engine (204) generates multiple alternative routes to travel from the origin to the destination of the track. Each route can include roads, sidewalks, bicycle paths, public transit, etc., each with its own attributes, e.g., road width, speed, travel time, speed limit, weather, bumpiness, etc. Each alternative route is associated with a score, referred to as a mirror score, that represents how well the alternative matches the actually selected user track: the higher the score, the better the match.

The goal of the IRL module (206) is to select a route from the routing engine which maximizes the mirror score. Optionally, the learned costs from the IRL module may then be used to tune the routing engine, which generates new improved alternates, and the entire procedure can be repeated. The generation of new alternatives from an increasingly tuned routing engine is referred to as outer-loop optimization. The entire optimization procedure may be done with differential privacy, ensuring that no user-track information leaks to the tuned routing engine, thereby preserving privacy. The reward function that the IRL converges to, whose optimum matches the observed paths actually selected by human users, can be used to answer route-finding queries posed by users.

Precision of routes and tracks

In contrast to existing techniques that maximize the likelihood of reproducing the observed, non-ambiguous routes, the described techniques maximize the expected mirror score of the alternatives. While existing techniques are predicated on the existence of expensive, high-quality datasets of routes to learn from, the described techniques decouple the learning process from the tracks used to score the alternatives and can handle ambiguous tracks that match different alternatives. Thus, the described techniques thus enable learning to occur from data sources with lower precision or quality of routes. In this manner, the described techniques can leverage relatively low-quality user tracks (e.g., as obtained from phones carried by users that permit use of track data) to build a dataset to learn from in a scalable and cost-effective way applicable to diverse regions of the world.

Learning

In contrast to existing techniques that maximize the likelihood of routes over all possible alternatives or a static subset of alternatives, the described techniques tune the routing engine against a small subset of alternatives and counterbalance the introduced suboptimality by continuously including (via the outer loop) new alternatives to re-optimize the routing engine. In this manner, the described techniques benefit from the scalability of tuning first against a smaller subset of alternatives, then account for factors unaccounted for in previous iterations. The described techniques have the advantages of both small subsets of routes as well as global optimization over all possible routes.

Naturalistic routing, as described herein, uses differential privacy and federated learning techniques to learn a routing model without disclosure of user-track (or other user) information. The use of federated learning allows learning while location data stays on the user device.

In this manner, the techniques of naturalistic routing described herein improve the quality of directions returned by map applications by learning from actual user demonstrations in a scalable, cost-effective, and privacy-preserving manner.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user’s routes through the physical world, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques, referred to as naturalistic routing (NR), that improve the quality of routes found by map applications by learning from users' real-world navigation actions, accessed with user permission. The techniques leverage the principle that users, in the aggregate, tend to travel on optimal routes to reach their destinations. A machine learning model is trained using inverse reinforcement learning and provides routes that are optimal by the users' definition of optimality, as determined from a dataset of navigation actions.

REFERENCES

[1] Abbeel, Pieter, and Andrew Y. Ng. "Apprenticeship learning via inverse reinforcement learning." In *Proceedings of the Twenty-first International Conference on Machine Learning*, p. 1. 2004.

[2] Ziebart, Brian D., Andrew L. Maas, J. Andrew Bagnell, and Anind K. Dey. "Maximum entropy inverse reinforcement learning." In *Aaai*, vol. 8, pp. 1433-1438. 2008.

[3] Snoswell, Aaron J., Surya PN Singh, and Nan Ye. "Revisiting Maximum Entropy Inverse Reinforcement Learning: New Perspectives and Algorithms." *arXiv preprint arXiv:2012.00889* (2020).