

Crossing intersections: A tool for investigating road user pathways

EPB: Urban Analytics and City Science
2023, Vol. 0(0) 1–7
© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23998083231215462

journals.sagepub.com/home/epb



Heather Anne Kath 

Bergische Universität Wuppertal, Germany

Abstract

The pathways used by cyclists, pedestrians, and users of micromobility to cross intersections do not always align with those planned by traffic engineers. Observing actual usage patterns could lead to a better understanding of the tactical behavior of users of active and micromobility, allowing planners and engineers to create urban environments specifically for these road users. An open-source Python tool is introduced that uses clustering to automatically identify the forms of pathways used by road users. The tool was used to cluster trajectories from five intersections in Germany. The exemplar of each cluster is selected to represent the average shape of each pathway type. The open-source Python tool *RoadUserPathways* is introduced, the case studies are examined and use cases are presented.

Keywords

Infrastructure design, urban space, desire lines, pathways, cyclists, clustering trajectories

Introduction

The concept of *desire lines* was coined by Gaston Bachelard in *The Poetics of Space* (1958). Desire lines are paths worn by pedestrians on unpaved surfaces. These worn-out pathways highlight mismatches between the planned (and paved) walking network and the preferred routes of those using the space. Although the situation is slightly different, Copenhagenize Design Co. pioneered the use of desire lines in the context of cycling and has implemented their manual, observational data collection approach to investigate the behavior of cyclists crossing intersections in many cities (Colville-Andersen et al., 2013). The movement patterns for cyclists crossing and turning at intersections are stipulated by traffic engineers and correspond to the paved footpaths in the pedestrian case. For example, cyclists in Germany should use the indirect style for left turns at intersections with bicycle lanes (see Figure 1, top left), are expected to use cycling infrastructure if provided, and should cycle in the given direction of travel. An example of the expected and stipulated movement

Corresponding author:

Heather Anne Kath, School of Architecture and Civil Engineering, Bergische Universität Wuppertal, Pauluskirchstraße 7, Wuppertal 42285, Germany.

Email: kaths@uni-wuppertal.de

Data Availability Statement included at the end of the article

patterns of cyclists crossing and turning at an intersection with on-road bicycle lanes in Germany is shown in Figure 1, top right. The regulations for infrastructure use differ based on the intersection design and country.

The actual movement patterns of cyclists at intersections is much more dispersed. In Figure 1 (bottom), the trajectories of cyclists crossing two intersections in Aachen, Germany from the inD dataset (Bock et al., 2020) are shown. Cyclists use unexpected parts of the infrastructure (e.g. sidewalk) and cross using pathways that differ from planned movement patterns. Akin to pedestrians treading new pathways, these actual movement patterns of cyclists can point to weaknesses in the intersection design and can be a source of inspiration for new planning or engineering measures.

Researchers studying the tactical pathfinding behavior of cyclists at intersections have identified the following trends:

- Cyclists in Copenhagen use the infrastructure as intended most of the time. Only a handful of cyclists carry out direct left turns rather than an indirect left turn as intended by planners (Colville-Andersen et al., 2013).

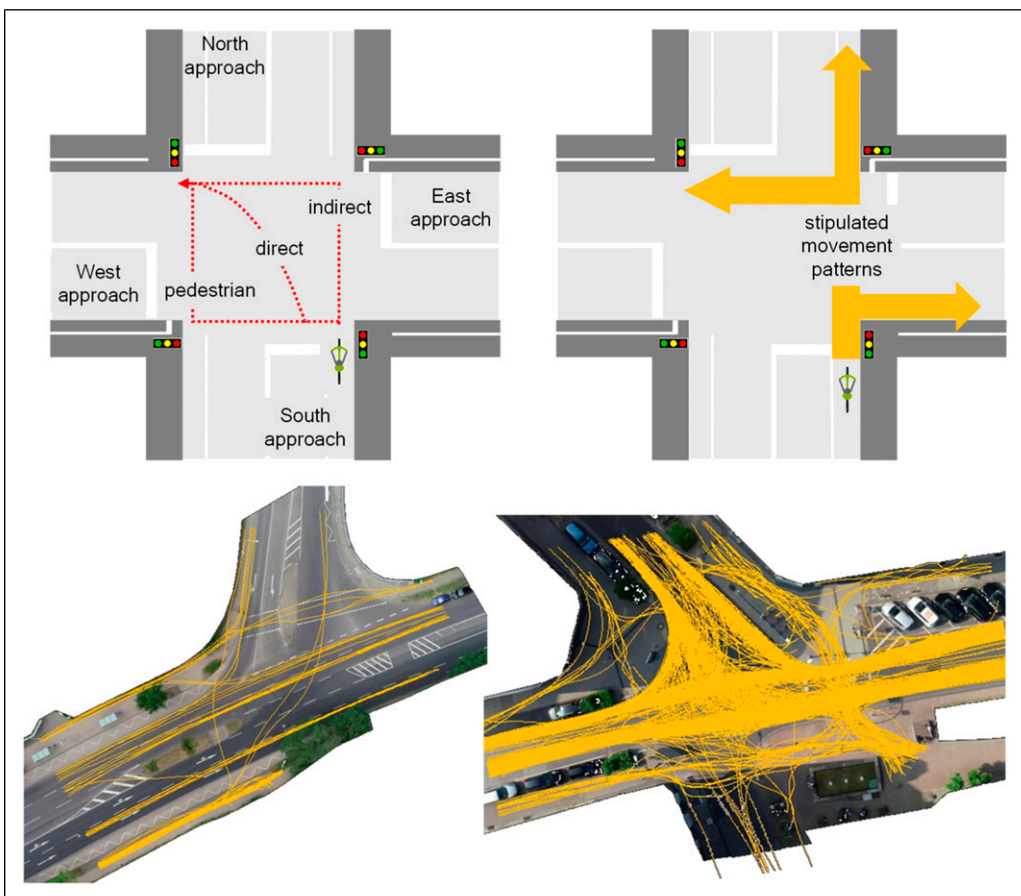


Figure 1. Commonly observed types of left turns (top left), examples of stipulated movement patterns for cyclists (top right) and trajectories from cyclists crossing two intersections from the inD dataset (bottom).

- The left-turn behavior of cyclists in Munich was found to be more diverse than in Copenhagen, with many cyclists using the pedestrian style left turn in addition to direct and indirect turns (Twaddle and Busch, 2019).
- A study of the desire lines of cyclists in Amsterdam showed that many different pathways are used by cyclists, particularly when turning left (University of Amsterdam and Copenhagenize Design Co., 2014).
- In Montreal, the number of typical pathways depends on the geometry of the intersection (Nabavi Niaki et al., 2019).
- In Barcelona, a desire line analysis showed that 78.9% of cyclists abide by traffic rules at intersections and the main reason for unexpected behaviour is avoiding stops (Lind et al., 2021).

Most of these findings were drawn from manually collected trajectory data. Manual data collection allows for the qualitative assessment of the situation by the observer. However, it greatly limits the number of trajectories that can be collected and processed. The contribution of this work lies in the development of an open-source Python tool for the automated identification of pathway types from raw trajectory data. Although it is possible for an analyst to manually cluster trajectories in cases with few observations or with dense cluster structures (e.g. intersection on the bottom left in Figure 1), this task is much more difficult in situations with many observations or dispersed infrastructure use (e.g. intersection on the bottom right in Figure 1). The automated identification of pathway types will enable researchers and practitioners to easily identify unexpected behaviors (desire lines) that differ from the intended movement patterns at particular intersections. This can be done time efficiently for large trajectory datasets and many intersections.

Methodology

Data

The input data are trajectories comprised of position coordinates. Trajectories saved in SQLite databases with the structure defined in the open-source project *Traffic Intelligence* (Jackson et al., 2013) are loaded. *RoadUserPathways* was developed using trajectories with 25 position coordinate observations per second. Trajectories with a lower frequency of position coordinates will work as long as the shape of the crossing maneuver is adequately captured. A local coordinate system (UTM) is used, in which the coordinate point (0,0) is the upper left-hand corner of the video frame from which the trajectories were extracted. The actual UTM coordinates of the given UTM zone can also be used.

Trajectory data is preprocessed to select complete trajectories and enable comparability. This is accomplished in *RoadUserPathways* using polygons drawn by the analyst to filter and trim the trajectory data. In addition, a method for analyzing the trajectories from each approach separately is included. More information about these functions can be found in the README file on GitHub.

Clustering approach

A trajectory from each crossing road user i is transformed into a one-dimensional feature vector F_i^* . The number of features in F_i^* must be uniform across the set of observations. Because trajectories range in length depending on the duration of the crossing maneuver, N position coordinates are sampled from each trajectory. Each trajectory is divided into segments of equal length and the dividing position coordinates form the feature vector. This sampling based on spatial progression eliminates the effect of stops and speed variation. No other features from the trajectory are included in the feature vector. The feature vectors from all observations are combined to form a pattern matrix \mathcal{A}^* . Euclidean is selected as a distance measure and \mathcal{A}^* is not normalized.

The expected clusters vary with regard to the number of observations in each cluster (cluster size) and the degree of similarity between observations in the same cluster (cluster density). Additionally, the number of different pathways across the intersection is not known in advance. The data are expected to be globular, or convex, meaning that any line drawn in d dimensional space between an observation and the cluster center is within the boundary of that cluster. Based on the attributes of the clustering problem and the availability of an openly available Python implementation (scikit-learn), the Affinity Propagation (AP) (Frey and Dueck, 2007) clustering algorithm is selected.

AP is a centroid-based clustering method, meaning that the clusters are built by measuring the distance of observation points from an assigned cluster centroid (referred to as an exemplar in AP). The algorithm rotates through trajectories as potential centroids and selects the set of centroid trajectories that results in the lowest overall distance between observation points and assigned cluster centroid. The use of a centroid-based method is important for *RoadUserPathways* because the centroid (or exemplar) represents the desire line of each pathway type.

Open-source tool

The Python code for loading and preprocessing raw trajectory data, clustering trajectories, and preparing the result is made available through the open-source Python project *RoadUserPathways*, which is hosted on GitHub. Detailed information about the dependencies, workflow, data input and preprocessing, parameters, output and clustering with AP are included in the README file.

The Python scikit-learn implementation of AP (Pedregosa et al., 2011) is used to cluster the trajectory data in the \mathcal{A}^* matrix. The \mathcal{A}^* matrix format described above is the basic input for (nearly) all of the clustering methods included in *scikit-learn*. Therefore, although AP is implemented in *RoadUserPathways* and has been found to provide useful results, it is straightforward to switch to a different centroid-based clustering method included in *scikit-learn*, such as k-means (although with this method the number of clusters has to be specified by the analyst).

The tool yields three main outputs:

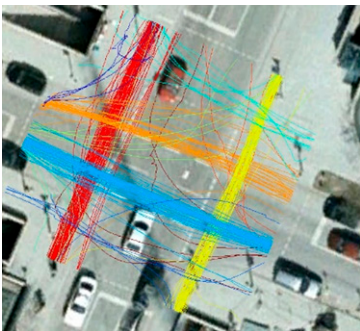

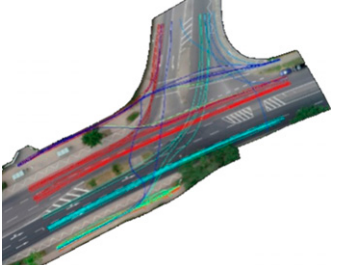
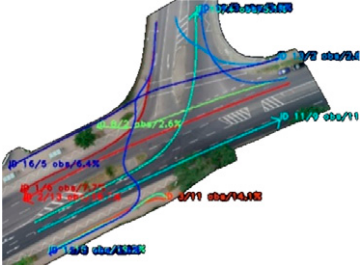
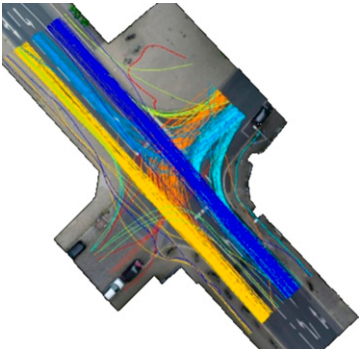
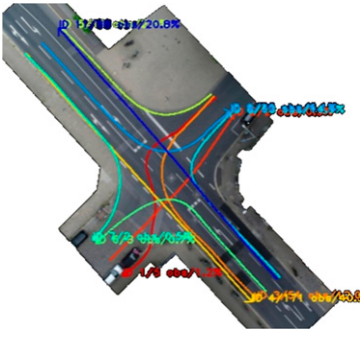
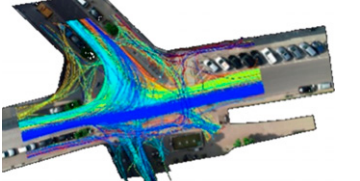
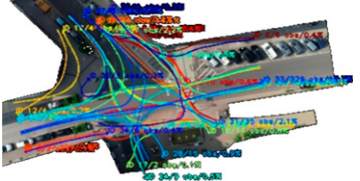


- The number of pathway types used by road users
- A representative trajectory that characterizes the average shape of each pathway type
- The number and percentage of road users who use each pathway type

Case studies

RoadUserPathways was tested using cyclists trajectories from one intersection in Munich (Twaddle, 2017) and four intersections in the inD dataset (Bock et al., 2020). The output from *RoadUserPathways* is shown in Table 1. The column “Clustered trajectories” shows all trajectories sorted by cluster (each color represents a cluster). Seeing all trajectories together is useful in analyzing the variation in behavior. Do all road users using a given pathway type follow similar trajectories? Or is there considerable variation in the form of pathways? Do trajectories cover much of the surface area (dispersed infrastructure use) or are they located in certain regions (concentrated infrastructure use)? The Munich intersection exhibits concentrated infrastructure use, with most cyclists using existing bicycle lanes on the north/south approaches. In contrast, cyclists at the inD intersection A use the roadway, sidewalk, and bicycle lane and various pathways to cross the intersection. An example of cyclists choosing different types of pathways to carry out the same maneuver is demonstrated by cyclists turning left from the north approach of the Munich intersection.

The images in the column ‘Representative trajectory’ show the exemplar trajectory, signifying the average shape of the pathways type. The direction of movement is indicated by an arrow on the exiting end of the trajectory. The number and percentage of cyclists using each pathway type are

Table I. Output of the pathway identification tool.

	Clustered trajectories	Representative trajectory	# of obs.	# of paths	\bar{s}
Munich			277	11	0.83
inD A			78	10	0.57
inD B			424	9	0.53
inD C			1666	37	0.32
inD D			39	5	0.64

displayed and output in a text file. This information is useful for examining the shape of average trajectories and how common unexpected pathway types are at a given intersection. The desire lines of cyclists using the inD intersection B match very closely with the planned movement patterns. In comparison, the desire lines at the inD intersection C show a wide divergence from the intended infrastructure use. The clustering results from this intersection could indicate that an infrastructure redesign is necessary. Given that many cyclists are using the sidewalks rather than the roadway, it may be beneficial to install bicycle lanes.

The clustering structures are evaluated using the average Silhouette Score, where $\bar{s} > 0.70$ indicates a strong clustering structure (Rousseeuw and Kaufman, 1990). The only case study intersection to surpass this threshold is the Munich intersection, while the inD intersection D nears it. The clusters results were also evaluated qualitatively to determine the occurrence of over-clustering, over-segmentation and otherwise erroneous clustering. Total errors ranged between 0.4 and 2.5%. This evaluation is subjective because the ‘correct clustering’ depends on the application and the opinion of the analyst. The preference parameter of the AP method can influence the resulting number of clusters, and analysts could adjust this value to rather err on the side of over-segmentation or over-clustering.

Discussion and conclusion

There are many possible use cases for the Python package *RoadUserPathways*:

1. Results provide planners and engineers with **count data** and **turning ratios**. The frequency of basic maneuvers (straight, left turn, right turn, U-turn) and the traffic volume can be derived from the output.
2. Clusters provide insight into the **tactics** employed to execute basic maneuvers. Such information can be valuable in (re)designing infrastructure that serves the needs of users of different modes.
3. **Before and after studies** to determine if a given measure influences tactical crossing behavior.
4. The **behavior of users of different modes** can be compared and contrasted.
5. The link between the pathway type and **traffic safety** could be explored using the resulting clustering structure and the surrogate safety indicators.

RoadUserPathways is a starting point for analyzing the movement patterns of many road users at different intersections. There are, however, some weaknesses and points for future development. First, a qualitative comparison of the clusters obtained using AP, k-means and Mean Shift indicated that the best results were obtained with AP. In addition to the good results, a strong advantage of this algorithm is that the number of clusters is determined as an output of the algorithm. However, a more systematic assessment is necessary. Second, the trajectories observed do not reflect the pure desired behavior. Road users are also reacting to elements in the environment; other road users, traffic lights, obstacles, etc. Analysts should be aware of this while drawing conclusions. Although automated data processing methods allow researchers to study a large datasets, trajectories remain a sterile form of data that do not include contextual information. Other methods of quantitative and qualitative data collection can lead to a more comprehensive understanding of behavior.

Acknowledgements

Professor Nicolas Saunier (<https://www.polymtl.ca/expertises/en/saunier-nicolas>) supported this work through his supervision of my doctoral thesis and through his technical assistance with the open-source tool for extracting road user trajectories from video data Traffic Intelligence.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the German Federal Ministry for Digital and Transport.

ORCID iD

Heather Anne Kaths  <https://orcid.org/0000-0003-2554-8243>

Data Availability Statement

All [data](#) generated or analysed during this study are included in this published article. [Bock et al. \(2020\)](#)

References

- Bachelard G (1958) *The Poetics of Space (La Poétique de l'Espace)*. Presses Universitaires de France.
- Bock J, Krajewski R, Moers T, et al. (2020) The ind dataset: a drone dataset of naturalistic road user trajectories at German intersections. *2020 IEEE Intelligent Vehicles Symposium 4*: 1929–1934.
- Colville-Andersen M, Madruga P, Kujanpää R, et al. (2013) *The Bicycle Choreography of an Urban Intersection. Desire Lines & Behaviour of Copenhagen Bicycle Users*. Denmark: Copenhagenize Design Co. Frederiksberg.
- Frey BJ and Dueck D (2007) Clustering by passing messages between data points. *American Association for the Advancement of Science* 315(5814): 972–976.
- Jackson S, Miranda-Moreno L, St-Aubin P, et al. (2013) Flexible, mobile video camera system and open source video analysis software for road safety and behavioral analysis. *Transportation Research Record: Journal of the Transportation Research Board* 2365: 90–98. DOI: [10.3141/2365-12](https://doi.org/10.3141/2365-12)
- Lind A, Honey-Rosés J and Corbera E (2021) Rule compliance and desire lines in Barcelona's cycling network. *Transportation letters* 13(10): 728–737.
- Nabavi Niaki MS, Saunier N and Miranda-Moreno LF (2019) Is that move safe? Case study of cyclist movements at intersections with cycling discontinuities. *Accident Analysis & Prevention* 131: 239–247. DOI: [10.1016/j.aap.2019.07.006](https://doi.org/10.1016/j.aap.2019.07.006)
- Pedregosa F, Varoquaux G, Gramfort A, et al. (2011) Scikit-learn: machine learning in Python. *The Journal of Machine Learning Research* 12: 2825–2830.
- Rousseeuw PJ and Kaufman L (1990) *Finding Groups in Data*. Wiley Online Library.
- Twaddle HA (2017) *Development of Tactical and Operational Behaviour Models for Bicyclists Based on Automated Video Data Analysis*. Technische Universität München.
- Twaddle H and Busch F (2019) Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour* 60: 47–57. DOI: [10.1016/j.trf.2018.10.002](https://doi.org/10.1016/j.trf.2018.10.002).
- University of Amsterdam and Copenhagenize Design Co. (2014) *The Desire Lines of Cyclists in Amsterdam*. Amsterdam: University of Amsterdam.

Heather Kaths is a professor at the University of Wuppertal. She leads the Department of Bicycle Traffic within the Center for Mobility and Transport. She holds a B.Sc. in Civil Engineering with a focus on Transportation Engineering from the University of Calgary (Canada), a M.Sc. in Transportation Systems, and a doctorate from the Technical University of Munich (Germany).