Predicting Perceived Level of Cycling Safety for Cycling Trips

Jiahui Wu
College of Information Studies
University of Maryland
College Park, Maryland, USA
jeffwu@umd.edu

Lingzi Hong
College of Information Studies
University of Maryland
College Park, Maryland, USA
lzhong@umd.edu

Vanessa Frias-Martinez
College of Information Studies
UMIACS
University of Maryland
College Park, Maryland, USA
vfrias@umd.edu

ABSTRACT

Cycling provides various benefits to cyclists and cities. Nevertheless, the growth of cycling is still hindered by the lack of citywide information about perceived cycling safety. Providing cyclists with information about the safest routes could help increase cycling activity. In this paper, we aim to predict the perceived level of cycling safety for a trip (trip-PLOCS). We utilize LSTM-based architectures to incorporate the sequential information of segments in a trip, and predict its cycling safety. Our proposed method can achieve up to 76% F1 micro (65% F1 macro) score, 10% (19%) better than the state-of-the-art baseline. Finally, we use SHAP to extract insights about trip-PLOCS, showing that social features contribute to perceived danger while cycling facilities contributes to the perceived safety.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Computing methodologies → Spatial and physical reasoning**; **Supervised learning**.

KEYWORDS

cycling safety prediction, spatio-temporal analysis, open data

ACM Reference Format:
Jiahui Wu, Lingzi Hong, and Vanessa Frias-Martinez. 2019. Predicting Perceived Level of Cycling Safety for Cycling Trips. In *27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19), November 5–8, 2019, Chicago, IL, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3347146.3359092

1 INTRODUCTION

A large body of literature demonstrates the benefits of cycling: reduction in air pollution [7]; savings in health care costs [1], among others. However, with more cyclists on the street, cycling safety becomes an increasingly important concern. One of the main obstacles to encourage cycling is the lack of information regarding perceived cycling safety [2]. Using surveys and mental maps, studies show that perceived cycling safety is related with various factors, such as built environment characteristics [6]. Build upon these insights, a recent study use machine learning models to predict perceived level of cycling safety (PLOCS) for each segment and create a perceived cycling safety map at the city level [16]. However, that study treats segments independently and ignores the sequential nature of a trip, and in consequence, the role that previous safety perceptions might play in the perception of segments visited in the future. In this paper, we aim to develop machine learning models that allow us to predict the PLOCS of a cycling trip (trip-PLOCS) with the objective of helping cyclists identify and select a route that is adequate to their level of comfort. Ultimately, we envision the cycling trips-PLOCS to be integrated into cycling route planning applications that would offer route and cycling safety information at the same time. We propose to characterize street segments in a similar way as [16], *i.e.,* a set of built environment and social features, and PLOCS labels crowdsourced from cyclists; and to capitalize on the sequential information of the segment trips, we use neural network models with long short-term memory (LSTM) as the core component.

Beyond building a good predictive model for trip-PLOCS, another goal of this paper is to understand the important features that determine predictions. Insights about important features to cycling safety can help urban planners build better environment to increase bicycle use. Here we use a method called SHAP (SHapley Additive exPlanations) [8] to understand features' importance in predictions and the contribution direction to predicting PLOCS labels.

The key contributions are: **(1)** We design LSTM-based models to predict trip-PLOCS and compare against the state-of-the-art baseline. **(2)** We extract insights about cycling safety by interpreting the contribution of features to PLOCS prediction using SHAP values.

2 RELATED WORK

Perceived cycling safety influences cyclists' decisions to ride bikes on the streets [2]. Most studies about perceived cycling safety focus on the determinants. Cycling facilities, collisions involving cyclists and the bicycle road network have been shown to influence cycling safety perception [5, 12, 14]. Besides determinants, a recent study try to provide citywide cycling safety information on the segment level to cyclists using machine learning models [16].

3 DEFINITIONS AND PROBLEM STATEMENT

Here we introduce related definitions and the problem statement.

Segment is a part of a street without any intersection inside, e.g., *T street NW* is divided into segments including S_2, S_3, and S_4.

Segment Features. Segments are represented by social features, such as crash accidents, and built-environment features, such as bike lanes, extracted from open and crowdsourced data. We use the complete list of features in [16]. These features will be useful to interpret how machine learning models make PLOCS predictions.

Cycling trip is the route the cyclist travels on a bike. It is represented as a list of segments after the GPS trace is mapped to a given road network using map matching algorithms [9].

SIGSPATIAL '19, November 5–8, 2019, Chicago, IL, USA

ACM ISBN 978-1-4503-6909-1/19/11.
https://doi.org/10.1145/3347146.3359092

Figure 1: An example Trip T_i near *15th* and *T street northwest* in Washington D.C.

Perceived Level of Cycling Safety (PLOCS) is the safety perception in the context of cycling. After watching a short cycling video, cyclists give their subjective safety perception on a 5-level scale, from 1 to 5 representing very dangerous, dangerous, normal, safe and very safe. Ratings are collected from local cyclists living in the same city where the videos are recorded, so that they are somewhat familiar with the social characteristics of the city.

Seg-PLOCS (SP_{S_i}) is the average of PLOCS ratings for S_i. The average is transformed into one of the five PLOCS labels by dividing the range from 1 the 5 evenly into five bins.

Trip-PLOCS (TP_i) is the average of the seg-PLOCS of all segments in that trip. The trip-PLOCS average is also transformed into one of the five labels in the same way as the segment average.

Problem Statement. Given a cycling trip T_i and the features for segments in T_i, we aim to classify the trip's PLOCS, TP_i, as a label among very dangerous, dangerous, normal, safe and very safe.

4 METHODOLOGY

In this section, we will explain three models for the trip-PLOCS prediction problem. Then we will introduce our proposed training data expansion approach, and describe how we use SHAP values to interpret important features for trip-PLOCS prediction.

4.1 Models

Baseline: The baseline setting is adapted from [16]. We predict a trip-PLOCS by first predicting all the seg-PLOCS in the trip independently, averaging all the predicted seg-PLOCS for that trip, and assigning the closest category. XGB is chosen as the machine learning method since it is the most accurate method for seg-PLOCS prediction [16]. However, this approach ignores the sequential nature of a trip. There are two reasons to consider and model the sequential information in a trip. First, the safety perception of the current segment might be temporally correlated with the perception of previous segments. Second, spatial auto-correlation is a common characteristic of geographical data, such as crimes [15] and road infrastructure [4]. The sequential information contains the spatial proximity of segments, i.e., segments next to each other in a cycling trip are adjacent. The next two approaches are an attempt to incorporate information for spatial and temporal correlations.

Sequence Labeling: The task of sequence labeling is to assign a label to each observation in the input sequence [11]. A commonly used family of models for this task is Long Short Term Memory (LSTM). Because the labeling is executed at each time step, a unidirectional LSTM makes predictions based on only information from previous time steps. To avoid this limitation, a bidirectional LSTM layer (BiLSTM) is often used to capture information from the whole sequence at each labeling step [3]. For our prediction problem, (Bi)LSTM model will iterate over Trip T with N segments and predict SP_{S_i} for each segment. With an LSTM layer, predictions are conditioned on the features of S_i and the sequential context of previous segments $\{S_1, ..., S_{i-1}\}$. With a BiLSTM layer, the sequential context of subsequent segments $\{S_{i+1}, ..., S_N\}$ is also considered.

Sequence Classification As our goal is to predict the PLOCS for a trip, we can skip predicting seg-PLOCS for each segment and instead optimize the model towards the final trip-PLOCS. In this case, we can frame it as a sequence classification problem: given a sequence of segments and their feature representations, predict the trip-PLOCS. The downside of this approach is that models in this setting do not have full access to the detailed information of the seg-PLOCS, since the training labels of trip-PLOCS is the average of the seg-PLOCS. Nevertheless, this setting might be useful where labels are not available for each segment in the road network.

4.2 Training Data Expansion

Collecting sufficient training labelled data is always a challenge for supervised machine learning models. It is even more challenging to collect PLOCS labels at the segment level. Here we propose to expand training data with plausible trips using sliding window with one parameter, the window size w. The length of generated trips, i.e., the number of segments covered in a trip, equals to w. Multiple window sizes can be applied to the training data to generate trips at various lengths. The sliding window approach with window size w works as follows. For a cycling trip T_i=S_1, .., S_N with N segments, we apply a window with size w to generate N-w+1 sub-trajectories (e.g. T_{i_j}=$S_j, ..., S_{j+w-1}$) to be added to the training dataset.

4.3 Important Features for Trip-PLOCS

Beyond a good predictive model for trip-PLOCS we also aim to understand the types of features that are determinant of safety perception. This information would also be highly relevant for urban planners willing to increase bicycle use and other micro-mobility services. There exist several tools to interpret deep learning and machine learning models. Scoot and Lee proposed SHAP (SHapley Additive exPlanations) values to explain the role that each feature plays in each prediction [8]. SHAP is proposed based on a sound theoretical framework and is shown to be more consistent with human's intuition than other approaches [13]. SHAP values are computed by an explanation model independent of the predictive model to be explained. This explanation model first computes the average model output over a background dataset, e.g., the training dataset for the predictive model, as the base value. Then it attributes to each feature the change in base value brought by adding that feature to the predictive model. This change is the SHAP value for that feature. In multi-class prediction setting, a SHAP value is computed for each input feature for each class. By aggregating SHAP values, we can understand the features' importance on and

Problem Setting	Model	Training Data Expansion	F1 Score Micro	Macro
Baseline	XGB	none	0.66	0.46
Sequence Labeling	LSTM	sw01-13	0.73	0.56
	BiLSTM	sw08-13	0.74	0.61
Sequence Classification	LSTM	sw01-13	0.76	0.65
	BiLSTM	sw01-13	0.74	0.62

Table 1: Experiment results.

contribution to each trip-PLOCS label. In this paper, we will use SHAP values to provide information about the role that social and built environment features might play in cycling safety perception.

5 EXPERIMENTS

5.1 Experiment Settings

5.1.1 Dataset. We use a dataset that contains 82 trips covering 453 segments in Washington D.C. Each street segment in the dataset was rated by multiple local cyclists that watched videos and provided their PLOCS. The dataset contains 2295 segment ratings from 235 local cyclists [16]. The dataset is relatively small compared to others described in the deep learning literature. This is mostly due to the cost of recording the cycling videos and recruiting local cyclists to provide the ratings. However, the training data expansion explained in section 4.2 will enlarge the dataset to up to ~1400 trips. Each segment is represented as a 231-dimension feature vector, of which 63 are built environment features and 148 are social features. On average, these trips have 6 segments and a duration of 188 seconds.

In our experiment, we randomly sample 80% of our dataset as the training data and the rest 20% as the testing data. To have robust results, we repeat the random sampling ten times and report average results across runs.

5.1.2 Evaluation metrics. The F1 score is widely used in binary classification. There are different ways to generalize F1 score to the multi-class setting. Here we use both F1-micro and F1-macro to evaluate our models. F1 micro score does not emphasize on rare classes, e.g., the performance in rare classes would have less impact on the F1 micro score than common classes with larger population. On the other hand, F1 macro score treats the F1 scores of each class equally and is popular in class imbalanced settings [10]. Therefore, the difference between F1-micro and F1-macro indicates the performance of the imbalanced dataset [16]. The smaller the difference is, the more similar performance in both rare and common classes, and the better the model handles the imbalanced dataset.

5.1.3 Data expansion setting. Since the maximum number of segments per trip in our dataset is 14, we set the largest window size of the sliding window to 13 and the smallest to 1. In principle, any combination of different window sizes can be applied to expand the training data. In our experiment, we fix the maximum window size as 13 and then gradually decrease minimum window size we apply to the training data. We use sww_{min}-w_{max} to denote the dataset expanded by window sizes w_{min}, w_{min+1}, .., w_{max}.

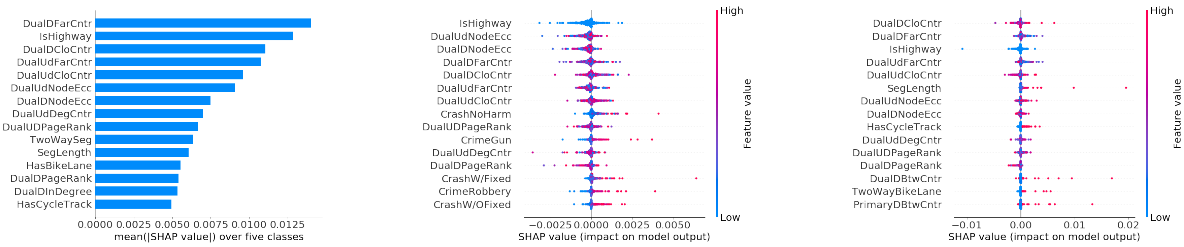
5.2 Experiment Results

In this section, we compare the performance of the three problem settings described above: baseline, sequence labeling and classification to predict the PLOCS of a trip. We report the average F1 micro and macro score across ten runs with different random train/testing splits. Table 1 shows the best results achieved by different models and data expansions. We present the following observations. **(1)** We confirm our hypothesis that incorporating sequential information in a trip improves the trip-PLOCS prediction. Models in both sequence labeling and classification have a substantial improvement over the baseline. The best performance is achieved by the LSTM model in the sequence classification setting with the largest training data set expanded. Although the F1-macro is still smaller than F1-micro, i.e., the model performs better in common labels with large number of trips, e.g., *normal* PLOCS, than rare labels with small number of trips, e.g., *very dangerous* PLOCS, the difference between the them reduces from 0.20 (0.66-0.46) in the baseline to 0.11 (0.76-0.65) in the best model. **(2)** Models in sequence classification setting generally outperform the ones in the sequence labeling setting with the help of data expansion. As mentioned in section 4, the advantage of sequence labeling is that it fully utilizes all the labels from all segments in the cycling trips, while the information of seg-PLOCS in the classification setting is averaged. Nevertheless, with the help of the data expansion, the models in the classification setting have access to more segment-PLOCS information in the training data and demonstrate their advantage in optimizing towards the final trip-PLOCS (10% better than baseline XGB, 2% better than sequence labeling setting). All models in the classification setting achieve the best results with the smallest window sizes (sw01-13). This also suggests that training models to learn hidden features with shorter trips does not hurt the models' ability to capture long sequential information and help them make better predictions at trips with various lengths. **(3)** Bidirectional LSTM performs better than unidirectional LSTM in the sequence labeling setting, because BiLSTM provides information of the subsequent segments at the time of making prediction for a segment. For the classification setting, bidirectional LSTM perform worse than unidirectional, probably because the lengths of trips are quite small and the models make predictions after observing complete sequences.

5.3 Important Features in Trip-PLOCS

To understand the most relevant features behind trip-PLOCS predictions, we take the best model in Table 1, i.e., LSTM trained with dataset sw01-13 in the sequence classification setting, and compute the SHAP values in predicting the testing data. SHAP value is a local explanation of each input feature for a prediction. To understand the contribution of each built-environment and social feature to the trip-PLOCS, we need to aggregate these SHAP values.

First, we compute the overall contribution of each feature across trip-PLOCS classes. This is computed as the average absolute SHAP values across classes and segments. The top 15 most important features are shown in Figure 2(a). All 15 are built-in environment features, which is in line with previous work on seg-PLOCS prediction, where built-environment features alone achieve better performance than social features alone[16]. Most of the important built environment features are topological (names start with "Dual"). Not being a highway segment is the second most important feature. Whether the segment is one- or two-way street and whether the segment has bike lanes or tracks are also very important. The importance of bike facilities to perceived cycling safety is well studied [12]. But whether a segment is one- or two-way street is seldom



(a) Overall contribution. (b) SHAP for trip-PLOCS of "very dangerous" . (c) SHAP for trip-PLOCS of "very safe".

Figure 2: Contribution of Top 15 features to trip-PLOCS prediction. (i) Features starting with "Dual" or "Primary". Features starting with "Primary" means it is a metric computed in a primary graph, where segments are the edges and intersections are the nodes, and the one starting with "Dual" are graph-based metrics computed in dual map, where the segments are the nodes and intersections are the edges. Following "Dual" or "Primary", "D" means the network is directed and "UD" means undirected. The rest of the name is the metric's name. (ii) Features starting with Crime, Crash and 311 are different types of features extracted from crimes, crashes and 311 requests datasets.

mentioned in the literature about the determinants of perceived cycling safety. Although being important in predictive model does not entail causality, the relationship between perceived cycling safety and this feature is worth further examination.

Next, we examine the contributions of each feature, i.e., the direction of contribution to each class of trip-PLOCS (e.g., Figures 2(b) and 2(c), SHAP for other 3 PLOCS levels are not reported due to space limit). A dot in the plot represents that feature for one segment. It is colored based on the feature value for that segment, pink for high and blue for low values. The x-axis represents the SHAP value. For a feature, if all the pink dots have positive values and all the blue dots have negative value, it means high values in that feature consistently favor that class, and low values consistently push the prediction away from that class. There is a clear distinction between important features for predicting *very dangerous* and *very safe*. **(1)** Social features are very important in predicting *very dangerous* trip-PLOCS. Large values in crimes (with guns or robberies) and crashes (whether crashes involve fixed objects, produce harms to other objects) determine that the trips are probably very dangerous. Although we did not provide any social features to cyclists when they watched and rated the cycling videos, local cyclists still picked up social features of the neighbourhood and gave proper safety rating to those segments. This observation reinforces that cities should start their local campaigns to collect ratings from local cyclists. **(2)** In contrast, immediate built-in environment features perceived by the cyclists directly, such as bike lanes, are important in predicting safe routes. Cycling trips involving segments with bike lanes or cycle tracks tend to be very safe.

6 CONCLUSION

In this paper, we use LSTM-based models to predict trip-PLOCS in two problem settings, namely sequence labeling and classification. To deal with the potentially small labeled datasets, we propose a sliding window approach to expand the training data. By modeling the sequential information in cycling trips, models in both problem settings perform substantially better than the baseline setting that treats segments independently (10% improvement in F1 micro score and 19% in F1 macro score). We also use SHAP values to understand the contribution of each feature to the trip-PLOCS prediction. Overall, built-environment features are most important, such as segment type and graph-based centrality features. For each class of trip-PLOCS, the contributions of these features vary. Notably, social features such as crashes and crimes are very important for determining perceived dangers and bike lanes and tracks are indeed useful to build a safe cycling environment.

Acknowledgements: This work was funded by the National Science Foundation under Grant NSF-1636915.

REFERENCES

[1] David R Bassett, John Pucher Jr, Ralph Buehler, Dixie L Thompson, and Scott E Crouter. 2008. Walking, cycling, and obesity rates in Europe, North America, and Australia. *Journal of Physical Activity and Health* 5, 6 (2008), 795–814.

[2] Kaplan S. Nielsen T.A.S. Chataway, E.S. and C.G. Prato. 2014. Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between Brisbane and Copenhagen. *Transportation Research Part F* 23 (2014), 32–43.

[3] Jason PC Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the ACL* 4 (2016), 357–370.

[4] BenoıĹČt Flahaut. 2004. Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. *Accident Analysis & Prevention* 36, 6 (2004), 1055–1066.

[5] Michael Klobucar and Jon Fricker. 2007. Network evaluation tool to improve real and perceived bicycle safety. *Transportation Research Record: Journal of the Transportation Research Board* 2031 (2007), 25–33.

[6] Anneka R Lawson, Vikram Pakrashi, Bidisha Ghosh, and WY Szeto. 2013. Perception of safety of cyclists in Dublin City. *Accident Analysis & Prevention* 50 (2013), 499–511.

[7] Graeme Lindsay, Alexandra Macmillan, and Alistair Woodward. 2011. Moving urban trips from cars to bicycles: impact on health and emissions. *Australian and New Zealand Journal of Public Health* 35, 1 (2011), 54–60.

[8] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *NeuralIPS*. 4765–4774.

[9] Mapbox. 2018. Map Matching API. www.mapbox.com. (2018).

[10] Harikrishna Narasimhan, Weiwei Pan, Purushottam Kar, Pavlos Protopapas, and Harish G Ramaswamy. 2016. Optimizing the Multiclass F-Measure via Biconcave Programming. In *2016 IEEE 16th ICDM*. IEEE, 1101–1106.

[11] Nam Nguyen and Yunsong Guo. 2007. Comparisons of sequence labeling algorithms and extensions. In *Proceedings of the 24th international conference on Machine learning*. ACM, 681–688.

[12] John Parkin and Ciaran Meyers. 2010. The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis & Prevention* 42, 1 (2010), 159–165.

[13] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd SIGKDD*. ACM, 1135–1144.

[14] Paul Schepers and Marjan et al. Hagenzieker. 2014. A conceptual framework for road safety. *Accident Analysis & Prevention* 62 (2014), 331–340.

[15] Michael Townsley. 2009. Spatial autocorrelation and impacts on criminology. *Geographical Analysis* 41, 4 (2009), 452–461.

[16] Jiahui Wu, Lingzi Hong, and Vanessa Frias-Martinez. 2018. Predicting Perceived Cycling Safety Levels Using Open and Crowdsourced Data. In *2018 IEEE International Conference on Big Data*. IEEE, 1669–1676.