Modelling Pedestrian Safety with respect to Road Traffic Crashes by Estimating the Safety of Paths

Charlotte Hannah*[1], Irena Spasić†[2] and Padraig Corcoran‡[2]

[1]School of Mathematics, Cardiff University
[2]School of Computer Science & Informatics, Cardiff University

Summary

We propose a novel linear model of pedestrian safety in urban areas that considers a single independent variable of pedestrian path safety. This variable is estimated by sampling pedestrian paths from an accurate model of such paths and in turn estimating the mean safety of these paths. In an analysis of 15 UK cities, the proposed model is demonstrated to accurately predict the number of pedestrian casualties per million population.

KEYWORDS: Pedestrian safety, modelling, urban areas, pedestrian paths, pavement network.

1 Introduction

Each year 1.25 million deaths are caused by road traffic crashes worldwide (World Health Organisation, 2015). According to the UK Department of Transport, pedestrians accounted for 24% of all road deaths in Great Britain in 2015. Therefore, identifying and subsequently addressing factors contributing to pedestrian road deaths would improve road safety significantly. Typically, such factors are identified by considering the number of pedestrian crashes as a dependent variable and unraveling its relationship to a set of independent variables using mathematical modelling. A sufficiently accurate model can then be used to identify those independent variables corresponding to the factors in question. Note that such a model may also be used in simulation experiments to evaluate the safety under proposed changes to the environment such as the introduction of an increased speed limit on a particular road (Olszewski et al., 2015). Models proposed to date consider independent variables from three main categories: the environment (e.g. the number of pedestrian crossings), the traffic (e.g. the volume of heavy vehicles) and the pedestrians (e.g. mean pedestrian age). In this article we propose a novel linear model that considers a single independent variable. This variable is estimated by sampling pedestrian paths from an accurate model of such paths

*HannahCE@cardiff.ac.uk
†SpasicI@cardiff.ac.uk
‡CorcoranP@cardiff.ac.uk

and in turn estimating the safety of these paths. We argue that this independent variable models the factors contributing to pedestrians crashes directly. This contrasts with previous approaches, which, by considering independent variables describing the environment, traffic and pedestrians, model these factors indirectly (Sze and Wong, 2007; Marshall and Garrick, 2011; Rifaat et al., 2012; Aziz et al., 2013; Guo et al., 2017; Osama and Sayed, 2017).

The remainder of the paper is organised as follows. In Section 2 we describe the proposed model of pedestrian safety. In Section 3, we evaluate the accuracy of the proposed model using statistical data on pedestrians crashes in a total of 15 UK cities.

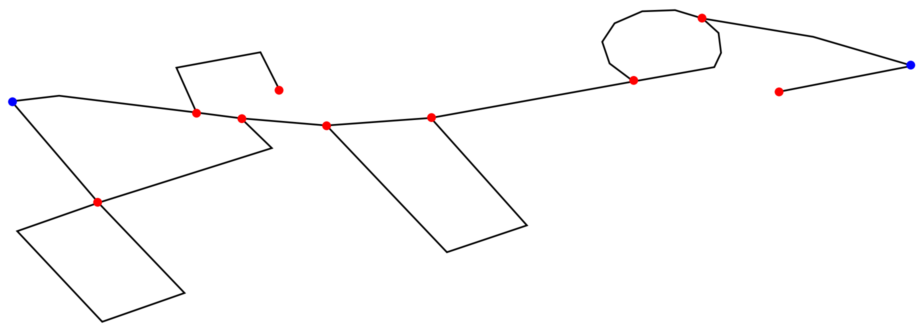
2 Model of Pedestrian Safety

In this section we describe the proposed model of pedestrian safety. The model assumes that a pedestrian always follows the safest path where the safety of a path is defined as the weighted sum of its length and inherent safety. The consideration of length is motivated by the fact there is a positive correlation between distance walked and probability of being involved in a pedestrian crash. Given this assumption, pedestrian safety is modelled in three steps. First the pavement network is modelled as a graph. Next, an edge labelling function is defined which maps each edge in the pavement network to a real number quantifying its safety. Finally, pedestrian safety is estimated as the mean safety of pedestrian paths. A linear regression model of this variable is used to predict the number of pedestrians crashes. These three steps are described in more detail in the following three subsections.

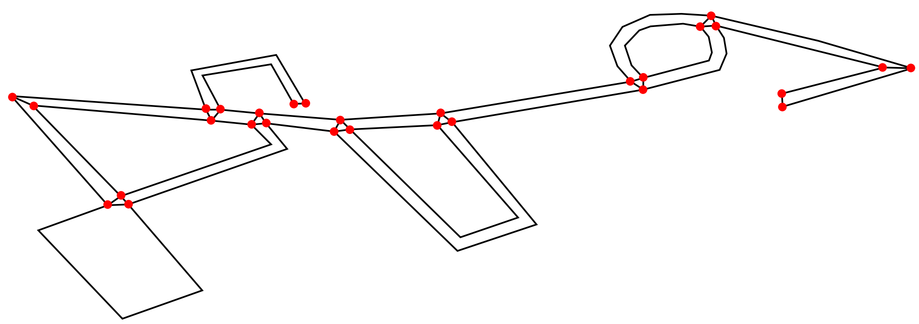
2.1 Pavement Network Construction

Given the fact that pedestrians walk on pavements we consider a network representation of the pavement structure which we refer to as a *pavement network*. We used data from OpenStreetMap (OSM), crowdsourced project that provides free geographic data such as street maps (Mooney and Corcoran, 2012). For most regions in the UK, OSM provides an accurate road network but not necessarily an accurate pavement network. To compensate for this, we automatically construct a pavement network from the corresponding road network based on a set of assumptions regarding the relationship between the roads and pavements. This is known as a network buffering approach to the construction of a pavement network (Karimi and Kasemsuppakorn, 2013).

First, an undirected labelled road network $G^s = (V^s, E^s)$ is constructed where the vertices from V^s correspond to pedestrian crossings, road intersections, and dead-ends, whereas the edges from E^s correspond to road segments connecting these vertices. The vertex labelling function $\lambda : V^s \rightarrow \{0, 1\}$ maps each vertex in V^s to a value in the set $\{0, 1\}$ indicating if the vertex in question is a pedestrian crossing or not. The edge labelling function $\mu : E^s \rightarrow (\mathbb{R}^+, \text{String})$ maps each edge in E^s to a positive real number equalling the length of the corresponding road segment and a string that indicates its type (e.g. 'primary', 'residential', etc.). Figure 1(a) provides a visual representation of a small road network G^s.



(a)



(b)

Figure 1: A sample road network $G^s = (V^s, E^s)$ is illustrated in (a). The elements of V^s are represented using blue and red dots depending if the vertex in question is respectively a pedestrian crossing or not. The corresponding pavement network $G^p = (V^p, E^p)$ is illustrated in (b).

Given a road network G^s, an undirected labelled pavement network $G^p = (V^p, E^p)$ is constructed based on two assumptions. First, each road segment is assumed to have a pavement on both sides. Second, each road segment is assumed to have a non-designated pedestrian crossing at the end; that is, pedestrians can jaywalk at the end of each road. The edge labelling function $\delta : E^p \to (\mathbb{R}^+, \text{String})$ maps each edge in E^p to a positive real number equalling the length of the corresponding pavement/pedestrian crossing and a string equalling its type (e.g. 'primary pavement', 'residential pavement', 'primary designated pedestrian crossing', 'residential jaywalk pedestrian crossing'). Figure 1(b) provides a visual representation of the pavement network automatically constructed from the road network given in Figure 1(a).

2.2 Computing Edge Safety

We assume that the safety of a path is defined as the weighted sum of its length and inherent safety. To model this formally, we construct an edge labelling function $W : E^p \to \mathbb{R}^+$ for a given pavement network G^p. This function is defined in Equation 1 and maps each edge in E^p to the weighted sum of the corresponding length and safety as defined by the functions $L : E^p \to \mathbb{R}^+$ and

$S : E^p \to \mathbb{R}^+$ respectively. The weighting in question is controlled by the parameter $\alpha \in \mathbb{R}^+$. The function L returns the length of the edge in question as specified by the edge labelling function δ of G^p. The function S returns a real value which is a function of the edge type as specified by the edge labelling function δ (e.g. 'primary designated pedestrian crossing'). Smaller values indicate higher safety.

$$W(e) = L(e) + \alpha S(e) \tag{1}$$

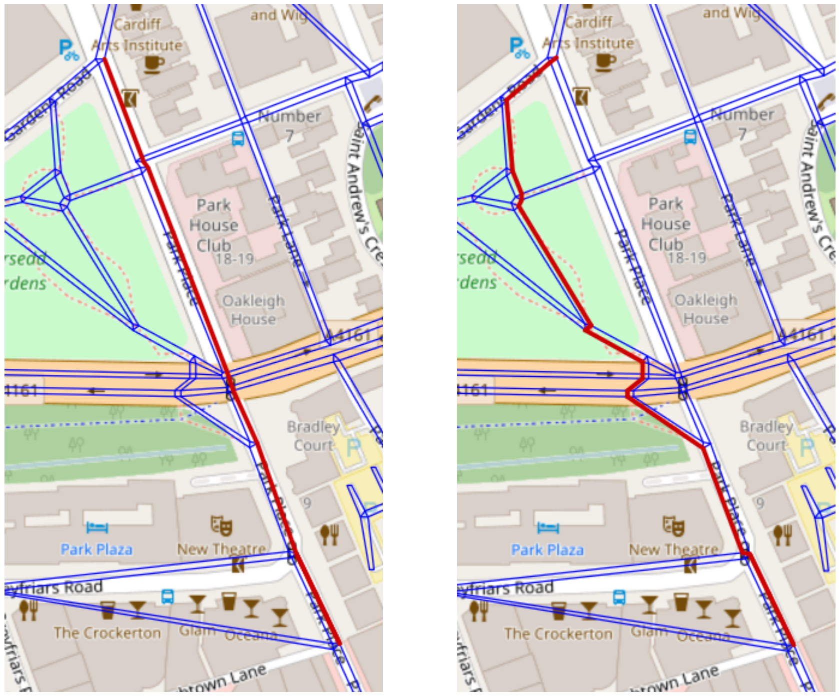
2.3 Pedestrian Safety Model

Given an origin-destination pair, the corresponding pedestrian path safety is estimated as the *safest path*, which is defined as the shortest path in the pavement network G^p with respect to the edge labelling function W. This safest path is calculated using Dijkstra's algorithm. To illustrate our pedestrian safety model we consider the problem of estimating the safest path for an origin-destination pair in the city of Cardiff. The shortest path with respect to geographical distance is indicated in Figure 2(a). It takes pedestrians along a tertiary pavement and across a relatively dangerous primary road jaywalk crossing. The safest path, as determined by the proposed model with $\alpha = 10$, is indicated in Figure 2(b). This path takes pedestrians via a footway instead of a tertiary pavement and across three sets of traffic light pedestrian crossings to avoid jaywalking across a primary road.

To estimate pedestrian path safety for a given urban area, as opposed to an origin-destination pair, the following approach was employed. Rejection sampling was used to generate a sample of origin-destination pairs that is representative and sufficiently large to provide statistically significant results. The pedestrian path safety of each pair in this sample was estimated using the above approach. Finally, pedestrian path safety for the entire area is determined as the mean of these set of values. Given this variable, the number of pedestrian crashes is estimated using a linear regression model.

3 Results

Using the pedestrian safety model described in section 2.3, we estimated the pedestrian safety for 15 cities in the UK. The corresponding number of pedestrian casualties per million population for the year 2015 was obtained from the Department for Transport (2016). These values are given in Table 1 and displayed as a scatter plot in Figure 3. It is evident from the scatter plot that the relationship between the estimated pedestrian safety and number of pedestrian casualties per million population is strongly linear. Using Pearson's correlation test, the null hypothesis that the statistics are uncorrelated is rejected with a p-value of 0.001. Furthermore, the Pearson's correlation coefficient between the statistics is 0.893. This demonstrates that, for a given city, a linear regression model of the estimated pedestrian safety accurately predicts the corresponding number of pedestrian casualties. The linear least squares regression model is illustrated in Figure 3.



(a) (b)

Figure 2: The shortest and safest paths from a location in the top to a location in the bottom is illustrated in (a) and (b) respectively. In both figures the pavement network is represented by a set of blue lines and the path in question is represented by a sequence of red lines.

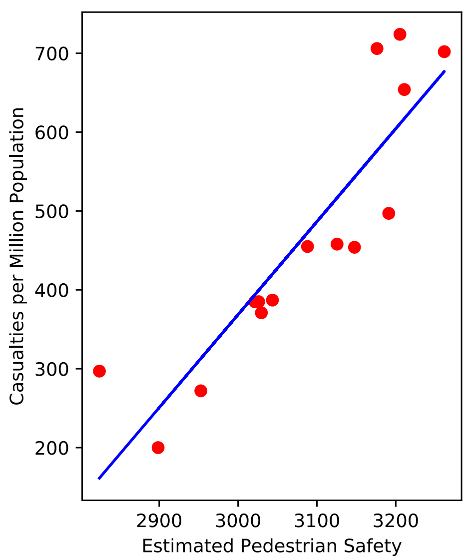


Figure 3: A scatter plot of the statistics contained in Table 1. Each city is represented by an individual red dot. The linear least squares regression model is represented by a blue line.

	Estimated Pedestrian Safety	Pedestrian Casualties per Million Population
Bath	2824.16	297
Bedford	3026.10	385
Blackpool	3205.25	724
Bristol	3147.62	454
Coventry	3021.24	385
Leeds	3191.08	497
Leicester	3210.84	654
Liverpool	3261.51	702
Manchester	3029.55	371
Nottingham	3176.16	706
Reading	3125.53	458
Salford	2898.63	200
Sheffield	3088.05	455
Swindon	2952.80	272
York	3043.59	387

Table 1: For 15 UK cities the corresponding estimated pedestrian safety and number of pedestrian casualties per million population for the year 2015 are stated.

Biography

Charlotte Hannah is a master's student in the School of Mathematics at Cardiff University. Her research interests lie in the areas of safety modelling and accident prevention.

Irena Spasić is a professor in the School of Computer Science & Informatics at Cardiff University. Her research interests lie in the areas of text mining and health informatics.

Padraig Corcoran is a lecturer in the School of Computer Science & Informatics at Cardiff University. His research interests lie in the areas of spatial data analysis and mining.

References

Aziz, H. A., Ukkusuri, S. V., and Hasan, S. (2013). Exploring the determinants of pedestrian–vehicle crash severity in new york city. *Accident Analysis & Prevention*, 50:1298–1309.

Department of Transport (2016). Reported Road Casualties Great Britain: 2015 Annual Report (pages 220-223), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/568484/rrcgb-2015.pdf (Accessed 16 November 2017).

Guo, Q., Xu, P., Pei, X., Wong, S., and Yao, D. (2017). The effect of road network patterns on pedestrian safety: a zone-based bayesian spatial modeling approach. *Accident Analysis & Prevention*, 99:114–124.

Karimi, H. A. and Kasemsuppakorn, P. (2013). Pedestrian network map generation approaches and recommendation. *International Journal of Geographical Information Science*, 27(5):947–962.

Marshall, W. E. and Garrick, N. W. (2011). Does street network design affect traffic safety? *Accident Analysis & Prevention*, 43(3):769–781.

Mooney, P. and Corcoran, P. (2012). Using OSM for LBS - an analysis of changes to attributes of spatial objects. In *Advances in Location-Based Services*, pages 165–179. Springer.

Olszewski, P., Szagała, P., Wolański, M., and Zielińska, A. (2015). Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accident Analysis & Prevention*, 84:83–91.

Osama, A. and Sayed, T. (2017). Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. *Accident Analysis & Prevention*, 107:117–125.

Rifaat, S. M., Tay, R., and de Barros, A. (2012). Urban street pattern and pedestrian traffic safety. *Journal of urban design*, 17(3):337–352.

Sze, N.-N. and Wong, S. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention*, 39(6):1267–1278.

World Health Organisation (2015). Global Status Report on Road Safety 2015. *Retrieved from World Health Organisation, http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/ (Accessed 29 October 2017).*