Analysing Cycle Accidents using a Relationship-Graph Technique

Narushige Shiode*[1], Shino Shiode[2], Ryoichi Shinkuma[3], Manuela Piasentin[4], Takashi Matsuura[5]

[1] Department of Geography, King's College London
[2] Department of Geography, Birkbeck, University of London
[3] Graduate School of Informatics, Kyoto University
[4] Transport Policy Team, Southwark Council
[5] European Centre London Office, Kyoto University

*Email: naru.shiode@kcl.ac.uk

Abstract

This study reports on an ongoing project to understand the causality behind fatal or serious-injury accidents involving cyclists in the Southwark Borough, London. The objective is to apply a text-mining technique that extracts information from unstructured data to identify the risk factors that tend to trigger such accidents. Preliminary results suggest that a specific combination of factors such as young male using mobile phones during daytime in clear weather condition poses the greatest threat, more than those recorded during night-time or under wet condition. These pieces of information were previously not revealed by conventional statistical or spatial analysis. The study gives an overview of the method used and explores its feasibility as a means to analyse larger datasets.

Keywords: Cycle accidents, human behaviour, qualitative text, road junctions, text mining

1. Introduction

Cycling is considered to be an affordable, sustainable and healthy mode of transport, yet its uptake is hampered in part by safety concerns. Transport for London reports that while cycling accounts for only 2% of all journey stages in London (TfL, 2015) and yet they make up 15% of casualties from all traffic accidents, 19% of serious injuries, and 7% of fatalities (Matters, 2016). In 2015, 25,193 traffic accidents were reported in London across all modes of transport with 30,182 casualties. The number of cyclists killed on the streets of London were 16 in 2011, 14 in 2012, 14 in 2013, 13 in 2014 and 9 in 2015, respectively. The Mayor's Office and TfL aim to reduce this figure by 50% by 2020.

The inspiration behind this study is to help reduce such accidents and, thereby, improve the safety of cycling in London. It requires a comprehensive understanding of the place, time and the circumstances of these accidents which would help the decision makers to eliminate or reduce such risk factors. For instance, the majority of traffic accidents are known to occur around junctions to a major road (Morgan, 2010), especially at non-signalised intersections between a major and a minor road (Schepers et al., 2011). Indeed, in recent years, most of the fatal cycle accidents in London have occurred at busy junctions and often in collisions with large vehicles.

A study by the Department of Transport (McRobbie and Wright, 2005) also suggests that the foremost risk factor for cycle accidents is failure to look by motor vehicle drivers, including "vehicle blind spots" of HGVs and "passing too close to the cyclist", as well as cyclists crossing a red light. Others have noted the impact of environmental factors such as the presence of cycle lanes, weather condition, poor lighting on streets, peak-hour traffic, streets with a higher speed limit, as well as demographic and some behavioural factors such as the age of cyclists and driving under influence (Kim et al., 2007; Kaplan, Vavatsoulas and Prato, 2014).

However, investigation of cyclists' behaviour immediately before an accident remains largely unexplored. Our preliminary analysis using regression suggests that physical and environmental factors explain only a small portion of the outcome. This study reports on an on-going project which aims to incorporate various risk factors associated with human behaviour that may have led to an accident.

2. Method

Traffic accidents are inherently spatial and their cause have been often pursued by way of statistical analysis of the variables found in the records. However, they are not designed to interpret free text that accompany such records, which is where much of the description on the particular circumstances is found. In order to extract such information, this study adopts a relationship-graph technique for identifying the strength of association and connectivity among various factors. Detailed description of the method can be found in Ding et al. (2014) and Shinkuma et al. (2014) but, broadly speaking, it is a text-mining technique that constructs a network of various elements of a dataset, and extract meaningful structure from that network though graph mining. In particular, it identifies various objects and attributes embedded in structured and unstructured time-series data as nodes, and describes the association among such objects in the form of a graph network-structure, where the length and the weight of each edge is determined by the number of contact between the objects, and the frequency and periodicity of such contacts. The fact that it gives a cumulative count of all possible connection between each object pair, be it the shortest path or otherwise, help improve the estimation of the extent of association between different objects, such as fatal cycle accidents and various physical, environmental as well as behavioural attributes, beyond what we could usually extract with standard OLS regression.

3. Data

The sample dataset was extracted from the Road Safety Data based on the STATS19 forms, consisting of approximately 300 cases of fatal or serious injury (SI) accidents that were recorded around junctions in Southwark between 2010 and 2014. During this period, Southwark saw the third highest number of cyclist casualties among London Boroughs after Westminster and Lambeth, despite that cycling makes up only 3.4% of the total resident trips (TfL, 2015). Figure 1 shows the distribution of fatal/SI cases across Southwark. Each record consists of (1) time and location of the accident, (2) 76 behavioural factors, (3) physical and environmental features such as junction types and junction controls, and (4) demographic attributes.

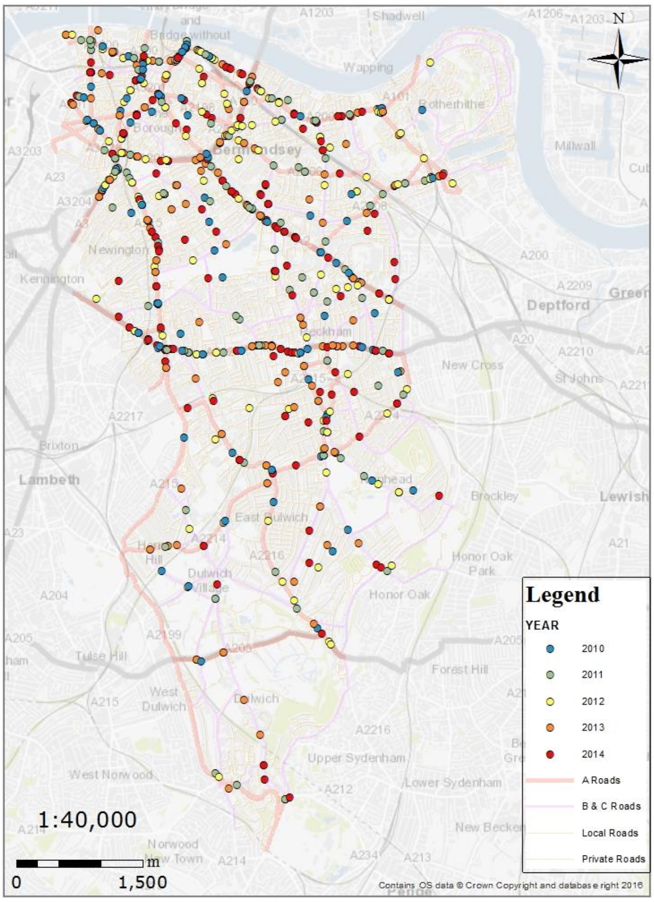
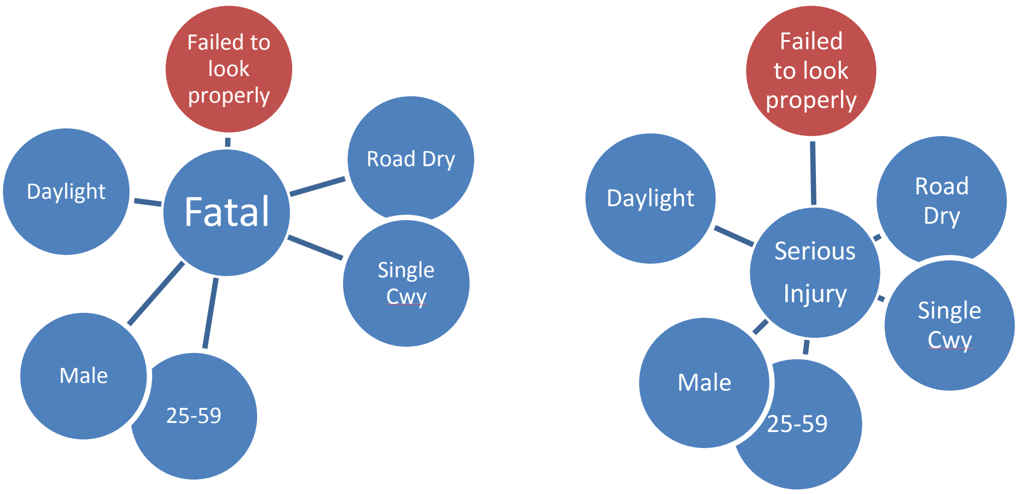


Figure 1: Distribution of fatal/SI cycle accidents across Southwark Borough by year (Source: STATS19).

4. Analysis



(a) (b)

Figure 2: Diagrams illustrating, in the form of a distance index, the association between the types of accidents and the relevant risk factors for (a) fatal and (b) SI cases.

The two diagrams in Figure 2 illustrate the extent that various factors are associated with fatal and SI cases, respectively. The strength of such association is indicated by their proximity to the relevant accident type. Both fatal and SI cases share a near-identical list of 10 most relevant risk factors. The greatest difference is that "failed to look properly (by cyclists)" is identified as the primary factor for fatal accidents whereas, for SI cases, it comes after several other factors (road conditions, demographic characteristics, weather).

Table 1. Behavioural factors associated with fatal and SI cases.

	Fatal	SI
1.	Failed to look properly	Failed to look properly
2.	Careless/Reckless/In a hurry	Careless/Reckless/In a hurry
3.	Poor turn or manoeuvre	Poor turn or manoeuvre
4.	Failed to judge others' path/speed	Failed to judge others' path/speed
5.	Disobeyed give way/stop/marking	Too close to cyclist/horse/pedestrian
6.	Too close to cyclist/horse/pedestrian	Disobeyed give way/stop/marking

Table 1 shows the 6 most relevant behavioural factors associated with fatal or SI cases. The lists are nearly identical between the two types, consisting of the same factors and are roughly in the same order. It should be noted, however, that the relative weight and the overall ranking of these behavioural factors among all variables including physical and environmental factors were different. Behavioural factors are often embedded within the descriptive, unstructured text of each accident report and, therefore, are not necessarily extracted well when using standard statistical or spatial analytical methods.

In terms of the types of junction, T-junctions and stagnated junctions saw the greatest number of fatal/SI cases (over others such as roundabout, mini-junction, crossroads, private drive). As for the methods of junction control, give way and uncontrolled junctions were the worst (over others such as those controlled by traffic signals or by an authorised person).

5. Discussion

The preliminary results suggest that the use of relationship-graph technique offers additional insights into the circumstances around the time of cycling accidents and is effective for interpreting the risks behind traffic accidents. In particular, it identified clearly that "failed to look properly" is the key factor that separates accidents with fatal outcomes from the SI cases.

As this is an on-going project, all results are preliminary and the method needs to be tested with larger data set to confirm these findings, as it would allow us to offer a more powerful explanation for the causes of cycle accidents. This would include the use of the full set of cycle accident cases from the road safety data across UK. Comparison of risk factors for minor injury cases would be another topic of interest. Adding other structured and unstructured data may also help describe in more detail the circumstances surrounding each accident.

6. References

Ding, Z., Omori, Y., Shinkuma, R. and Takahashi, T. (2014) Designing mobility models based on relational graph. *The IEICE Transactions on Information and Systems* 97(12): 3007-3015.

Kaplan, S., Vavatsoulas, K. and Prato, C. (2014) Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of Safety Research* 50: 75-82.

Kim, J., Kim, S., Ulfarsson, G. and Porrello, L. (2007) Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis & Prevention* 39(2): 238-251.

Matters, T. (2016) Casualties in Greater London during 2015. *Transport for London.* [Online]. [Accessed on 1 April 2017]. Available from https://tfl.gov.uk/corporate/publications-and-reports/road-safety.

McRobbie, S. and Wright, A. (2005) *TTS Research*. Wokingham, Berkshire: TRL Limited.

Morgan, A., Dale, H., Lee, W. and Edwards, P. (2010). Deaths of cyclists in London: Trends from 1992 to 2006. *BMC Public Health* 10(1).

Schepers, J., Kroeze, P., Sweers, W. and Wüst, J. (2011). Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. *Accident Analysis & Prevention*, 43(3): 853-861.

Shinkuma, R., Sawada, Y., Omori, Y., Yamaguchi, K., Kasai, H. and Takahashi, T. (2014) A socialized system for enabling the extraction of potential values. In Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmäki, P. and Pironneau, O. (Eds.) *Modeling and Optimization in Science and Technologies*, Springer International Publishing, pp.385- 404.

Transport for London (2015) Travel in London Report 8. *Transport for London.* [Online]. [Accessed on 1 April 2017]. Available from http://content.tfl.gov.uk/travel-in-london-report-8.pdf