Accident Analysis and Prevention 65 (2014) 105–113

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Network-level accident-mapping: Distance based pattern matching using artificial neural network☆

Lipika Deka*, Mohammed Quddus[1]

School of Civil and Building Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 2 December 2013
Accepted 5 December 2013

Keywords:
Accident-mapping
Pattern-matching
Artificial neural network

ABSTRACT

The objective of an accident-mapping algorithm is to snap traffic accidents onto the correct road segments. Assigning accidents onto the correct segments facilitate to robustly carry out some key analyses in accident research including the identification of accident hot-spots, network-level risk mapping and segment-level accident risk modelling. Existing risk mapping algorithms have some severe limitations: (i) they are not easily 'transferable' as the algorithms are specific to given accident datasets; (ii) they do not perform well in all road-network environments such as in areas of dense road network; and (iii) the methods used do not perform well in addressing inaccuracies inherent in and type of road environment. The purpose of this paper is to develop a new accident mapping algorithm based on the common variables observed in most accident databases (e.g. road name and type, direction of vehicle movement before the accident and recorded accident location). The challenges here are to: (i) develop a method that takes into account uncertainties inherent to the recorded traffic accident data and the underlying digital road network data, (ii) accurately determine the type and proportion of inaccuracies, and (iii) develop a robust algorithm that can be adapted for any accident set and road network of varying complexity. In order to overcome these challenges, a distance based pattern-matching approach is used to identify the correct road segment. This is based on vectors containing feature values that are common in the accident data and the network data. Since each feature does not contribute equally towards the identification of the correct road segments, an ANN approach using the single-layer perceptron is used to assist in "learning" the relative importance of each feature in the distance calculation and hence the correct link identification. The performance of the developed algorithm was evaluated based on a reference accident dataset from the UK confirming that the accuracy is much better than other methods.

1. Introduction

In 2012 Great Britain saw 1754 deaths, 23,039 seriously injured and a total of 195,723 casualties in reported road accidents (DoT, 2013). World Health Organisation estimates over 1 million deaths world-wide as a result of road accidents (WHO, 2013). To make roads safer and save life and money, understanding the safety performance of the underlying road network and identifying link-level accident hot-spots so as to design engineering countermeasures are critical. Hence, accurate assigning of accidents to the correct road segments where the accidents occurred is a vital precursor for safety related applications such as accident risk modelling, risk mapping and accident hot-spot identification. Accident risk modelling and link feature identification are also essential for the design and manoeuvring of intelligent, self-driving vehicles of the future.

Data and information on traffic accidents (such as their geographical references in terms of road name, district name, accident location denoted as *x*- and *y*-coordinates, number of casualties and their characteristics, number of vehicles/types involved) in most countries are recorded by the police by either visiting the place of accident or by conducting remote inquiries. Due to reasons such as the situation at the accident site, accuracy issues related to positioning methods/instruments such as GPS or national grid reference (Quddus et al., 2007), mistakes on part of the police, etc., errors exist in the police recorded accident data (Shinar et al., 1983; Levine et al., 1995; Austin, 1995; Aptel et al., 1999; Loo, 2006; Tarko et al., 2009; Khan et al., 2004). For example, in the UK (Austin, 1995) as well as in Abu Dhabi (Khan et al., 2004), location of the accident has been identified as the most inaccurately recorded data item. Shinar et al. (1983) reported that in the United States, highway feature data such as gradient, speed limit, surface composition and curvature were the most inaccurately reported information, whereas accident location, date, passenger and vehicle information were the

* Corresponding author. Tel.: +44 01509223401.
E-mail addresses: l.deka@lboro.ac.uk (L. Deka), m.a.quddus@lboro.ac.uk (M. Quddus).
[1] Tel.: +44 01509228545.

0001-4575/$ – see front matter. Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.12.001

most reliable data. More recently Tarko et al. (2009) reports that large number of missing accident data entries, spelling mistakes in road names, presence of alternate road names poses issues in the US.

Such data inaccuracies invalidate and significantly affect any analysis rooted in it. Hence, accident records must be validated and accidents should be mapped on to the correct links before being analysed for the purpose of enhanced road safety applications. The objective of an accident-mapping algorithm is to snap traffic accidents onto the correct road segments and correct position on the selected segment, given inaccurately recorded location information.

Much of the accident-mapping research effort in the past has been towards identification of mistakes in road accident data records (Shinar et al., 1983; Levine et al., 1995; Austin, 1995). This was initially done manually and through the use of computer validation techniques (Shinar et al., 1983) and later through the use of Geographical Information System (Austin, 1995). Natural language understanding techniques have also been used to retrieve information from accident reports written in free format plain English and used to validate the accompanied records in pre-defined formats (Wu and Heydecker, 1998). In the near past, significant progress has been made in not only identifying mistakes in police reported accident records, but also in correcting those mistakes to identify correct road segments where the accident took place. The underlying concept of most of these accident-mapping endeavours has been towards integration of the accident database of police accident reports with the road network database. These accident positioning attempts include GIS-based approaches of snapping accidents to nearest road segment or junction and then iteratively validating and correcting associated variables such as district and road name (Loo, 2006). Dutta et al. (2007) used the information of accident location, direction and distance from the accident location to develop an approach of mapping the accident position on the correct intersection or local road segment. Probabilistic record linkage methods have also been used to link erroneous accident record database with the road-network database, thereby positioning accidents on the correct road segments (Tarko et al., 2009). Although these efforts have greatly improved the quality of accident location data, but positioning accuracy is compromised in complex road network scenarios such as presence of multiple parallel roads, roundabouts and other types of junctions. Such compromises are due to the limitation in the heuristic techniques (Loo, 2006) and probabilistic formulas (Tarko et al., 2009) used as well as due to the limitations in the type of data available.

The aim of this paper is therefore to develop a new accident mapping algorithm based on the common variables observed in most accident databases (e.g. road name and type, direction of vehicle movement before the accident and recorded accident location). Our approach involves representing an accident as well as all road links as feature vectors of these variables and a distance based pattern matching technique is employed to map an accident to the link with which its “pattern” or feature vector matches most closely. Since, each feature do not contribute equally towards the identifications of the correct link (Quddus et al., 2007; Tarko et al., 2009; Velaga et al., 2009; Greenfeld, 2002), an artificial neural network is employed to “learn” the relative significance of the above stated features. Once the correct link has been identified, the accident position on the link is determined through perpendicular projection of the accident location on the selected link. In the case where the perpendicular projection of the accident location falls outside the link, the closest end point of the link is fixed as the point of accident.

The performance of approach was evaluated against a reference accident dataset that was compiled through manual mapping of accidents onto links through the use of GIS software. Additional variables such as vehicle position at time of accident, second road name (in the case of junction accidents) were used during manual mapping apart from the variables used in our algorithm. In the absence of correct reference data on accident location, any form of mapping must be treated with caution. Manual mapping is extensively labour intensive and hence we evaluated our results against only a subset of 560 accidents from the accident data set on UK's strategic road network for the year 2012.

The remaining sections present the detailed description of the approach (Section 4) and evaluation (Section 5) result by first presenting a brief literature review (Section 2) and overview of the existing challenges (Section 3).

2. Related work

The police department of different countries collect data on traffic accidents with subtle difference in the type of data from country to country. For example, an accident location in Wisconsin is recorded as the direction and distance from a junction (Dutta et al., 2007) whereas an accident location in the UK is recorded in terms of its geographic co-ordinates (Austin, 1995). This section presents the existing accident-mapping techniques that utilise the location specific available accident data.

Loo (2006) developed a GIS-based spatial data validation methodology to map accident locations in Hong Kong to a precise road section. The methodology snaps an accident to the nearest junction if the accident occurred at a junction else snaps it to the nearest road. The approach then checks to validate the road and district name of the mapped accident location to that of the original recorded accident. In the case of a mismatch in the district name, the algorithm amends the incorrect field with the correct data associated with the accident-mapped location. In the case where the road name does not match, the algorithm maps the accident location to the next nearest road or junction, this time amending the accident record with the current mapped road name in case the original road name still does not match.

Tarko et al. (2009) employed the concept of probabilistic record-linkage using the Fellegi–Sunter model (Fellegi and Sunter, 1969) with the Expectation-Maximization (EM) method to map accidents to road segments. The features used for accident-mapping were County ID, Township ID, City ID, main road name, reference road name, shoulder type, median presence and junction type. The Fellegi–Sunter model estimates the probability of the occurrence of an accident on a road through pair-wise matching of features in records from respective datasets. Each feature is adjusted with weights, where weights determine an attributes relative contribution towards the final decision of match/no-match. Probabilities above a certain threshold will translate to a match, below a second threshold translates to a no-match and any probability in-between suggests human intervention for the final decision of a match/no-match. The EM method is used to estimate the respective feature weights. The method was evaluated using a test sample of 137 real and simulated accident data from the state of Indiana in the USA and saw that even though almost all accidents were mapped, 80% of accidents were mapped to more than one link and mapping to intersections were not very efficient.

Dutta et al. (2007) developed a tool to digitally plot Wisconsin's local road accidents on a GIS map integrating it with complete information on the mapped accident. Two primary data sources namely the Wisconsin Accident Database and the Wisconsin's Information System for Local Roads were used and accident mapping for intersection accidents and segment accidents were done separately. Accident mapping methodology mainly involved parsing portions of street names (i.e. prefix, name type and suffix component) of each accident record and matching them against records in the road