Combining traffic accident data, crowdsourced data about dangerous spots and vehicle kinematic data

Jörg Ehlers [1], Arno Wolter [2]*

[1] RWTH Aachen University, Institute of Highway Engineering, Germany

[2] Initiative für sichere Straßen GmbH, Bonn, Germany

** Corresponding author: ehlers@isac.rwth-aachen.de*

Abstract

In today's digital age, more and more data is collected every day in the field of transportation. Still, for evaluating road safety, authorities often solely rely on historical traffic accident data. In the presented project, three data sources are aggregated into a single hazard score that allows a proactive assessment of road safety. The accident data is enriched with data on safety-critical driving events and crowdsourced user reports on dangerous locations. The novel data sources allow to identify dangerous locations before a large number of accidents occur and to evaluate the effectiveness of road safety measures more quickly. The network-wide hazard score for the whole German road network is displayed on a map which is available at www.gefahrenstellen.de.

Keywords

Proactive road safety; Safety assessment; crowdsourcing; vehicle kinematic data

Introduction

In today's digital age, more and more data is collected every day in the field of transportation. Still, for evaluating road safety, authorities often solely rely on historical traffic accident data. This leads to several shortcomings. Accidents must happen before measures are taken. This is ethically problematic, particularly if mobility patterns are changing quickly. For example, in Germany, the share of cycling has been rising steadily for two decades (Nobis and Kuhnimhof, 2018) and has again increased significantly during the Corona pandemic. 25% of people in Germany say they cycle more often in 2020 than in the same period of the previous year (Sinus Institut, 2021). However, the bicycle infrastructure is not adapted to these changes in mobility patterns. Sections at risk therefore remain for a longer period of time, as it may take several years to obtain a large enough number of accidents to evaluate traffic safety. Additionally, underreporting of accidents amplifies this issue as up to 70 % of accidents are not reported to the police (Layreshyn et al., 2010; Juhra et al., 2012). As a result, road users are put at an unnecessary risk if measures are taken too late.

There is therefore an urgent need to take new paths in the field of road safety work. The website www.gefahrenstellen.de aims to build a framework to detect potential dangers in road traffic at an early stage by combining existing safety-relevant data with new data sources. Moreover, the platform addresses all stakeholders from authorities and the industry to the road user itself.

New data sources for a proactive Assessment of road safety

Overview of data sources

The platform www.gefahrenstellen.de addresses the above-mentioned problems. Three different data sources are aggregated into a single hazard score that is displayed on a map. The idea for the hazard

score and the early detection is based on the theory of Hyden's (1987) so-called Safety Pyramid. At the top of the pyramid are the traffic accidents, which are usually easily visible through statistics and which must be avoided. The lower, much larger part of the pyramid represents other traffic events such as near-misses and conflict situations. The theory is that a frequent occurrence of near-misses could point to future accidents if the connection between accidents and the occurrence of near-misses is known. Thus, these critical traffic events and other surrogates safety measures (SSM) are collected in a database. The information in the database can be then used for the 3Es of road safety (engineering, education, enforcement).

For the identification of hazardous areas on the road network, information from three data sources is analysed and aggregated:

- accident data,
- crowdsourced road user reports,
- vehicle kinematic data.

Accident data is essential for road safety work. With this data, the fundamental safety analyses (e. g. detection of black spots) are carried out and priorities for measures are set. This data represents the most valid source for safety assessment in Germany as there exists a standardized procedure for data collection. Accident data therefore form the basis for the hazard score. A data set on accidents with personal injury for all German states is used. This data set is available as Open Data (unfallatlas.statistikportal.de). For data protection reasons, the data attributes are limited. However, the features provided are sufficient for the identification of hazardous locations as the location and severity of accidents are specified and further data sources enrich this information.

Reports from road users

Road users are often aware of dangers on their daily routes. Recent studies show that the perception of subjective risks is an indicator for objective risks as well (Stülpnagel & Krukar, 2018; Stülpnagel et al. 2022). Reports from road users on deficiencies in the infrastructure can therefore enhance the road safety assessment. To collect this kind of information systematically and on a large scale, the website www.gefahrenstellen.de leverages the principle of crowdsourcing. Users can report dangerous locations on an interactive map for the whole German road network. Predefined prompts help the users to describe the hazard. The response options are also standardized to make the submission of reports as user-friendly as possible so that a large number of reports can be collected. Furthermore, this ensures that the information can be easily merged with the other data sources. For this reason, the categories and response options were derived from the police report form for accidents. For example, the category "Danger for" gives an indication of the types of traffic involvement and "Type of danger" enables a selection of possible causes of danger. Aside from these mandatory fields, users can also give supplemental information about the cause of danger in more detail as well as write comments und upload photos. In addition, users can support existing danger spots and thus signal their approval.

Overall, 9.000 user reports have been created and 30.000 interactions have been made on the website from May 2018 until September 2022. Bicyclists are the road user group most often reported as being at risk. In 63 percent of all user reports, bicyclists were at danger. This can be explained by the fact that most of the reports were submitted in urban areas and that Germany cities often lack a connected cycling infrastructure. The most common type of hazard is “confusing traffic layout” followed by “driving errors from drivers”. A striking observation is that there are a particularly large number of reports for two cities: Aachen and Bonn. These cities were part of a feasibility study in which the website www.gefahrenstellen.de were initially developed and tested. In this context, the website was promoted extensively there.

In order to use the crowdsourced reports for safety assessments of the road infrastructure, the quality of the reports must be known in advance. Since the reports do not directly address the individual subjective risk perceptions, but rather identify deficits in the infrastructure, the reported danger spots are evaluated according to objective criteria. For this purpose, on-site audits were carried out in four major cities in different parts of Germany. To conduct road safety audits in Germany, the Federal Highway Research Institute (Bundesanstalt für Straßenwesen) has developed a deficit inventory. An abbreviated form of this deficit inventory is used for the validation user reported danger spots. The abbreviated form contained 70 negatively polarized items clustered into 12 scales. Based on these items, an overall rating was determined: “deficit confirmed”, “uncertain”, “deficit not confirmed”. The overall rating was finally matched with the road user report. Every on-site audit was conducted by at least two researchers to minimize bias.

For this pilot study, on site audits at 77 reported danger spots were conducted. Overall, half of the danger spots were be confirmed as dangerous. For another 27 % of the audited danger spots, the rating was “uncertain”, i.e. the audits had to be conducted at another time – e. g. at darkness or during peak hour. The results further showed that the rating of the danger spots as hazardous was related to an increased number of supporters and comments. This means that danger spots with a high number of supports and comments could likely be an indicator of objective risks. An association with accidents could not be found. The user reports can thus reveal danger spots before a large number of accidents occur. For further information on the method and results of the pilot study, see Olma et al. (2022). Additional on-site audits are conducted for further research on crowdsourced danger reports. This research aims to gain further insights into the usage of subjective reports as an indicator for traffic safety.

Vehicle kinematic data

The third data source for the hazard score is vehicle kinematic data. Vehicle kinematics can indicate safety critical driving manoeuvres, e. g. emergency brakings or evasive manoeuvres. This data is an important addition to the early detection of danger spots because these critical traffic events occur far more frequently in everyday traffic than accidents. According to the Safety Pyramid (cf. “Overview of data sources”), a high number of critical events, taking into account the number of trips per segment, could indicate at hazards on a specific road segment before accidents occur.

Critical traffic events are derived from acceleration data that is collected with acceleration sensors. GPS sensors provide coordinates of the critical events which can then be assigned to a road section. Based on threshold values for the acceleration (in longitudinal and lateral direction), both the type of driving manoeuvre and the severity of the critical event can be differentiated. In test series and in the literature (Perez et al., 2017) it could be shown that this methodology for the acquisition of the critical traffic events provides reliable results.

Vehicle kinematic data thus can contribute the data base for early detection of hazardous locations. Moreover, this data can also be used to evaluate measures to increase traffic safety. Since, compared to accident data, a shorter observation period is sufficient to draw conclusions about a change in the safety situation due to the larger volume of data, the effect of a measure can also be determined more quickly.

Vehicle kinematic data can therefore complement accident data when collected on a large scale. For www.gefahrenstellen.de, vehicle kinematic data is provided via partner companies that already collect this data on all roads in Germany. Vehicle kinematics, however, are not yet integrated into the live version of gefahrenstellen.de. This is planned for 2023.

Hazard score

Accident data, crowdsourced user reports and safety critical driving manoeuvres are blended together into a hazard score for each element of the road network. For this purpose, the road network was divided into individual 25 m segments and junctions. The calculation of the hazard score comprises two steps. First, a baseline hazard score is calculated based on accident data. In order to take into account the relationships between accidents on neighbouring segments, these point events are smoothed over several segments, whereby events closer together are given more weight than events further away. In the second step, safety critical driving manoeuvres and user reports on dangerous spots enhance the hazard score. For the user reports, the increase of the score depends on the number of reports as well as the number of supports and comments. For the critical driving events, the increase depends on the share of critical events to the total number of trips per road element.

The resulting safety score thus combines a reactive and proactive perspective to further enhance current traffic safety work. In this way, measures can be taken early on to minimize hazards. At the same time, citizens can be involved and the acceptance of measures can be increased.

Use Cases and further research

The hazard score is calculated for all road categories in Germany, i.e. for urban and rural street networks as well as motorways and residential roads. This allows for broad use cases when made available to those involved in road safety work such as local authorities, police, science, engineering offices, navigation providers and car manufacturers.

Use Case: analysis tool for municipalities

A separate, access-restricted analysis tool for municipalities and police authorities was built as a first professional use case for the hazard score and the underlying data sources. This tool aims to help authorities in their road safety work along with communicating with citizens on safety related issues.

With the analysis tool, dangerous locations can be identified in addition to accident black spots. The vehicle kinematic data and the user report help to decide where measures could be taken in order to improving traffic safety proactively. For this purpose, road segments are highlighted where the proactive data indicate a high danger level but few accidents have occurred. Additional information on the hazards are provided as well as options for filtering, for e. g. specific road user groups.

In addition to identifying new hazards, the tool can also be used for citizen participation processes. Citizens' feedback can be obtained via the gefahrenstellen.de website. Municipalities can then enter into a dialog with citizens via the analysis tool. Public comments can be made about the current status or about planned measures. In contrast, municipalities in Germany today often receive comments on road infrastructure related issues through several channels. Often the answers by the agencies are directed only to the individual and not made publicly available. With the gefahrenstellen.de tool, inquiries from citizens can be collected centrally and answered more efficiently, as the answers can be viewed by the public.

Future research

In a new research project, an enhanced methodology will be developed for collecting and using vehicle kinematic data on a network-wide level from different road user groups. Safety-critical events and traffic flow data from bicycles, e-scooters and motor vehicles will be used to determine the behavior of road users and the influence of the infrastructure. The critical events and the contextual and infrastructural data will be fed into a statistical model to better assess safety impacts for future traffic developments.

References

Juhra, C., Wieskötter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., Malczyk, A., Heckwolf, M., Raschke, M. (2012). Bicycle Accidents – Do We Only See the Tip of the Iceberg?: A Prospective Multi-Centre Study in a Large German City Combining Medical and Police Data. Injury, 43 (12), 2026–2034. https://doi.org/10.1016/J.INJURY.2011.10.016.

Laureshyn, A., Svensson, Å., Hydén, C. (2010). Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accident Analysis & Prevention, 42(6), 1637-1646. https://doi.org/10.1016/j.aap.2010.03.021.

Nobis, C. & Kuhnimhof, T. (2018). Mobilität in Deutschland – MiD Ergebnisbericht. Studie von infas, DLR, IVT und infas 360 im Auftrag des Bundesministers für Verkehr und digitale Infrastruktur (FE-Nr. 70.904/15). Bonn, Berlin. www.mobilitaet-in-deutschland.de

Olma, J., Bode, T., Ehlers, J., Sutter, C. (submitted). Road Users' Reports on Danger Spots - The Crowd as an Underestimated Expert?. *Safety.*

Perez, M.A., Sudweeks, J.D., Sears, E., Antin, J., Lee, S., Hankey, J. M., Dingus, T. A. (2017). Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data, Accid. Anal. Prev. 103, 10–19. https://doi.org/10.1016/j.aap.2017.03.005.

Sinus Institut (2021). Fahrrad-Monitor Deutschland - Corona-Befragung 2020 - Ergebnisse einer repräsentativen Online-Befragung. https://www.bmvi.de/SharedDocs/DE/Artikel/StV/Radverkehr/fahrradmonitor-2020.html

von Stülpnagel, R., Krukar, J. (2018). Risk perception during urban cycling: an assessment of crowdsourced and authoritative data. Accid. Anal. Prev. 121, 109–117. https://doi.org/10.1016/j.aap.2018.09.009.

von Stülpnagel, R., Petinaud, C., Lißner, S. (2022). Crash risk and subjective risk perception during urban cycling: Accounting for cycling volume. Accid. Anal. Prev. 164, 109–117. https://doi.org/10.1016/j.aap.2021.106470.