



POLITECHNIKA WARSZAWSKA

ISSN 1230-9265

PRACE NAUKOWE • TRANSPORT • z. 88

Marzena Nowakowska

**MODELOWANIE ZWIĄZKÓW
MIĘDZY CECHAMI DROGI
A ZAGROŻENIAMI W RUCHU
NA DROGACH ZAMIEJSKICH**



**OFICyna WYDAWNICZA POLITECHNIKI WARSZAWSKIEJ
WARSZAWA 2013**

Marzena Nowakowska

Wydział Zarządzania i Modelowania Komputerowego
Politechnika Świętokrzyska

MODELOWANIE ZWIĄZKÓW MIĘDZY CECHAMI DROGI A ZAGROŻENIAMI W RUCHU NA DROGACH ZAMIEJSKICH

Rękopis dostarczono 30.01.2013 r.

Elementy systemu *człowiek-pojazd-droga* mogą stwarzać okoliczności, które sprzyjają powstawaniu zagrożeń w ruchu drogowym. Analizy wskazują na decydujące znaczenie czynnika ludzkiego w zbiorze cech opisujących przyczyny wypadków drogowych. Droga i jej otoczenie, wymieniana jako drugi istotny element wpływający na zagrożenia, to coraz częściej podnoszony przedmiot badań bezpieczeństwa ruchu drogowego (brd). Praca wpisuje się w nurt tych badań.

Podjęto analizę cech środowiska drogi (projektowych, otoczenia, środowiska) w celu wskazania ich znaczenia w definiowaniu okoliczności wypadku drogowego. Badania zrealizowano na podstawie danych z rejestrów policyjnych, a diskutowaną jednostką był wypadek drogowy. Wykorzystując źródła dodatkowe, udostępnione przez zarządcę dróg, utworzono zbiór danych charakteryzujący drogę w miejscach wystąpień wypadków drogowych i scalono go ze zbiorem danych o wypadkach. Zagrożenie brd zdefiniowano za pomocą trzech cech charakteryzujących zdarzenie drogowe. Były to: przyczyna wynikająca z błędnego zachowania kierującego pojazdem, rodzaj zdarzenia oraz status zdarzenia (największy stopień uszkodzenia wśród ofiar). W analizowanym zbiorze danych szczegółowych (niezagregowanych) występowały cechy ilościowe i jakościowe. Przedmiotem badań były wypadki bez udziału pieszych, z udziałem pojazdów silnikowych, z wybranego regionu kraju, zarejestrowane na zamiejskich drogach krajowych jednojezdniowych dwukierunkowych. Kompleksowy naukowy i poznawczy cel pracy obejmował aspekt merytoryczny i metodologiczny.

Zwrócono uwagę na znaczenie jakości danych oraz możliwości poprawy tej jakości w drodze weryfikacji i czyszczenia. Uwypuklono wpływ struktury danych na wiarygodność wyników. Ważnym elementem badań był zaproponowany sposób modyfikacji zasobów i oceny analizowanych cech z uwagi na potrzeby badawcze. Zasadnicze zadania merytoryczne koncentrowały się na następujących zagadnieniach:

- wyodrębnienie wzorców wypadków drogowych oraz wskazanie na te spośród analizowanych cech, które miały istotne znaczenie w definiowaniu tych wzorców,
- zbudowanie modeli prognostycznych w celu określenia, w jaki sposób istniejące warunki drogowe wpływały na zagrożenie na drodze. Zadania modelowania objęły:

- klasyfikację zachowania sprawcy zdarzenia i rodzaju zdarzenia dla wypadku drogowego za pomocą logistycznych modeli nominalnych,
- klasyfikację statusu zdarzenia dla wypadku drogowego za pomocą logistycznych modeli porządkowych,
- wprowadzenie metod metauczenia się, w szczególności połączenie warstwowych metod samoinicjowania z losowaniem niwelującym, w celu wzmocnienia klasyfikatorów dla identyfikacji związków i prognozowania,
- wykorzystanie bootstrapowych rodzin klasyfikatorów logistycznych do wyznaczenia aposteriorycznych rozkładów zachowania sprawcy, rodzaju i statusu zdarzenia dla wypadków drogowych reprezentujących profile wzorców wypadków drogowych.

Analizy zostały przeprowadzone niezależnie dla wypadków drogowych z udziałem jednego pojazdu i dla wypadków drogowych z udziałem co najmniej dwóch pojazdów.

W zakresie metodologicznym przedstawiono różne sposoby analizowania tych samych zasobów danych: za pomocą grupowania pojęciowego w postaci metody *K-średnich* i map Kohonena oraz za pomocą klasyfikatorów statystycznych w postaci wielokrotnej regresji logistycznej. Zarówno w zakresie identyfikacji wzorców wypadków drogowych, jak i klasyfikowania cech zagrożenia bezpieczeństwa ruchu drogowego wykonano szereg eksperymentów, które umożliwiły porównanie różnych narzędzi badawczych oraz wskazanie ich słabości i zalet. Przedstawiono możliwości zastosowań i wzbogacenia metod drażenia danych w analizach zagrożeń w ruchu drogowym.

Wykonane badania pozwoliły na sformułowanie uwag i wniosków dotyczących każdego etapu pracy: przetwarzania i eksploracji danych, metod grupowania pojęciowego i regresji logistycznej oraz wyników analiz. Zwrócono uwagę na znaczenie regionu, dla którego prowadzi się badania oraz na potrzebę rozłącznych analiz danych o wypadkach drogowych z udziałem jednego pojazdu i o wypadkach drogowych z udziałem co najmniej dwóch pojazdów. Dla każdego z tych zbiorów danych związki między cechami drogi i jej otoczenia oraz cechami zagrożenia były różne, a zidentyfikowane wpływy środowiska drogi na zachowanie sprawcy, rodzaj i status zdarzenia miały różne nasilenie lub różne kierunki.

MODELLING THE RELATIONSHIPS BETWEEN ROAD FEATURES AND TRAFFIC THREATS ON NATIONAL ROADS

Summary

The elements of a human-vehicle-road system can create circumstances that contribute to formation of road traffic threats. Analyses indicate the crucial meaning of the human factor among factors causing road accidents. The road and its vicinity, regarded as the second element of the system influencing the threat, is the subject repeatedly investigated in road traffic safety research. The work follows this research trend.

The analysis of road features (design features, road surroundings and environmental considerations) has been undertaken in order to identify their importance in defining the specific circumstances of the road accident phenomenon. The research is conducted on the basis of actual data from police records and a road accident is the unit under discussion. Using additional data sources, obtained from a road administration office, a database has been created containing information on technical road parameters as well as on a road traffic environment at the road accident location. Then road data and accident data were joined, using computer programs elaborated by the author. The road threat was defined through three accident attributes. They were as follows: the causative accident factor resulting from a driver's faulty behaviour, the accident type and the accident severity. Both quantitative and qualitative features defined the analysed detailed (disaggregated) dataset. The research was carried out on non-pedestrian motor vehicle accident data recorded on two-way two-lane rural national roads from a selected region of Poland. The complex scientific and cognitive character of the work has covered both merit and methodological aspects.

The importance of data quality and the possibilities of quality improvements by verification and cleaning were discussed. The influence of the data structure on the reliability of the results was also pointed out. Both the proposed way of the dataset modification and the assessment of the analysed variables from the research point of view were key elements of the research. The substantial merit tasks were as follows:

- the identification of road accident patterns and the indication of the factors that play the most important role in defining the patterns,
- the development of predictive models in order to assess to what extent existing road conditions influence the threat. The modelling tasks covered:
 - the classification of a driver's behaviour and accident type using nominal logistic models,
 - the classification of accident severity using ordinal logistic models,
- the introduction of meta-learning methods, in particular combining stratified bootstrapping with levelling sampling, in order to strengthen the classifiers,
- the application of bootstrap logistic classifier ensembles to the calculation of posterior distributions for the driver's behaviour, the accident type, and the accident severity for sample observations that were depicted by the road accident patterns.

All the analyses mentioned above were conducted separately for single vehicle and multi vehicle accidents.

In terms of methodology, different approaches to the same data were presented: conceptual clustering by the *K-means* method and the Kohonen maps and statistical classifiers by multiple logistic regression. In accident patterns identification as well as in threat modelling, a series of experiments was performed, which allowed the comparison of different research tools and the indication of their advantages and disadvantages. The possibilities of applications and the enrichment of methods of data mining in the analysis of road traffic threats were presented.

The conducted research made it possible to formulate some remarks and conclusions concerning each work stage, i.e. data processing and exploration, methods of both conceptual grouping and logistic regressions as well as analysis results. Attention was paid to the significance of a region character for which the research is conducted and to the need of separate analysis for single vehicle and multi vehicles accident data. For each of the datasets, the relationships between the road and its surrounding features and the road accident threat features were different. Also identified influences on the driver's behavior, type and severity of accident had different intensity or different directions.

SPIS TREŚCI

Wprowadzenie	5
Ważniejsze oznaczenia.....	11
1. Synteza wyników wybranych prac związanych z analizami zagrożeń w ruchu drogowym ..	14
1.1. Wstęp.....	14
1.2. Wzorce zdarzeń drogowych i ich profilowanie	14
1.3. Modele prognostyczne w analizach bezpieczeństwa ruchu drogowego	19
1.3.1. Klasyfikacja modeli prognostycznych.....	19
1.3.2. Ilościowe modele zagregowane	19
1.3.3. Możliwości i ograniczenia w prognozowaniu jakościowych cech zdarzenia drogowego	20
1.4. Wnioski do ukierunkowania badań własnych	30
2. Cel i zakres pracy	35
3. Baza danych dla wybranego obszaru badań	39
3.1. System Ewidencji Wypadków i Kolidzji jako źródło danych o zdarzeniach drogowych..	39
3.2. Struktury danych o zdarzeniach drogowych.....	40
3.3. Czyszczenie, uzupełnienie i uporządkowanie danych o zdarzeniach drogowych	45
3.4. Obszar badań.....	52
3.5. Dane o drodze i charakterystyka źródeł	54
3.6. Złączenie danych o wypadku i o drodze w miejscu wystąpienia wypadku drogowego ..	57
4. Wybór zmiennych do analiz	61
4.1. Eksploracja danych o wypadkach drogowych	64
4.2. Eksploracja danych o wypadkach drogowych z udziałem jednego pojazdu	66
4.3. Eksploracja danych o wypadkach drogowych z udziałem co najmniej dwóch pojazdów	71
4.4. Badanie korelacji i współliniowości	76
5. Narzędzia grupowania pojęciowego do identyfikacji wzorców wypadków drogowych	78
5.1. Przedstawienie zagadnienia	78
5.2. Algorytm <i>K-średnich</i>	80
5.3. Samoorganizujące się mapy Kohonena	81
5.4. Miary oceny wyników podziału danych na grupy	83
6. Identyfikacja i profilowanie wzorców wypadków drogowych	87
6.1. Identyfikacja wzorców wypadków drogowych z udziałem jednego pojazdu	93
6.1.1. Ogólna charakterystyka wzorców wypadków drogowych z udziałem jednego pojazdu	96
6.1.2. Profile zagrożeń dla wypadków drogowych z udziałem jednego pojazdu	100
6.2. Identyfikacja wzorców wypadków drogowych z udziałem co najmniej dwóch pojazdów	107
6.2.1. Ogólna charakterystyka wzorców wypadków drogowych z udziałem co najmniej dwóch pojazdów	111
6.2.2. Profile zagrożeń dla wypadków drogowych z udziałem co najmniej dwóch pojazdów	115

6.3. Różnice we wzorcach wypadków drogowych z uwagi na liczbę uczestniczących pojazdów i statusu zdarzenia	123
7. Modele logistyczne w identyfikacji zagrożeń w ruchu drogowym	127
7.1. Regresja logistyczna dwuwartościowa	127
7.2. Modele logitowe z kategorią bazową	129
7.3. Logistyczne modele porządkowe	130
7.4. Dobór modelu do klasyfikacji zagrożeń na drodze	133
7.5. Rozwiązywanie problemów w budowaniu modeli logistycznych do klasyfikowania cech zagrożenia brd	135
7.6. Budowanie i ocena modeli logistycznych	139
7.7. Rodzina modeli i jakość klasyfikacji	141
7.8. Prezentacja modeli	147
8. Identyfikacja cech zagrożenia brd dla wypadku drogowego z udziałem jednego pojazdu ...	149
8.1. Identyfikacja zachowania kierującego sprawcy dla wypadku drogowego z udziałem jednego pojazdu – dwuwartościowe klasyfikatory logistyczne	153
8.2. Identyfikacja rodzaju zdarzenia dla wypadku drogowego z udziałem jednego pojazdu – klasyfikatory logitowe z kategorią bazową	162
8.3. Identyfikacja statusu zdarzenia dla wypadku drogowego z udziałem jednego pojazdu – logistyczne klasyfikatory porządkowe	168
8.3.1. Status zdarzenia drogowego w modelu proporcjonalnych szans	169
8.3.2. Status zdarzenia drogowego w logitach kontynuowanych	174
8.4. Wykorzystanie rodziny modeli do klasyfikacji zagrożeń brd dla wypadków drogowych z udziałem jednego pojazdu	182
9. Identyfikacja cech zagrożenia brd dla wypadku drogowego z udziałem co najmniej dwóch pojazdów	190
9.1. Identyfikacja zachowania kierującego sprawcy dla wypadku drogowego z udziałem co najmniej dwóch pojazdów – klasyfikatory logitowe z kategorią bazową	194
9.2. Identyfikacja rodzaju zdarzenia dla wypadku drogowego z udziałem co najmniej dwóch pojazdów – klasyfikatory logitowe z kategorią bazową	201
9.3. Identyfikacja statusu zdarzenia dla wypadku drogowego z udziałem co najmniej dwóch pojazdów – logistyczne klasyfikatory porządkowe	208
9.3.1. Status zdarzenia drogowego w modelach proporcjonalnych szans	208
9.3.2. Status zdarzenia drogowego w logitach kontynuowanych	212
9.4. Wykorzystanie rodziny modeli do klasyfikacji zagrożeń brd dla wypadków drogowych z udziałem co najmniej dwóch pojazdów	220
10. Dyskusja wyników modelowania dla wypadków drogowych bez udziału pieszych na drogach zamiejskich	228
Podsumowanie i wnioski	243
Bibliografia	255
Summary	265
Załączniki	267