

UDK: 614.8:519.237.8

PRIMENA K-MEANS METODE U DUBINSKOJ ANALIZI SAOBRAĆAJNIH NEZGODA

USING K-MEANS CLUSTERING METHOD FOR IN-DEPTH ANALYSIS OF TRAFFIC ACCIDENT

Dalibor Pešić¹, Milica Šelmić², Dragana Macura³ i Nenad Marković⁴

Rezime: Preventivno delovanje u bezbednosti saobraćaja podrazumeva kvalitetno prikupljanje podataka o faktorima koji utiču na nastanak i posledice saobraćajnih nezgoda. U praksi najbolje prikupljanje ovih podataka omogućava metod dubinskih analiza saobraćajnih nezgoda, na koji način se analizira uticaj puta i putne okoline, vozila i čoveka na saobraćajnu nezgodu. Kako je ovaj metod izuzetno složen i prevashodno skup za sprovođenje na širem području, to je neophodno razviti model koji bi prilagodio i pojednostavio sprovođenje dubinskih analiza u praksi, a na osnovu već postojećih podataka prikupljenih veštačenjem saobraćajnih nezgoda. U radu je razvijen model za grupisanje uzročnika saobraćajnih nezgoda na osnovu sličnosti karakteristika vozača, vozila i puta. Razvijeni model je zasnovan na metodu klasterovanja. Klasterovanje je proces klasifikacije objekata u različite grupe, skupove, tako da elementi jednog skupa imaju izvestan stepen približnosti i jednakosti meren po unapred definisanim kriterijumima. Sličnost elemenata se utvrđuje na osnovu njihovih rastojanja. Dva elementa su sličnija ukoliko se nalaze na bližem rastojanju. Najviše korišćena metoda klasterovanja je metod K-means, koji je ujedno korišćen kao alat u ovom radu.

Ključne riječi: klaster analiza, veštačenje, saobraćajne nezgode, faktori

Abstract: Preventive actions in the field of road traffic safety implies quality collection of data related to factors which are contribute to traffic accident occurrence and its consequences. In practice, the best collection of these data is provided by using in-depth studies of traffic accidents, where influence of road, environment, vehicle and man (driver) is being analysed. Whereas this model is significantly complex and expensive for wide implementation, it is necessary to create a model capable to simplify and adapt in-depth studies process by using data already collected through traffic accident expertises. In this paper is developed model for grouping – clustering of the causes of traffic accidents, based on similarities of driver, vehicle and road. Clustering is process where different objects are classified in groups and sets in way to be equal or similar according to predefined criteria. Similarity of the elements is determined according to the „distance“ between them and two elements are more similar if the „distance“ is shorter. One of the most applied clustering methods is K-means and this method is explained in this paper.

Keywords: Cluster analysis, Traffic accident expertise, Traffic accident, Impact factors

1. UVOD

Analizama saobraćajnih nezgoda može se doći do odgovarajućih zaključaka koji stvaraju mogućnost upravljanja stanjem bezbednosti saobraćaja. U zavisnosti od toga koji nivo analiza je neophodan upravljaču, moguće je sprovoditi različite vrste analiza saobraćajnih nezgoda. U naučnoj i stručnoj praksi poznate su analize saobraćajnih nezgoda koje se mogu podeliti u dve grupe: fenomenološke i etiološke. Fenomenološke analize saobraćajnih nezgoda imaju zadatak da utvrde globalne probleme bezbednosti saobraćaja i najčešće se sprovode korišćenjem tzv. zbirnih podataka o saobraćajnim nezgodama. Sa druge

¹ dr Dalibor Pešić, dipl. inženjer saobraćaja, docent, Saobraćajni fakultet u Beogradu, Vojvode Stepe 305, 11000 Beograd, Srbija, e-mail: d.pesic@sf.bg.ac.rs

² dr Milica Šelmić, dipl. inženjer saobraćaja, docent, Saobraćajni fakultet u Beogradu, Vojvode Stepe 305, 11000 Beograd, Srbija, e-mail: m.selmic@sf.bg.ac.rs

³ dr Dragana Macura, dipl. inženjer saobraćaja, docent, Saobraćajni fakultet u Beogradu, Vojvode Stepe 305, 11000 Beograd, Srbija, e-mail: d.macura@sf.bg.ac.rs

⁴ MSc Nenad Marković, dipl. inženjer saobraćaja, asistent, Saobraćajni fakultet u Beogradu, Vojvode Stepe 305, 11000 Beograd, Srbija, e-mail: n.markovic@sf.bg.ac.rs

strane, etiološka analiza otvara tzv. crnu kutiju i omogućava da se utvrde uzroci saobraćajnih nezgoda. Sa preciznim poznavanjem uzroka saobraćajnih nezgoda moguće je definisati odgovarajući set mera za unapređenje stanja bezbednosti saobraćaja. U okviru etioloških analiza saobraćajnih nezgoda prepoznaju se najčešće: ekspertize, odnosno veštačenja saobraćajnih nezgoda i dubinske analize saobraćajnih nezgoda, odnosno nezavisne ocene saobraćajnih nezgoda.

Detaljnou analizom podataka iz etioloških istraživanja saobraćajnih nezgoda mogu se pouzdano utvrditi uzroci i okolnosti koji su doveli do nastanka svake konkretne saobraćajne nezgode. Analizom ovako prikupljenih podataka moguće je prepoznati najvažnija obeležja saobraćajnih nezgoda koja ih izdvajaju ili poistovećuju sa drugim nezgodama. Prepoznavanje ovakvih faktora omogućava usmeravanje dalje analize na konkretnu oblast ili konkretni faktor, kako bi se on istražio u svakom sledećem slučaju i utvrdio pouzdan metod za njegovu eliminaciju u budućnosti.

U radu je razvijen model za grupisanje uzročnika saobraćajnih nezgoda na osnovu sličnosti karakteristika vozača, vozila i puta. Razvijeni model je zasnovan na metodu klasterovanja. Klasterovanje je proces klasifikacije objekata u različite grupe, skupove, tako da elementi jednog skupa imaju izvestan stepen približnosti i jednakosti meren po unapred definisanim kriterijumima. Sličnost elemenata se utvrđuje na osnovu njihovih rastojanja. Dva elementa su sličnija ukoliko se nalaze na bližem rastojanju. U dosadašnjoj stručnoj i naučnoj praksi u oblasti bezbednosti saobraćaja klasterovanje je primenjivano u mnogo slučajeva (Fleury and Brenac, 2001; Casaer et al, 2004; Ma and Kockelman, 2006; Habibović et al, 2013; de Ona et al., 2013; Fernades and Neves, 2013; Bil et al, 2013; Lenard et al, 2014; Huang and Wong, 2015; Amirian et al., 2015). Međutim, najviše korišćena metoda klasterovanja je metod *K-means* (Teodorović et al., 2010 i Šelmić et al., 2012), koji je u ovom radu predstavljen kao jedan od modela klasterovanja sa mogućnošću primene na analize saobraćajnih nezgoda, posebno kao pomoć pri preliminarnim analizama u procesu dubinskih analiza saobraćajnih nezgoda.

2. KLASTEROVANJE – K-MEANS ALGORITAM

Klasterovanje je metod klasifikacije objekata u različite grupe, skupove, tako da elementi jednog skupa imaju izvestan stepen približnosti i jednakosti meren po unapred definisanim kriterijumima. Sličnost elemenata se utvrđuje na osnovu rastojanja. Dva elementa su sličnija ukoliko se nalaze na bližem rastojanju. Postoje mnogobrojne metode za računanje rastojanja, kao što su: Euklidsko, Menhetn, Maksimalno rastojanje, itd. Najviše korišćena metoda klasterovanja je metod *K-means* koju je razvio Hartigan 1975. godine. Četiri godine kasnije metod je dopunjen doprinosom autora Hartigana i Vonga.

K-means algoritam dodeljuje element grupi, čiji je centar najbliži posmatranom elementu. Rastojanje, odnosno udaljenost, elementa od drugih elemenata može da se računa različitim metodama, kao što je već spomenuto. U ovom radu korišćeno je Euklidsko rastojanje. Konačno rešenje se prikazuje binarnom matricom U . Element u_{ij} matrice U je jednak 1 ukoliko je j -ti element dodeljen i -toj grupi. Vrednost elementa je jednaka 0 u suprotnom slučaju, tj.:

$$u_{ij} = \begin{cases} 1, & \text{ako je } \|x_j - c_i\|^2 \leq \|x_j - c_k\|^2, \text{ za } \forall k \neq i \\ 0, & \text{u ostalim slučajevima} \end{cases} \quad (1)$$

gde je:

x_j – koordinate (karakteristike) j -tog elementa

c_i – centar posmatrane i -te grupe;

Funkcija cilja (2) koju treba minimizirati predstavlja zbir rastojanja svih elemenata od centara grupe kojoj su dodeljeni. Matematička formulacija problema klasterovanja glasi:

Minimizirati

$$F = \sum_{i=1}^c F_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \quad (2)$$

pri ograničenjima:

$$\sum_{i=1}^c u_{ij} = 1, \text{ za } \forall j=1, 2, \dots, n \quad (3)$$

$$\sum_{i=1}^c \sum_{j=1}^n u_{ij} = n \quad (4)$$

где је:

F_i – вредност критеријумске функције у оквиру i -те групе;

x_k – координате (карактеристике) k -тог елемента;

c_i – центар посматране i -те групе;

c – број група ($i=1,2,\dots,c$);

G_i – i -та група.

Ограничење представљено релацијом (3) показује да један елемент може да припада само једној групи, док се ограничењем (4) проверава да ли су сви елементи распоређени по групама. Поčetni центар сваке групе бира се на случајан начин. Након сваке завршене итерације алгоритма бира се нови центар кластера и врши се поновно придруживање елемената. Нови центар, c_i , за сваку групу се дефинише тако што се вредности претходно дефинисаних елемената, као чланова i -те групе, сумирају и деле са бројем чланова саме групе, тј.:

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \quad (5)$$

при ограничењу:

$$|G_i| = \sum_{j=1}^n u_{ij} \quad (6)$$

где је:

$|G_i|$ – укупан број елемената i -те групе

n – број елемената које треба груписати

Ограничење (6) обезбеђује да се не прекораči максимални дозвољени број елемената у свакој групи. Неопходни кораци приликом решавања посматраног модела су (Šelmić et al, 2012):

- Дефинисати елементе које треба груписати, n ;
- Изабрати број група, c ;
- Изабрати број елемената сваке групе, G (овaj корак може да се изостави, уколико је број чланова група неограничен);
- Одредити c центара група;
- Одредити најближи центар за сваку групу;
- Одредити нове центре;

Понављати претходне кораке, док се не постигне неки унапред дефинисан критеријум. Критеријуми заустављања могу бити: унапред задати број итерација, нова вредност функције циља је већа од претходне; нова вредност функције циља је променјена у односу на претходну за мање од 20%.

Основна предност овог модела јесте његова једноставност и брзина обраде великог броја података. Неки од недостатака су: зависност квалитета излазног решења од случајног избора почетних, улазних, вредности центара чворова; атрибути су искључиво нумеричке вредности; израчунати локални optimum може да буде веома различит од глобалног optima; и др. Примена овог модела не гарантује изналажење optimalног решења (Teodorovic et al, 2010).

3. PRIMENA METODE K-MEANS ZA GRUPISANJE UZROČNIKA SAOBRAĆAJNIH NEZGODA NA OSNOVU SLIČNOSTI KARAKTERISTIKA VOZAČA, VOZILA I PUTA

У примеру који је развијен у овом раду, посматран је случај када су учествовала два возила у саобраћајној незгоди. Сваки учесник саобраћајне незгоде је окарактерисан са следећих осам карактеристика:

- време и дан када се незгода догодила;
- стање пута;

- poravnanje puta;
- pol i godine vozača;
- brzina;
- vidljivost;
- površina kolovoza;
- kategorija puta.

Алгоритам за груписање узрочника саобраћајних незгода се састоји из следећих корака:

Korak 1. За задати број итерација, број узорака незгода и величине група генерисати улазне податке (координате) за све незгоде које се анализирају.

Korak 2. На случајан начин генерисати центре група. Укупан број кластера (група) је 15 и то у складу са узроцима незгоде. Под узроцима се подразумевају фактори (пропусти) који су утицали на настанак саобраћајне незгоде, односно који су је проузроковали, док су под доприносом сматрани фактори (пропусти) који нису директно проузроковали настанак саобраћајне незгоде, али су допринели последицама настале незгоде. Наиме, доприноси се огледају у повећаним последицама саобраћајне незгоде у односу на последице које би биле да није било ових околности.

Мogućност избегавања саобраћајних незгода представља пропуст nekог од наведених фактора (најчешће везана за фактор човек) где је начинjen пропуст не предузимањем одређених радњи у циљу избегавања саобраћајне незгоде. Пропusti везани за тежину последица су они пропусти који немају везу са настaнком или могућношћу избегавања настaнка саобраћајне незгоде, али су допринели већим последицама незгоде (већа оштећења и/или теже повреде).

Korak 3. Израчунати Еуклидска растојања свих незгода од свих генерисаних центара. Формирати матрицу растојања у којој врсте представљају незгоде, а колоне центре кластера.

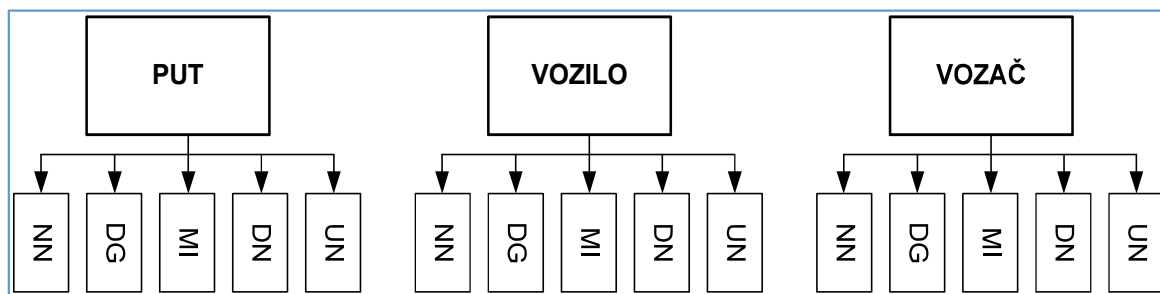
Korak 4. Пронаћи најмањи елемент у матрици растојања. Одређена незгода којој одговара вредност најмањег растојања додељује се одговарајућем кластеру и искључује се из даљег разматрања. Ажурирати матрицу растојања и понављати поступак све док се све незгоде не доделе кластерима (сврстају по узроцима саобраћајних незгода). С обзиром да не постоји ограничење капацитета кластера свака незгода се може сврстати у кластер коме припада.

Korak 5. По додељивању свих незгода кластерима израчунати вредност критеријумске функције.

Korak 6. Одредити нове центре група користећи релације (5) и (6). Вратити се на корак 2.

Korak 7. Понављати алгоритамске кораке док се не достигне унапред задати број итерација.

Применом изложеног алгоритма различити утицаји пута, возила и вожача у појединачној саобраћајној незгоди се могу класификовати, а са циљем обављања дубинске анализе. Сваки од ових фактора може учествовати у незгоди као: узрок незгоде (UN), допринос појави незгоде (DN), могућност избегавања (MI), допринос настaнку губитака (DG) и ниједан од наведених (NN).



Слика 1. 1- 15 могућих кластера

Validacija ostvarenih rezultata se vrši primenom Silhouette tehnike (Rousseeuw 1987). Silhouette je metoda koja se koristi za interpretaciju klasterovanih podataka i omogućava grafički prikaz kako se svaki element (nezgoda) uklapa u klaster (uzrok nezgode) kome je dodeljen. Na y osi bi bili predstavljeni klasteri, a na x osi vrednost Silhouette metode. Ova vrednost može da bude u rasponu od -1 do +1. Ukoliko je vrednost bliža +1 to znači da je nezgoda dobro klasifikovana, dok suprotno ukoliko je Silhouette vrednost bliža -1 to znači da klasterovanje nije precizno urađeno.

4. ZAKLJUČAK

Uzimajući u obzir negativne posledice saobraćaja, a posebno nedopustivo veliko stradanje ljudi u saobraćajnim nezgodama, jasno je da je uspešno upravljanje stanjem bezbednosti saobraćaja izuzetno značajan društveni proces. Brojna su istraživanja koja su dokumentovala značaj strateškog upravljanja bezbednošću saobraćaja, kao i primenu sofisticiranih naučnih metoda u rešavanju ključnih problema bezbednosti saobraćaja. Posebno je važno izdvojiti alate koji se sa jedne strane bave otkrivanjem i analizom uzroka nastanka saobraćajnih nezgoda, jer se neutralisanjem uzroka na najjednostavniji i najekonomičniji način sprečavaju najteže posledice saobraćajnih nezgoda, ali i alate koji služe za rešavanje već uočenih uzroka – problema.

Jedan od međunarodno priznatih i u većini zemalja zakonom propisanih alata za unapređenje bezbednosti saobraćaja jesu i dubinske analize saobraćajnih nezgoda (metod nezavisne ocene) koje su bile predmet ovog rada. S obzirom na to da je postupak realizacije dubinskih analiza saobraćajnih nezgoda dugotrajan i veoma složen, jer zahteva prikupljanje i analizu velikog broja podataka, u ovom radu su predstavljene mogućnosti izdvajanja i grupisanja podataka iz ekspertiza saobraćajnih nezgoda, primenom K-means algoritma. Ekspertize saobraćajnih nezgoda predstavljaju riznicu podataka vezanih za nastanak konkretne saobraćajne nezgode, pa se korišćenjem tih, već prikupljenih, podataka za potrebe dubinskih analiza saobraćajnih nezgoda ostvaruje niz koristi i postupak pojednostavljuje. Pored toga, najvažnija prednost primene K-means algoritma je što se na jednostavan i brz način može obraditi velika količina podataka čime se trajanje postupka dubinske analize i utvrđivanja uzroka nastanka nezgode bitno skraćuje.

Prethodno iznete prednosti i dati zaključci ukazuju na značaj opisanog metoda za unapređenje bezbednosti saobraćaja u lokalnim zajednicama. Posebno je važno istaći da samouprave, a posebno donosioci odluka u subjektima značajnim za upravljanje stanjem bezbednosti saobraćaja, primenom ovog metoda imaju značajnu podršku u odlučivanju. Naime, rezultati se dobijaju brzo, pa se u slučaju uočenog problema u bezbednosti saobraćaja relativno brzo mogu otkriti uzroci tog problema, a potom i realizovati one mere koje na najbolji način neutrališu uočene uzroke. Pored toga, primenom opisanog metoda je moguće po hijerarhiji kreirati listu uzroka, odnosno listu problema i na taj način alocirati sredstva predviđena za unapređenje bezbednosti saobraćaja u lokalnoj zajednici.

Uzimajući u obzir prednosti dubinskih analiza saobraćajnih nezgoda, kao i sposobnosti K-means algoritma, u budućnosti bi podaci iz ekspertiza saobraćajnih nezgoda mogli biti dopunjeni i podacima iz osiguravajućih kompanija, zdravstvenih ustanova, servisnih ustanova i drugih subjekata koji raspolazu podacima od značaja za utvrđivanje uzroka (i okolnosti) saobraćajnih nezgoda, čime bi se još više unapredio proces upravljanja bezbednošću saobraćaja na konkretnom području, jer bi doneti zaključci bili potkrepljeni analizom većeg broja podataka.

5. LITERATURA

- [1]. Amirian, E., Leung, J.Y., Zanon, S., Dzurman, P. 2015. Integrated cluster analysis and artificial neural network modeling for
- [2]. steam-assisted gravity drainage performance prediction in heterogeneous reservoirs, *Expert Systems with Applications* 42, 723–740
- [3]. Bíl, M., Andr, R., ek Janoška, Z. 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation, *Accident Analysis and Prevention* 55, 265–273
- [4]. Casaer, F., Geert, W., Isabelle, T. 2004. Road Traffic Accident Clustering With Categorical Attributes, *TRB 2004 Annual Meeting*
- [5]. de Ona, J., Lpez, G., Mujalli, R., Calvo, F.J. 2013. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks, *Accident Analysis and Prevention* 51, 1– 10
- [6]. Fernandes, A., Neves, J. 2013. An approach to accidents modeling based on compounds road environments, *Accident Analysis and Prevention* 53, 39– 45
- [7]. Fleury, D., Brenac, T. 2001. Accident prototypical scenarios, a tool for road safety research and diagnostic studies, *Accident Analysis and Prevention* 33, 267 – 276
- [8]. Habibovica, A., Tivesten, E., Uchidab, N., Brgman, J., Ljung Aust, M. 2013. Driver behavior in car-to-pedestrian incidents: An application of the Driving Reliability and Error Analysis Method (DREAM), *Accident Analysis and Prevention* 50, 554– 565

- [9]. Hartigan, J. A., Clustering Algorithms. Wiley, 1975.
- [10]. Hartigan, J. A., Wong, M. A., Algorithm AS 136: A K-means Clustering Algorithm. Journal of the Royal Statistical Society, Series C (Applied Statistics) 28 (1), 100-108, 1979.
- [11]. Huang, S.H., Wong, J.T. 2015. A multinomial choice model approach for dynamic driver vision transitions, Accident Analysis and Prevention 74, 107–117
- [12]. Lenard, J., Badea-Romero, A., Danton, R. 2014. Typical pedestrian accident scenarios for the development of autonomous emergency braking test protocols, Accident Analysis and Prevention 73, 73–80
- [13]. Ma, J., Kockelman, K. 2006. Crash Frequency and Severity Modeling Using Clustered Data from Washington State, IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September 17-20
- [14]. Rousseeuw, P., "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics, Vol. 20, 53–65, 1987.
- [15]. Teodorović D., Macura D., Šelmić M. 2010 Povezivanje zajedničkih vožnji primenom metode k-means, Simpozijum o operacionim istraživanjima SYM-OP-IS.
- [16]. Šelmić, M., Macura, D., and Teodorović, D. (2012). "Ride Matching Using K -means Method: Case Study of Gazela Bridge in Belgrade, Serbia." J. Transp. Eng., 138(1), 132–140.