

Identifikácia typu streliva pomocou zhlukovej analýzy

Anotácia: V článku analyzujeme dáta z [1], ktoré tvoria vzorky štyroch typov povýstrelových splodín. Úlohou je na základe chemického zloženia zvyškov po výstrele identifikovať typ streliva. Využitím pokročilých metód v zhlukovej analýze sme správne klasifikovali 2 typy streliva, zvyšné 2 nemožno zo získaných dát správne identifikovať.

Kľúčové slová: typ streliva, povýstrelové splodiny, štatistické metódy.

Úvod

Správne určenie typu streliva, ktoré útočník použil na streľbu, môže byť dôležitý krok pri vyšetrowaní trestného činu. V situáciách, keď páchatel zanechá na mieste činu príliš veľa stôp, nie je ťažké určiť typ či konkrétny model zbrane, z akej sa strieľalo. Sú prípady, keď na mieste skutku zostane len stopové množstvo povýstrelových splodín (gunshot residues – GSR). Chemickou expertízou je možné tento typ stopy analyzovať tak, že je možné zistiť prítomnosť určitých chemických prvkov, ktoré sa môžu nachádzať vo vzorke povýstrelových splodín ako jeho prirodzená súčasť.

Avšak aj keď poznáme chemické zloženie konkrétnej vzorky, je veľmi ťažké určiť typ streliva, z ktorého môže pochádzať, keďže obsahy prvkov a ich zlúčenín bývajú v mnohých typoch streliva veľmi podobné.

Zadora spolu s kolegami preto skúmali, či možno nástrojmi matematickej štatistiky rozlíšiť typy streliva, ktorými mohol strelec strieľať. S týmto cieľom vykonali nasledujúci experiment. Štyrmi typmi streliva (Browning 7.65 mm, Luger 9 mm, Makarov 9 mm a Tokarev 7.62 mm) vystrelili jeden až trikrát, čím získali celkovo 30 vzoriek povýstrelových splodín (pre typ Browning vzorky B1-B3, pre typ Luger L1-L14, pre typ Makarov M1-M7 a pre typ Tokarev T1-T6) odobraných z rúk strelca ihneď po výstreloch.

Každá z 30 vzoriek bola následne analyzovaná metódou rastrovacej (skenovacej) elektrónovej mikroskopie s energiovo-disperzným röntgenovým fluorescenčným spektrometrom (SEM – EDX) v automatickom režime na účel namerania obsahov niektorých častíc prítomných v povýstrelových splodinách. Týmto časticami boli zlúčeniny olova, antimónu a bária prítomné v látke. V každej vzorke bolo zmerané relatívne zastúpenie siedmich skupín častíc povýstrelových splodín – označených ako PbSbBa, PbBa, PbSb, SbBa, Pb, Sb a Ba. Pôvodný dataset možno nájsť¹.

Po chemickej analýze boli dáta ďalej analyzované štatistickými metódami, konkrétne Mann-Whitneyho testom a hierarchickým zhlukovaním. Mann-Whitneyho test mal ukázať, ktoré

¹ BROZEK-MUCHA, Z., ZADORA, G. (2003) Grouping of ammunition types by means of frequencies of occurrence of GSR, *Forensic Science International* 135, 97-104.

premenné dokážu najlepšie rozlíšiť jednotlivé typy streliva. Následná zhluková analýza ukázala, že jednoznačne možno rozlíšiť typ Browning a Luger od všetkých ostatných typov.

V článku analyzujeme rovnaké dáta pomocou korelačnej analýzy a následne aplikujeme na dáta neurónovú sieť, ktorá dokáže identifikovať zhluky v dátach efektívnejším spôsobom ako hierarchické zhlukovanie.

Metodika

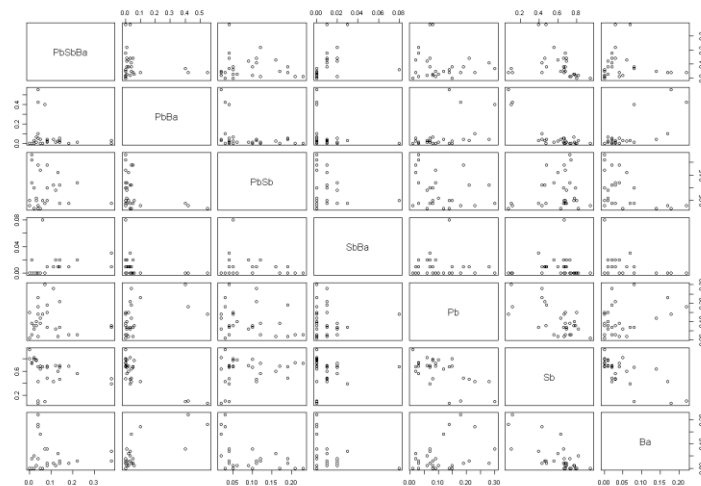
Dáta sme najskôr načítali do štatistického open-source programu R. Pred samotnou viacrozmernou štatistickou analýzou dát treba dáta najskôr predspracovať. Vzhľadom na to, že premenné (v našom prípade 7 skupín častíc povýstrelových splodín) sú vo forme relatívnych frekvencií f_i , existuje medzi nimi lineárny vzťah. Platí totiž, že $f_1 + f_2 + \dots + f_6 = 1$.

Vypočítali sme si preto Pearsonovu korelačnú maticu premenných (tab. 1) a zobrazili scatterploty všetkých dvojíc (obr. 1). Pearsonova korelácia je mierou lineárnej závislosti medzi dvojicou premenných. Môže nadobúdať hodnoty z intervalu $[-1,1]$. Hodnoty blízke 1 znamenajú silnú priamu lineárnu závislosť, naopak hodnoty blízke -1 silnú nepriamu lineárnu závislosť. Hodnoty blízke nule svedčia o nulovej lineárnej závislosti medzi premennými. Najsilnejšia korelácia je medzi premennými *Sb* a *PbBa* (-0.8228105). Jednu z premenných budeme musieť vyradiť. Na tento účel sme zobrazili boxploty všetkých 7 premenných (obr. 2). Najväčšiu variabilitu dosahuje premenná *Sb*, všetky ostatné premenné majú podobné stredné hodnoty a rozptyly. S cieľom odstrániť silnú závislosť medzi premennými preto odstránime práve premennú *Sb*.

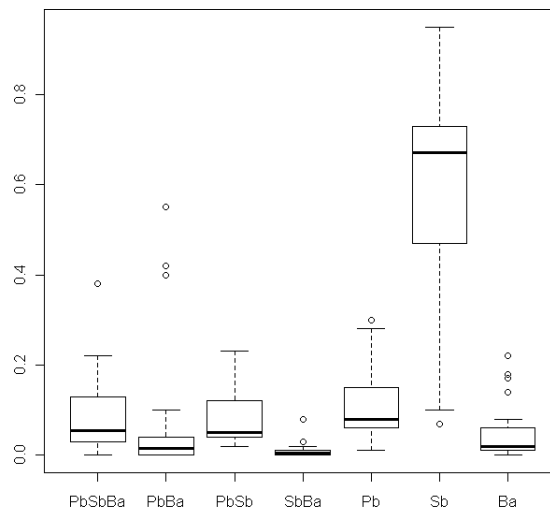
Zostávajúcich 6 premenných musíme teraz štandardizovať, keďže ich relatívne zastúpenia nemajú rovnaké rozdelenie. V takomto prípade by najväčšiu úlohu pri hľadaní zhlukov zohrávali tie premenné, ktoré majú najväčší rozptyl. Ak ale dáta štandardizujeme, všetky premenné budú mať v modeli rovnakú dôležitosť. Samotná štandardizácia pozostáva z dvoch krokov. V prvom odpočítame od každej premennej odhad jej strednej hodnoty (teda aritmetický priemer každého stĺpca) a následne vydáme každú premennú odhadom jej smerodajnej odchýlky. Dekorelované a štandardizované dáta sú pripravené na ďalšiu analýzu.

	PbSbBa	PbBa	PbSb	SbBa	Pb	Sb	Ba
PbSbBa	1.00000000	-0.1466755	-0.17328666	0.31735106	-0.10375476	-0.2800323	-0.03285642
PbBa	-0.14667551	1.00000000	-0.32318870	-0.22450360	0.46099295	-0.8228105	0.73580603
PbSb	-0.17328666	-0.3231887	1.00000000	-0.09733349	-0.24292850	0.1972818	-0.38722816
SbBa	0.31735106	-0.2245036	-0.09733349	1.00000000	-0.01080449	0.0239114	-0.21264426
Pb	-0.10375476	0.4609929	-0.24292850	-0.01080449	1.00000000	-0.6232143	0.35061148
Sb	-0.28003229	-0.8228105	0.19728183	0.02391140	-0.62321432	1.00000000	-0.71241233
Ba	-0.03285642	0.7358060	-0.38722816	-0.21264426	0.35061148	-0.7124123	1.00000000

Tab. 1: Pearsonova korelačná matica



Obr. 1: Scatterploty všetkých dvojíc premenných



Obr. 2: Box-Whiskerove grafy pre všetky premenné v modeli

Zhluková analýza

Zhluková analýza je jednou z najpoužívanějších štatistických metód, používa sa v situáciách, keď objekty v datasúbore vytvárajú zhluky. Zhlukom sa rozumie množina podobných objektov na základe určitých vlastností, ktoré sme na každom objekte namerali. Podobnosť je v matematickom zmysle slova definovaná ako blízkosť daných objektov v ich priestore. Každý objekt si možno predstaviť ako bod v priestore a objekty s veľmi podobnými vlastnosťami si budú ako body veľmi blízke.

Samoorganizujúce sa mapy

Samoorganizujúce sa mapy (skrátene SOM)² sú typom umelej neurónovej siete bez učiteľa (teda príslušnosť objektov k zhlukom nemusíme zadať vopred, hovoríme, že sieť sa učí sama z dát). Navrhol ich v roku 1982 Fín Teuvo Kohonen. Ich hlavný význam spočíva v zobrazení viacrozmerného priestoru (najčastejšie viac ako trojrozmerného, teda takého, ktorý sa nedá priamo vizualizovať) do priestoru s nižšou dimenziou (najčastejšie dvojrozmerného). Takéto zobrazenie zachováva lokálnu topológiu, teda objekty, ktoré sú si blízke v priestore s vyššou dimenziou, si budú blízke aj v priestore s nižšou dimenziou (týmto priestorom je spomínaná Kohonenova mapa).

Pôvodná myšlienka vizualizovať dáta sa neskôr použila na zhlukovanie objektov. Vďaka zachovávanju sa topológii môžeme pomocou Kohonenovej mapy identifikovať, ktoré objekty sú si blízke (teda podobné). Výhoda SOM oproti iným zhlukovacím metódam je najmä to, že nepotrebujú žiadne parametre, akými sú napríklad počet zhlukov. Nevýhodou je najmä ťažšia interpretácia výsledkov zo samotnej mapy. SOM preto vyžadujú omnoho viac času na pochopenie a realizáciu v počítačovom programe.

Algoritmus, akým sa mapa organizuje, možno popísať takto: Používateľ najskôr zvolí veľkosť mapy, ktorú predstavuje počet neurónov, a typ mapy. Štatistické balíky pritom najčastejšie ponúkajú list (na dvojrozmernú vizualizáciu), toroid alebo cylinder (oba na trojrozmernú vizualizáciu). Ďalej treba nastaviť typ okolia, teda ktoré neuróny má algoritmus považovať za blízke danému neurónu. Rozšírené je najmä šesťuholníkové a pravouhlé okolie.

Hlavným vizualizačným výstupom algoritmu SOM je tzv. U-matica (obr. 3), ktorá slúži na identifikáciu počtu zhlukov. Vznikne tak, že k pôvodnej mape, ktorej veľkosť si určíme na začiatku, sa pridávajú ďalšie bunky, a to tak, že medzi každými dvomi susednými neurónmi pribudne bunka, ktorej bude zodpovedať vzdialenosť medzi týmito neurónmi (v zmysle vzdialeností ich váhových vektorov na konci algoritmu). Farba buniek U-maticy teda odzrkadľuje vzdialenosti medzi neurónmi.

Hľadanie zhlukov v dátach pomocou samoorganizujúcej sa mapy

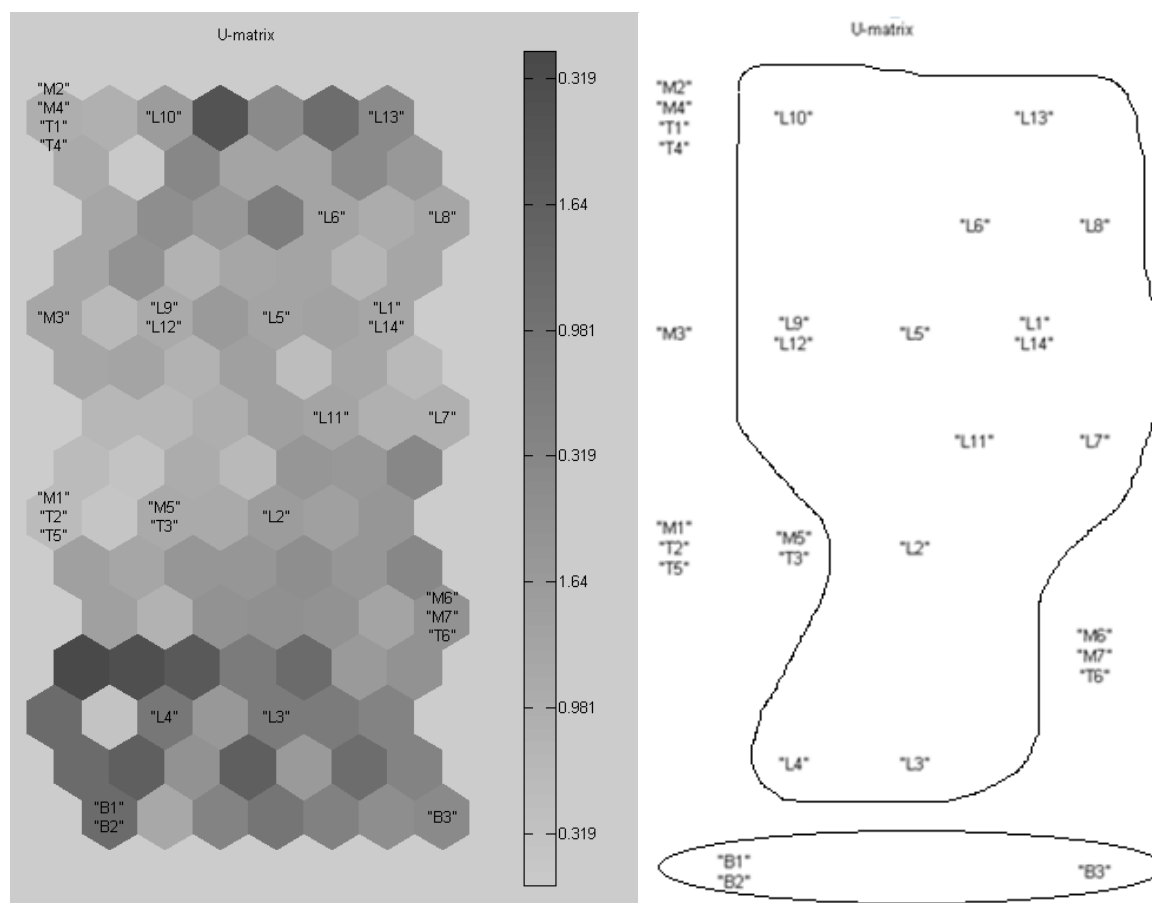
Dekorelované a štandardizované dáta sme načítali do programu Matlab. V porovnaní s programom R totiž poskytuje lepšie vizualizačné možnosti pri práci so samoorganizujúcimi sa mapami (nutné doinštalovať si balík SOM toolbox). Na obrázku 3 možno vidieť U-maticu vytvorenú z našich dát (písmená zodpovedajú typom streliva a čísla vzorkám).

V zmysle algoritmu samoorganizujúcich sa máp by sa mali podobné objekty (u nás tie isté typy streliva) zobrazovať blízko pri sebe na výslednú U-maticu (za predpokladu, že sa dajú rozlíšiť pomocou chemického zloženia povýstrelových splodín (GSR)). V spodnej časti možno vidieť všetky 3 vzorky streliva Browning, ktoré ležia oddelene od všetkých iných vzoriek (tento

² KOHONEN, T. Self-organized formation of topologically correct feature maps, IN *Biological Cybernetics* 43.

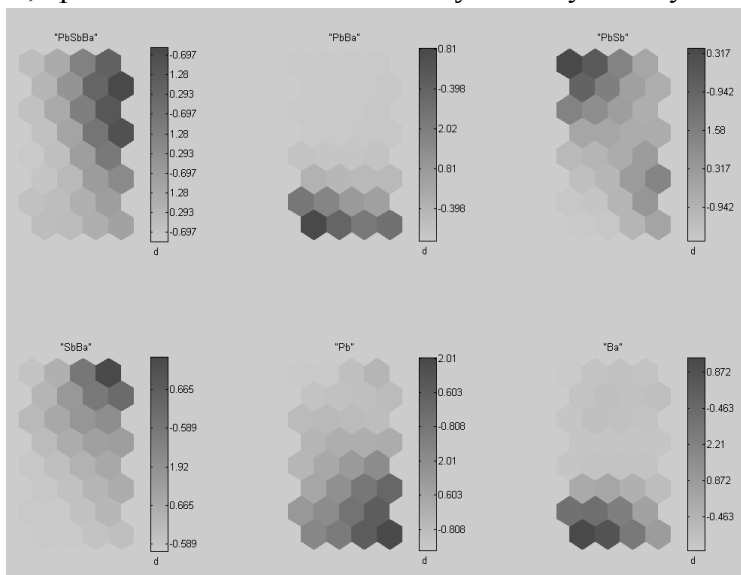
výsledok potvrdzuje záver publikovaný v ³. Vo vnútornej časti U-matice sa nachádza veľký zhhluk vzoriek povýstrelových splodín (GSR) pochádzajúcich zo streliva Luger. Všetky vzorky sa projektovali blízko k sebe, čo značí, že tento typ streliva sa dá tiež dobre identifikovať. Pomocou hierarchického zhlukovania dostali v ³ mierne odlišné výsledky (niektoré vzorky povýstrelových splodín GSR Luger sa kryli so vzorkami z iných typov povýstrelových splodín (GSR)). Všetky zvyšné objekty na mape sú vzorky povýstrelových splodín (GSR) pochádzajúcich z typu streliva Makarov a Tokarev. Tieto vzorky nie sú dobre separované, keďže sa na mape prekrývajú. Z toho usudzujeme, že tieto dva typy povýstrelových splodín (GSR) nemôžeme spoľahlivo rozlíšiť.

Na obrázku 4 sú matice jednotlivých premenných použitých v modeli. Farby odzrkadľujú hodnoty danej premennej, ktorú nadobúdajú všetky vzorky, ktoré sa zobrazili na daný neurón. Napríklad z grafu premennej *PbBa* vidíme, že dosahuje najväčšie hodnoty v spodnej časti mapy, do ktorej sa projektovali 3 vzorky pochádzajúce z typu streliva Browning. To znamená, že najdôležitejšou premennou na identifikáciu tohto typu streliva je práve *PbBa*.



³ BROZEK-MUCHA, Z., ZADORA, G. (2003) Grouping of ammunition types by means of frequencies of occurrence of GSR, *Forensic Science International* 135, 97-104.

Obr. 3: U-matica (vľavo U-matica zobrazujúca rozdiely medzi neurónmi vo forme farebných odtieňov, vpravo tá istá U-matica bez farby a so zvýraznenými zhlukmi)



Obr. 4: Matice jednotlivých premenných

Záver

Pomocou korelačnej analýzy a analýzy zhlukov frekvencie výskytu častíc povýstrelových splodín sme zistili, že možno pomerne dobre rozlíšiť iba dva typy streliva z uvažovanej vzorky, a to Browning a Luger. Namerané hodnoty chemických častíc povýstrelových splodín neumožňujú pri zvyšných dvoch typoch streliva (Makarov a Tokarev) identifikovať, ktorým sa strieľalo. Najmä zhlukovanie pomocou neurónovej siete však poskytuje sľubné možnosti na podobné analýzy v budúcnosti. Presnejšie výsledky by sme mohli dostať pri použití väčšieho počtu vzoriek, prípadne väčšieho počtu typov streliva.

V súlade s ⁴ odporúčame rovnako analyzovať pôvod streliva nástrojmi zhlukovej analýzy, ktorá zo všetkých dostupných metód dokáže najlepšie identifikovať podobné vzorky. Namiesto hierarchického zhlukovania použitého v ⁴ ale odporúčame používať modernejšie zhlukovacie metódy. Výhodou použitia neurónovej siete je jednoduchá interpretácia využívajúca zrozumiteľný grafický výstup.

Pod'akovanie: Táto práca bola podporená grantom VEGA 1/0143/11.

⁴ BROZEK-MUCHA, Z., ZADORA, G. (2003) Grouping of ammunition types by means of frequencies of occurrence of GSR, *Forensic Science International* 135, 97-104.

Literatúra

BROZEK-MUCHA, Z., ZADORA, G. (2003) Grouping of ammunition by means of frequencies of occurrence of GSR, *Forensic Science International* 135, 97-104.

KOHONEN, T. (1982) Self-organized formation of topologically correct feature maps, *Biological Cybernetics* 43, pp. 59–69.

MATLAB version 7.6. Natick, Massachusetts: The MathWorks Inc., 2010.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

VESANTO, J., HIMBERG, J., ALHONIEMI, E., PARHANKANGAS, J. (2000) SOM toolbox for Matlab 5, Helsinki University of Technology, Espoo.

Key words: ammunition, gunshot residues, statistics analysis.

Summary

In this paper, we analyze data from [1], that consist of gunshot residues from four different ammunition types. The aim is to identify ammunition type using only chemical composition of each sample. Via cluster analysis, we correctly classified 2 types of ammunition. It is not possible to distinguish between rest 2 ammunition types.

*Mgr. Miroslav Sabo
Katedra matematiky a deskriptívnej geometrie
Stavebná fakulta STU v Bratislave
Radlinského 11, 813 68 Bratislava
email: sabo@math.sk
Tel. č: 0902 167 061*

*RNDr. Tatiana Hajduková
Katedra manažmentu a informatiky
Akadémia Policajného zboru v Bratislave
Sklabinská 1, 835 17 Bratislava 35
email: Tatiana.hajdukova@minv.sk
Tel. č.: 0961 057 297*

Recenzent: plk. Ing. Peter Obšitník