VYUŽITIE DATA MININGU V BOJI PROTI TERORIZMU

Beáta Stehlíková[*)]

ABSTRAKT

Cieľom príspevku je prezentovať prehľad prostriedkov data miningu navrhnutých, resp. realizovaných pre účely boja proti terorizmu vo svete. Väčšina programov využívajúcich analytické nástroje data miningu má tendenciu poskytovať nielen nástroje data miningu, ale tiež nástroje, ktoré vedú užívateľa pri ich analýze a podporujú jeho rozhodovanie. Dátové sklady obsahujú dáta upravené takým spôsobom, ktorý umožňuje čo najľahšie získavanie analytických informácií z nich. Konečným cieľom data miningu využívaného v podnikaní je predikcia ľudského správania. Tento cieľ je možné modifikovať tak, aby šlo o odhaľovanie správania sa teroristov.

Kľúčové slová:
Data mining, terorizmus, zdroje údajov, analytické nástroje

ABSTRACT

The aim of this paper is to present an overview of the data mining tools proposed, respectively implemented to counter terrorism in the world. Most programs have a tendency to not only provide data mining tools, but also tools that lead the user in analyzing them and support his decision. Data warehouses contain data arranged in such a way that allows for the easiest acquisition of analytical information from them. The ultimate objective data mining utilized in business is forecast human behavior. Objective can be modified to detecting behavior of terrorists.

Key words:
Data mining, terrorism, data sources, analytical tools

1 ÚVOD

Slovo terorizmus (z latinského terrere - vydesiť, vystrašiť) sa prvýkrát začalo používať v 14. storočí vo Francúzsku. Prvý pokus definovať terorizmus bol oveľa

*) prof. RNDr. Beáta Stehlíková, CSc. Paneurópska vysoká škola, Tematínska 10, 851 05 Bratislava, e-mail: stehlikovab@gmai.com

neskôr – až v roku 1937 z iniciatívy Spoločnosti národov. Terorizmus nie je ideológia, ale zhluk násilných taktík, ktoré sú teroristickými trestnými činmi, ak sa používajú proti legitímnemu ústavnému poriadku. Dodnes sa používajú odlišné definície terorizmu. Členské štáty Európskej únie sa dohodli na spoločnej definícii terorizmu. Rámcové rozhodnutie Rady 2002/475/SVV z 13. júna 2002, v znení rámcového rozhodnutia Rady 2008/919/SVV z 28. novembra 2008, definuje trestné činy, ktoré by mali byť kriminalizované podľa vnútroštátneho práva členských štátov ako teroristické trestné činy - patrí medzi nich násilná trestná činnosť, únosy, branie rukojemníkov, zničenie infraštruktúry, výroba a používanie zbraní, výbušnín alebo CRBN látok alebo hrozba spáchať niektorý z týchto činov.

Terorizmus sa bezprostredne dotýka každého z nás – každý z nás sa môže stať jeho obeťou. Preto nie je prekvapujúce, že spoločnosť je po kybernetickej kriminalite najcitlivejšia práve na terorizmus [7]. Pátranie po teroristických skupinách je však zložité. Mnohé organizácie sú dlhodobo latentné. Teroristické organizácie sú často rozdelené na bunky, ktoré pracujú nezávisle od seba a jedna o druhej prakticky nevedia. Ani sledovanie finančných tokov nevedie k ich odhaleniu. Financovanie teroristov je vo väčšine prípadov veľmi komplikované. Peniaze zo štátnych zdrojov krajín, ktoré terorizmus podporujú, ako aj financie zo strany záujmových skupín sa zložito distribuujú po svete cez rôzne nastrčené osoby alebo firmy. Skúsenosti zo sveta ukazujú, že práve data mining je jedným z prostriedkov, ktorý umožňuje odhalenie teroristov.

2 DATA MINING

Data mining je možné charakterizovať ako dolovanie dát, objavovanie znalostí v databáze. Data mining je definovaný ako identifikácia zaujímavých štruktúr v dátach, ktoré stanovujú vzory, štatistické alebo prediktívne modely z údajov ako aj vzťahy medzi časťami údajov [6]. Data mining je špecifický proces získavania nových užitočných informácií. Výsledkom modelovania je popis vzorov a vzťahov v údajoch. Zastrešuje existujúce techniky analýzy dát ako aj získavania znalostí, informácií. Objavovanie znalostí v databáze má interdisciplinárny charakter. V rámci data miningu sa využíva zhluková analýza, regresia, logistická regresia, rozhodovacie stromy, neurónové siete [11] a ďalšie metódy. Metódy data miningu sa neustále vyvíjajú [21]. Podľa [8] data mining je jedným z komponentov – a to kľúčovým – extrakcie poznatkov z databáz KDD (Knowledge Discovery from Databases).

Problematiku využitia data miningu v boji proti terorizmu môžeme rozdeliť do dvoch, navzájom sa prelínajúcich a navzájom sa ovplyvňujúcich oblastí – zdroje údajov data miningu a modelovanie.

3 ZDROJE ÚDAJOV PRE DATA MINING

Nutným predpokladom data miningu v komerčnej sfére je existencia veľkej trénovacej množiny. Existujú milióny záznamov o spotrebiteľskom správaní ľudí a nie

je preto ťažké nájsť pomocou prostriedkov data miningu skupinu ľudí, na ktorú majú firmy zamerať svoje marketingové aktivity. Rovnako aj pri podvodoch s telefónnymi kartami, je možné na základe tisícov prípadov modelovať znaky podvodného správania. Podobne je tomu pri odhaľovaní prania špinavých peňazí. Príkladov terorizmu v porovnaní s uvedenými príkladmi je relatívne málo a na prvý pohľad sa môže zdať, že správanie teroristov nemá spoločné znaky.

V prácach [19], [20] sa doporučuje využitie data miningu pre detekciu neobvyklého správania sa. Mnohé z potrebných údajov nemusia byť špeciálne zhromažďované. Treba však o nich vedieť a treba vedieť ako ich pre potreby boja proti terorizmu použiť. Využiť môžeme v podstate akékoľvek dáta, ktoré je možné spracovať k vytvoreniu prediktívnych modelov podozrivého konania. Či už ide o dáta z bankových transakcií, telefónnych hovorov, emailov, webov, chatov, sociálnych sietí a podobne. Mnohé informácie nie je potrebné zhromažďovať zvláštnym spôsobom – obsahujú ich databázy určené pre marketingové účely. Prítomnosť teroristov môže signalizovať napríklad nákup dusíkatého hnojiva v oblastiach, kde sa hnojivo obvykle nepoužíva, atypické nákupy množstva potravín neobvyklých pre daný región.

Nie všetky prejavy neobvyklého správania súvisia s terorizmom. Na druhej strane činnosť teroristov je sofistikovaná a často latentná, preto je potrebné sledovať zmeny v čase. Pre identifikáciu znakov, ktoré treba sledovať sa môže použiť zhluková analýza. Systémy pre detekciu teroristov, bioteroristov či chemoteroristov často používajú údaje, ktoré majú priestorový a priestorovo-časový charakter.

Pre úplnosť treba spomenúť aj právnu stránku problému zhromažďovania údajov. Experti boja proti terorizmu a obhajcovia ľudských práv vedú dlhé diskusie o práve na súkromie, zásahoch do súkromia zhromažďovaním údajov [19]. Ale to je iná stránka problému a nie je predmetom tohto príspevku. Navyše existujú už systémy [12] zamerané na spracovanie citlivých údajov bez porušenia súkromia dotknutých osôb.

Štúdia [10] obsahuje zoznam organizácií zaoberajúcich sa výskumom terorizmu a 28 zdrojov rôznych databáz a archívov terorizmu. Sú cenným zdrojom informácií a môžu slúžiť na overenie správnosti algoritmov.

Svet okolo nás využíva výdobytky techniky a svet terorizmu je tiež digitálny. Autori článku [17] identifikovali 795 kľúčových slov a základe nich našli 160 tisíc webových stránok súvisiacich s činnosťou teroristov. Zosieťovaním domén ich počet sa zvýšil ma 360 tisíc. Data mining môžeme použiť aj v prípade, keď zdrojom údajov sú multimediálne databázy a skladiská [14]. Mimoriadne cenným zdrojom údajov a informácií v boji proti terorizmu sú sociálne siete [1], [2].

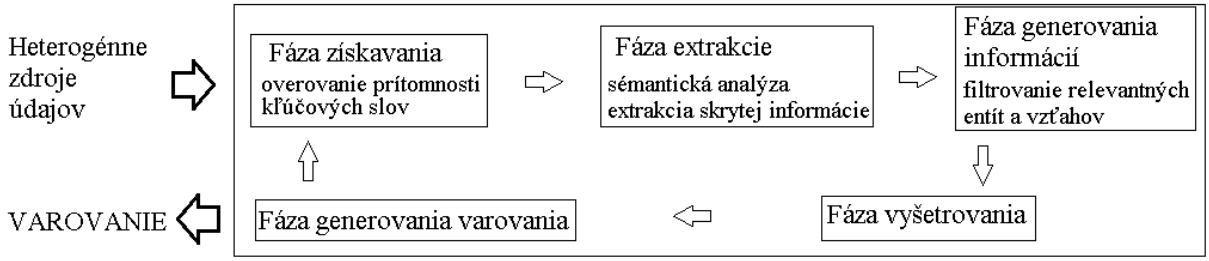
4 ANALYTICKÉ NÁSTOROJE DATA MININGU

Okrem veľkého množstva článkov popisujúcich zhromažďovanie relevantných údajov či nástrojov [18] existuje tiež množstvo zjednodušených kritických postojov

k využitiu data miningu pre detekciu teroristov. Tieto kritiky argumentujú s potrebou značného množstva informácií pri nízkej presnosti data miningu – neprijateľne vysokým podielom chýb II. typu – okolo 1 %. To znamená, že výsledok data miningu je pozitívy (osoba je označená za teroristu), ale skutočnosť je negatívna (osoba teroristom v skutočnosti nie je) v jednom zo 100 prípadov. Autori príspevku [9] tieto neisté výsledky data miningu neváhajú označiť za plytvanie národnými zdrojmi. Argument, že 1 % chyba predstavuje 3 milióny potencionálnych teroristov z 300 miliónov obyvateľov USA je neprofesionálny. Čisto zo štatistického hľadiska by to znamenalo, že údajová základňa data miningového systému nedisponuje o žiadnom obyvateľovi žiadnou informáciou okrem jeho existencii. Ale informácie v USA, a nielen tam, v štátnych databázach ako aj súkromných databázach o ľuďoch existujú. Veď práve data miningový nástroj Clementine aplikovaný na databázy z informačného hľadiska excelentné, bol jedným z prostriedkov, ktorý identifikáciou závažných problémov regiónov umožnil Obamovi pripraviť si cielené prejavy a vyhrať prezidentské voľby.

Okrem toho je potrebné zdôrazniť, že na rozdiel od klasického využívania data miningu, kde je kritériom kvality modelu jeho presnosť, v prípade boja proti terorizmu je kritériom kvality je jeho profitabilita, t.j. pravdepodobnosť odhalenia teroristov.

Výsledky data miningu nie sú okamžite aplikované - prechádzajú rukami odborníkov, analytikov. Až oni v spolupráci s ďalšími zložkami v konečnom dôsledku rozhodnú o relevantnosti výsledku – potencionálnej hrozbe, resp. vyslaní varovného signálu, či ďalšom overovaní a zbieraní informácií. Naše tvrdenie podporujú napríklad práce [16], [22]. Tiež v [3] sa konštatuje, že výsledky data miningu sú podrobované ďalšej analýze a vyšetrovaniu predtým, než by mali mať falošne pozitívne výsledky negatívny dopad na osoby.



Obrázok 1 Schéma cyklu spracovania informácií [15]

Súčasťou data miningu je aj tzv. extrakcia dôkazov, pomocou ktorej získame štrukturované údaje z neštrukturovaného dokumentu v prirodzenom jazyku. Metódy detekcie, objavovania spojení (link discovery) identifikujú multirelačné vzory [5], ktoré následne identifikujú potenciálne hroziace aktivity. Pri analýze spojení sa študujú siete, presnejšie ich reprezentácie pomocou aparátu teórie grafov.

Posledné roky v oblasti využívania data miningu v boji proti terorizmu sa nesú v znamení analýzy sociálnych sietí [1] ako aj intenzívneho využívania neurónových sietí [15].

Do popredia sa dostáva text mining – data mining textových dokumentov [15]. Spomeňme ešte špeciálne metódy detekcie signálov teroristov na zahájenie útoku [4].

Tabuľka 1 Prehľad programov používaných v boji proti terorizmu

Oblasť	**Systém**
USA	CAPPS II, Secure Flight, ATS, DARPA, MATRIX, NIMD, Analyst Notebook I2, SCOPE, Verity K2 Enterprise, PATHFINDER, Autonomy, CI-AIMS, CARDS, BioSense, Foreign Terrorist Tracking Task Force Activity (FBI), NETLEADS, ICEPIC, I2F, PAINT, VACE, ADVISE, TALON, TIDE, IDW (FBI)
Európa	CAHORS, ODYSSEY, INDECT, SAMURAI, ADABTS, SCION, CHRISTINA

Analytické nástroje v spojení s príslušnými zdrojmi údajov sú mocným nástrojom využívaný pri tvorbe programov a systémov pre detekciu teroristov. Vynikajúci prehľad nájdeme v [13].

5 ZÁVER

Ekonomické, sociálne aj kultúrne nerovnosti vytvárajú pôdu pre vznik terorizmu. Zároveň sa v niektorých oblastiach sveta posilňujú separatistické snahy. Znamená to, že ešte niekoľko desiatok rokov zostane terorizmus súčasťou medzinárodných vzťahov. A to je výzva pre vývoj nových účinných špeciálnych metód data miningu namierených proti terorizmu.

LITERATÚRA

[1] Ball, L.: Automating social network analysis: A power tool for counter-terrorism. Security Journal 2013.

[2] Ball, L., Craven, M.: Automated Counter-Terrorism. In Intelligence and Security Informatics Conference (EISIC), 2013 European, 216-216. IEEE.

[3] De Rosa, M.: Data mining and data analysis for counterterrorism. CSIS Press, 2004.

[4] Drozdova, K., Samoilov, M.: Predictive analysis of concealed social network activities based on communication technology choices: early-warning detection of attack signals from terrorist organizations. Computational and Mathematical Organization Theory, 16(1), 2010, 61-88.

[5] Džeroski, S.: Relational data mining. 887-911. Springer US, 2010.

[6] Fayyad, U., Uthurusamy, R.: Evolving data into mining solutions for insights. Communications of the ACM, 45(8), 2002, 28-31.

[7] Chau, M., Xu, J. J., Chen, H.: Extracting meaningful entities from police narrative reports. In Proceedings of the 2002 annual national conference on Digital government research, 2002, 271-275. Digital Government Society of North America.

[8] Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P.: Fundamentals of Data Warehouses, Berlin: Springer, 2000.
[9] Jonas, J., Harper, J.: Effective counterterrorism and the limited role of predictive data mining. Cato Institute, 2006.
[10] Kennedy, L. W., Lunn, C. M.: Developing a Foundation for Policy Relevant Terrorism. Research in Criminology, 2003
[11] Lee, N.: Artificial Intelligence and Data Mining. In Counterterrorism and Cybersecurity , 63-80. Springer New York, 2013.
[12] Liu, K., Kargupta, H., Ryan, J.: Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. Knowledge and Data Engineering, IEEE Transactions on, 18(1), 2006, 92-106.
[13] Moeckli, D., Thurman, J.: Survey of Counter-Terrorism Data Mining and Related Programmes. DETECTER Deliverable D, 8, 2009.
[14] Petrushin V.A., Khan, L.: Multimedia Data Mining and Knowledge Discovery, New York: Springer-Verlag, 2006
[15] Qureshi, P. A. R., Memon, N., Wiil, U. K.: Detecting Terrorism Evidence in Text Documents. In Social Computing (SocialCom), 2010 IEEE Second International Conference, 2010, 521-527.
[16] Qureshi, P. A. R., Wiil, U. K., Memon, N.: Modeling Early Warning System to Predict Terrorist Threats. In the Proceedings of the 10th International Symposium on Knowledge and System Sciences, Hong Kong, China, 2009, 179-190.
[17] Reid, E., Qin, J., Chung, W., Xu, J., Zhou, Y., Schumaker, R., ... a Chen, H.: Terrorism knowledge discovery project: A knowledge discovery approach to addressing the threats of terrorism. In Intelligence and Security Informatics. 2004, 125-145. Springer Berlin Heidelberg.
[18] Subrahmanian, V. S.: Handbook of Computational Approaches to Counterterrorism. Springer, 2013.
[19] Thuraisingham, B.: Data mining, national security, privacy and civil liberties. ACM SIGKDD Explorations Newsletter, 4(2), 2002, 1-5.
[20] Thuraisingham, B.: Web data mining and applications in business intelligence and counter-terrorism. CRC Press, 2003.
[21] Venkatadri, M., Reddy, L. C.: A review on data mining from past to the future. International Journal of Computer Applications, 15(7), 2011, 19-22.
[22] Wiil, U. K., Memon, N., Gniadek, J.: Knowledge Management Processes, Tools and Techniques for Counterterrorism. In KMIS 2009, 29-36.

Článok recenzovali dvaja nezávislí recenzenti.