UDK: 343.985:004.8
351.746.2

Pregledni rad

VEŠTAČKA INTELEGENCIJA U PRIKUPLJANJU I ANALIZI PODATAKA U POLICIJI

Kristijan Kuk[1]
Kriminalističko-policijska akademija, Beograd

Sažetak: Kompleksni realni problemi sve češće zahtevaju inteligentne sisteme koji kombinuju znanje, tehnike i metodologije iz različitih izvora. Inteligentni sistemi bazirani na tehnikama veštačke inteligencije koje asociraju na ponašanje ljudi mogu da obavljaju procese učenja, zaključivanja i rešavanje raznovrsnih problema. Ovakvi sistemi, koji automatski mogu da izvrše zadatke zadate od strane korisnika ili drugih softvera, danas se sreću pod imenom inteligentni agenti. Samostalno, inteligentni agenti na Internetu mogu veoma uspešno da izvode neki pretraživački posao u ime i za potrebe raznih korisnika. Zbog efikasnog sakupljanja, manipulisanja i upravljanja podacima, ovakvi softveri mogu biti veoma interesantni sa stanovišta inteligentne analize podataka u mnogim oblastima policije. Analiza podataka sakupljenih od strane inteligentnog agenta (softverskog robota – bota) može se uspešno iskoristiti, između mnogih poslova u policiji, i na polju kriminala i naročito pojavnog oblika sajber kriminala, bezbednosti saobraćaja, vanrednih situacija itd. Kako bi sakupljanje i analiza podataka iz kriminalnih aktivnosti na Internetu bila efikasna, neophodno je sagledati postojeće tehnike veštačke inteligencije koje se koriste za zaključivanje u inteligentnim agentima. S druge strane, treba iskoristiti metode veštačke inteligencije u pronalaženju podataka pri inteligentnoj analizi podataka (*data mining*-u) koja je našla široku primenu u oblasti poslovanja preduzeća, ekonomije, mehanike, medicine, genetike, saobraćaja i sl.

Ključne reči: veštačka inteligencija, inteligentni agenti, botovi, *data mining*, analiza podataka u policiji.

1 e-mail: kristijan.kuk@kpa.edu.rs

Uvod

Inteligentni računarski sistemi pripadaju vrsti računarskih sistema koji imaju sposobnost da otkrivaju nevidljive veze i oblike podataka koji mogu uspešno da se primene prilikom donošenja poslovnih odluka. Veštačka inteligencija (*artificial intelligence – AI*) jeste deo nauke o računarima koji se bavi dizajniranjem računarskih sistema, tj. sistema koji poseduju karakteristike koje asociraju na ponašanje ljudi, kao što su razumevanje jezika, učenje, zaključivanje, rešavanje problema i slično[2]. Sistemi veštačke inteligencije nemaju iste mogućnosti učenja kao ljudi, međutim mogu da imaju sposobnosti mehaničkog učenja koje se naziva mašinsko učenje i koje omogućava sistemu da prilagodi svoje ponašanje i da reaguje na promene u spoljašnjem okruženju. Za izgradnju inteligentnih sistema postoji više metoda mašinskog učenja, uključujući induktivno učenje, veštačke neuronske mreže i genetski algoritam. Mašinsko učenje je oblast veštačke inteligencije koja se bavi izgradnjom prilagodljivih računarskih sistema koji su sposobni da poboljšavaju svoje performanse koristeći informacije iz iskustva[3].

Hibridni inteligentni sistemi su računarski sistemi koji integrišu različite inteligentne tehnike. Integracija raznih pristupa veštačke inteligencije dovodi do razvoja koncepta inteligentnog agenta. Inteligentni agenti predstavljaju softver koji automatski može da izvrši zadatak koji mu postavi korisnik ili drugi softver (agent). Kada se jednom podese, oni izvršavaju svoje zadatke, automatski bez dalje intervencije korisnika. Najčešće se upotrebljavaju u automatskom traganju za informacijama, pružaju odgovore na postavljena pitanja u domenu svog znanja i mogu da informišu korisnike o interesantnim događajima (o pojavi novog članka na Internetu, prikazuju informacije o eventualnoj pojavi problema na putu između početne i krajnje destinacije, da li se zadati pojam pojavljuje negde na *Web*-u i slično[4]). Inteligentnim agentima za pretraživanje *Web*-a nazivaju se računarski programi koji samostalno izvode neki pretraživački posao u ime i za račun korisnika. Smešteni su u računaru vlasnika, što ne mora nužno biti računar krajnjeg korisnika, već neko *web* mesto. Korisnik ih mora „napuniti" informacijama o domenima svog interesovanja, pravilima pretraživanja, prioritetima, i eventualnim vremenskim ograničenjima. Nakon što agent obavi postavljeni zadatak, analiziraju se rezultati; ako rezultati nisu zadovoljavajući prema nekom unapred postavljenom kriterijumu, agent će ispraviti sam sebe. Inteligentni agenti za pretraživanje *Web*-a mogu imati različite stepene samostalnosti u izvršavanju zadataka u odnosu prema korisniku i njegovim potrebama. Neki inteligentni agenti mogu „samo" prikupljati informacije, neki mogu filtrirati poruke primljene elektronskom poštom, a neki sarađivati sa drugim agentima i na taj način obavljati vrlo složene i specifične zadatke. Do sada, razvijene su tri vrste takvih agenata[5]:

– *Web crawler* – pokušavaju da daju korisniku celovit pregled informacija, šetajući *Web*-om i izveštavajući korisnika o onome šta su pronašli;

2 S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, 2010.
3 T. M. Mitchell, *Machine Learning*, McGraw-Hill, Inc., New York, USA, 1997.
4 V. Ilić, Systems based on agent's technology, *InfoM – Journal of Information Technology and Multimedia Systems*, No. 3–4, 2002, Belgrade.
5 T. N. Nguyen, C. J. Lakhmi (Eds.) *Intelligent Agents in the Evolution of Web and Applications*, Vol. 167.

– *Web pauci* (*Web spider*) – programi koji se pomoću „crvića" uvlače u sadržaj *web* stranica, odnosno hipermedijskih dokumenata, i pronađene informacije indeksiraju i čuvaju u vlastitoj bazi podataka, koju korisnik kasnije može jednostavno lokalno pretraživati i analizirati.
– *Web* roboti (*Web robot*) – programi koji mogu u potpunosti nezavisno od korisnika izvršavati kompletne transakcije u skladu sa instrukcijama koje im je korisnik dao, kao što su kupovine na daljinu, rezervacije avio-karata, novčanih transakcija itd.

Inteligentni agenti imaju široku oblast primene u pretraživanju tako da se oni koriste za razne oblike kreiranja statističke analize, indeksiranje, tj. prikupljanje podataka preko ključnih reči kao i za pronalaženje uzorka. Inteligentni *Web* agenti imaju veliki potencijal u pronalaženju podataka (*data mining* – proces pronalaženja uzorka u velikoj količini podataka kroz niz pretraživanja), takođe mogu donositi odluke zasnovane na prethodnim pretraživanjima kako bi obavili što kompleksnije pretraživanje.

1. *Web* inteligentni agenti – botovi

Bot (skraćenica od „robot") jeste softver koji ima ulogu softverskog robota odnosno inteligentnog softverskog agenta i obično u sebi ima implementiranu veštačku inteligenciju zbog efikasnog sakupljanja, manipulisanja u upravljanja podacima[6]. Bot je regularan softverski alat koji koriste kompanije da bi prikupile njima bitne informacije spakovane u bazi podataka, informacije se koriste radi ostvarivanja materijalne dobiti, bilo na legalan ili nelegalan način, obično puštanjem reklama, kao što radi, na primer, kompanija *Google*. Postoje i botovi koji su tu da pokupe informacije potrebne za spamovanje e-mail adrese, da postavljaju komentare po raznim sajtovima ili forumima (obično sa reklamama ili malicioznim linkovima), da se ponašaju kao čovek pomoću veštačke inteligencije, mogu se dopisivati, „četovati" sa vama, i teško je proceniti da li pričate sa pravim čovekom ili robotom. U *Web* okruženju pod pojmom „bot" podrazumijevaju se softverski agenti koji interreaguju sa mrežnim servisima namenjenim ljudima kao da su i oni sami ljudi. Najčešće se primenjuju za prikupljanje informacija, kada se obično nazivaju programima za indeksiranje *Web* stranica, puzačima (*crawler*) ili paucima (*web spider*). Neki botovi mogu komunicirati sa ljudima putem standardnih internet servisa kao što su servisi za četovanje ili trenutne razmene vesti. Primer takvih botova su roboti za dopisivanje kojima ljudi mogu postavljati pitanja iskazana u prirodnom jeziku, najčešće engleskom, i dobijati pisane i verbalne odgovore. U praksi, takvi se roboti koriste za odgovaranje na pitanja o vremenskoj prognozi, o sportskim rezultatima, o poštanskim ili telefonskim brojevima, o raznim valutama, o redovima vožnje ili redovima letenja, itd. Koriste se i u elektronskom poslovanju kada na zahtev korisnika prikupljaju informacije o proizvodima i uslugama određene vrste, upoređuju cene i uslove prodaje, nabavke ili isporuke, pronalaze optimalne kombinacije

6 D. Vuletić, Napadi na računarske sisteme, *Vojnotehničkih glasnik / Military Technical Courier*, Vol. LX, No. 1, 2009, str. 235–249.

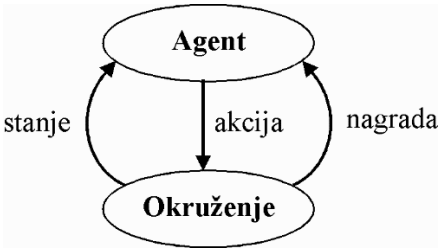
proizvoda ili usluga, npr. u industriji putovanja i u turizmu, pronalaze *Web* stranice sa posebnim ponudama, itd. Važnu ulogu imaju u obrazovanju kada mogu preuzeti ulogu instruktora ili učitelja, kao i u računarskim igrama gde oponašaju prijateljskog igrača ili protivnika, tj. neprijatelja.

Postoji više vrsta botova, a najpoznatijih od njih su: *Spambot*, *Chatbot* i *Web crawling bot*.

- *Spambot* ostavlja reklame, maliciozne linkove po sajtovima, šalje neželjene mejlove na e-mail adrese – spamuje mesta za postavljanje komentara koje pronalazi na stranicama sajta, može da se registruje na sajtu ili forumu i da vrši spamovanje u tom okruženju. Ovaj tip bota spada u maliciozne.
- *Chatbot* se obično koristi na sajtovima sa pornografskim sadržajem. Posetiocu se pojavi *chatbox* poput onoga koji se pojavi kad neki korisnik kontaktira drugog korisnika na društvenoj mreži Fejsbuku. U tim slučajevima, *chatbot* poziva korisnika na detaljnije upoznavanje pri čemu mu postavlja niz opštih pitanja, a korisnik ne znajući da priča sa robotom ostavlja svoje poverljive podatke. Postoje i regularni *chatbot*-ovi, koji se obično koriste na stranicama vezanim za tehničku podršku. Oni odgovaraju na pitanja koja su im većinom poznata, a pitanja koja im nisu poznata prosleđuju dalje odgovornom licu ili ga upućuju na drugu *web* stranicu.
- *Web crawling bot* su programi koji služe da „pretresu" *web* sajt u potrazi za bitnim podacima, pa da te podatke pošalju nazad u bazu podataka koja će se upotrebljavati za dalje akcije. Jedan od najpoznatijih *Web crawlera* je *Googlebot*, bot napravljen od strane kompanije *Google* koji ima zadatak da osvežava indeks sajtova na *Google*-u i podataka sa tih sajtova, kao i da se automatski poveže na drugi sajt čiji link pronađe na sajtu. Osim *Googlebota* i druge velike kompanije koriste svoje botove, pa tako na primer *Yahoo* koristi *Slurp* dok *Microsoft Bing* koristi *Bingbot*. *Web crawling bot* se još zove i pauk jer preko linkova ide sa sajta na sajt i tako pravi svoju mrežu.

U velikom broju metoda veštačke inteligencije koje se koriste u botovima, primenjuju se metod Markovljevog lanca, kao jedne od nezaobilaznih komponenata u procesu odlučivanja. Svaki stohastički proces čije trenutno stanje je jedino bitno za dalje odvijanje procesa naziva se Markovljev proces. Poznajući trenutno stanje, može se odredili kako će se proces odvijati u budućnosti, ali i kako se proces odvijao u prošlosti. Pri ovom procesu nije bitno kroz šta je sistem prošao i kakva su bila prethodna stanja. Drugim rečima, sistem nema memoriju, ne postoji memorijski efekat prethodnih stanja ali se neposredno „ovde i sada" generiše dalji tok događanja. Ovakav proces je poznat kao proces bez memorije ili Markovljev lanac.

Uopšteno gledano, interakcija bilo koje vrste agenta sa bilo kojim okruženjem, može se predstaviti kao na slici 1.



Slika 1: Agent u interakciji sa okruženjem

Interakcija između okruženja i agenta koji se u njemu nalazi, može se prikazati kao:

$$S_0 \xrightarrow[r_0]{a_0} S_1 \xrightarrow[r_1]{a_1} S_2 \xrightarrow{a_2} \cdot r_2 \cdot \tag{1}$$

gde je na početku agent u stanju sistema S_0 i deluje akcijom a_0. Sistem vraća nagradu r_1 agentu i prelazi u stanje S_1. Zatim, u stanju S_1 agent deluje akcijom a_1, prima nagradu r_2 i sistem prelazi u stanje S_2 i tako dalje.

Markovljev model je matematički model koji opisuje stohastičke procese, odnosno procese koji generišu slučajnu sekvencu ishoda u skladu sa određenim verovatnoćama[7]. U stohastičkom svetu akcija a_t koji se nalazi u stanju S_t nije rezultat novog jednoznačnog stanja nego raspodela verovatnoća $P(S_t+1|S_t,a_t)$ po svim mogućim stanjima S_t+1. Verovatnoća prelaska stanja S u stanje S' akcijom a, data je jednačinom:

$$P_{\mathbf{S}}^{a} = \Pr\{S_{t+1} = S' \mid S_t = S, a_t = a\} \tag{2}$$

Skriveni Markovljevi lanci su stohastički konačni automati kod kojih je prelazak iz stanja u stanje obeležen verovatnoćom; kod skrivenih je čak nemoguće unapred odrediti prelaske stanja već samo posledice. Skriveni Markovljevi modeli omogućuju jednostavno modeliranje procesa iz stvarnog sistema; skriveni lanac Markova (*Hidden Markov models – HMM*) deli događaje, odnosno odgovarajuće promenljive u skrivena X i posmatrana Y stanja. Kao osnova sistema za prepoznavanje koristi se HMM sa konačnim brojem stanja $\{S_1,\ldots,S_M\}$. Model se može opisati uređenom petorkom[8] $\lambda=(A,b,\pi,M)$, gde je M broj stanja modela, $A=[a_{ij}]$ matrica verovatnoća prelaza između stanja Markovljevog modela, π vektor inicijalnih verovatnoća stanja, $b=[b_1,..,b_M]$ vektor uslovnih gustina raspodela $b_j(0)$ opservacija po stanjima $Sj, j \in \{j,\ldots,M\}$ (emitujuće gustine raspodela stanja). Matrica verovatnoće prelaza i vektor inicijalnih verovatnoća stanja zadovoljavaju ograničenja:

$$\sum_{j=1}^{M} a_{ij} = 1, \forall i \in \{1,\ldots, M\}, \sum_{j=1}^{M} \pi_j = 1 \tag{3}$$

7 E. Fosler-Lussier, *Markov Models and Hidden Markov Models: A Brief tutorial,* Technical Report (TR-98-041), 1998.

8 L. R. Rabiner, A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, 1989, str. 257–286.

HMM predstavlja grafički model verovatnoće, na kojem su bazirani mnogi programi za predviđanje, koji sa velikom tačnošću prepoznaju statičke obrasce i klasifikuju statičke podatke. HMM modeli su razvijeni u cilju prevazilaženja nedostatka metoda zasnovanih na veštačkim neuronskim mrežama.

2. Inteligentna analiza podataka i *data mining*

Dubinska analiza podataka je način obrade podataka koji podrazumeva razne postupke koji imaju za cilj dobijanje korisnog znanja iz podataka. Osnovno svojstvo po kojem se dubinska analiza podataka razlikuje od tradicionalne ili „obične" analize je primena postupaka mašinskog učenja. Inteligentna analiza podataka (*intelligent data analysis*) drugi je naziv za dubinsku analizu podataka. Pridev „inteligentna" naglašava da je to analiza podataka zasnovana na postupcima veštačke inteligencije, pre svega mašinskog učenja. Tehnike koje se najčešće primenjuju uglavnom su izvedene iz tri glavne oblasti: statistike, mašinskog učenja i baza podataka. Mašinsko učenje je zanimljivo zbog svoje težnje da se približi ljudskom učenju po efikasnosti, kao i da ga objasni, odnosno pruži teorijski model za njega. Neka od najvažnijih pitanja mašinskog učenja[9] su:

- šta se može naučiti i pod kojim uslovima?
- kako se povećava efikasnost učenja u zavisnosti od obima iskustva?
- koji su algoritmi pogodni za koje vrste problema?

Odgovore na ova najvažnija pitanja mašinskog učenja upravo leže u teoriji modela učenja. U tom smislu, treba sagledati vrste zadataka učenja koje se često pojavljuju. Jedan od najčešćih zadataka učenja koji se javlja u praksi je klasifikacija, tehnika kojom se vrši prepoznavanje vrste objekata. S druge strane, regresija je zadatak mašinskog učenja u kome objektima odgovaraju vrednosti iz skupa realnih brojeva, kao što je, na primer, predviđanje potražnje robe u zavisnosti od raznih faktora koji na nju utiču. Razvoj nekih primena induktivnog mašinskog učenja, pre svega u izgradnji sistema zasnovanih na znanju i otkrivanju znanja u podacima (*Data Mining – DM*) dovelo je do naglašenog zahteva za razumljivošću naučenog znanja. *Data mining* se razlikuje od klasičnih statističkih metoda po tome što se ne odvija po unapred utvrđenim pravilima, već pokazuje kreativnost u analizi podataka i na taj način može da otkrije nova, neočekivana pravila. Osnovni cilj DM jeste otkrivanje do sada nepoznatih odnosa između podataka. Analizom ogromnih baza podataka upotrebom DM definišemo relacije, obrasce ili forme ponašanja, neophodne za odlučivanje i predviđanje.

U inteligentnoj analizi podataka postoje dva modela za istraživanje:

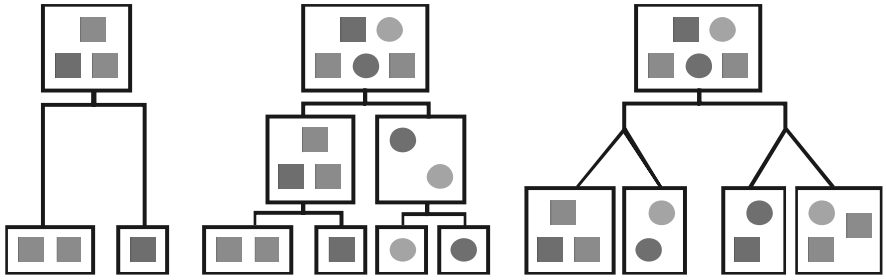
- statistički i matematički model kao što su multivariciona analiza (*multivariate analysis –MVA*) i analiza glavnih komponenti (*principal component analysis – PCA*);
- metodi mašinskog učenja kao na primer učenje klasifikacija i pravila regresije, učenje grupisanja – klastering, učenje višestrukih modela (*ensemble*), selekcija i estimacija atributa (*attribute/ feature selection*) i otkrivanje pravila udruživanja (*association*).

9 P. Janičić, M. Nikolić, *Veštačka inteligencija*, Matematički fakultet, Beograd, 2010, str. 161.

Prema standardu CRISP-DM (*CRoss-Industry Standard Process for Data Mining*) metodologije tipični zadaci u inteligentnoj analizi podataka su svakako deskripcija, estimacija, predikcija, klasifikacija, klastering i analiza asocijacija. U odnosu na prirodu problema i zadataka mogu se izdvojiti dve nazastupljenije tehnike u procesu inteligentne analize podataka: klasifikacija i klasterovanje.

Klasifikacija (*Classification*)

Klasifikacija je jedna je od najzastupljenijih metoda istraživanja podataka. U tu grupu spadaju metode za svrstavanje entiteta u jednu od nekoliko prethodno definisanih grupa ili klasa. U postupku istraživanja formiraju se klasifikacioni modeli, ispitivanjem prethodno klasifikovanih podatka (slučajeva). Ovo je primer nadgledanog modela, jer zahteva postojanje skupa podataka u kojem je za svaki ulazni slučaj definisana klasa kojoj pripada. Svaki slučaj sadrži niz atributa, od kojih je jedan specijalan atribut određen za oznaku klase. Suština klasifikacije je pronalaženje modela koji opisuje atribut koji označava klasu kao funkciju ulaznih atributa. Najčešći algoritmi klasifikacije su stabla odlučivanja, neuronske i Bajesove mreže.

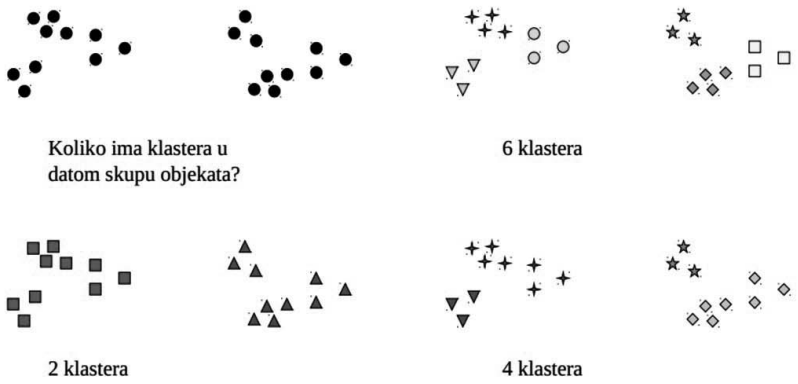


Slika 2: Vrste klasifikacije

Klasterovanje – grupisanje (*Clustering*)

Ovom metodom se pronalazi prirodno grupisanje slučajeva na osnovu niza atributa, tako da atributi unutar jedne grupe imaju prilično slične vrednosti, a među grupama postoji značajna razlika. Logičke celine, odnosno dobijene grupe, nazivaju se klasteri. Za razliku od klasifikacije, kod koje postoje predefinisane klase, ovde to nije slučaj. Pošto ne zahteva skup podataka za treniranje, klasifikacija pripada nenadgledanim metodama istraživanja podataka. Svi ulazni atributi se podjednako tretiraju. Čak se od korisnika ne zahteva ni određivanje ulaznih atributa, niti izlaza, već samo eventualno, broj klastera. Većina algoritama klasterovanja se razvija kroz veći broj iteracija, dok se granice klastera ne stabilizuju. U skladu sa osnovnim definicijama istraživanja podataka, može da se kaže da je suština klasterovanja otkrivanje skrivene vrednosti i promenljivih

koje precizno klasifikuju podatke. Metode klasterovanja imaju široku primenu kao što je e-uprava[10] (*e-government*), jer dosta efikasno rade sa različitim tipovima podataka (diskretne, numeričke, kategoričke vrednosti). Često predstavljaju početan korak u istraživanju podataka, koji prethodi klasifikaciji. Često je u upotrebi i naziv segmentacija. Dva najčešća pristupa problematici grupisanja rezultata pretraživanja poklapaju se sa pristupima u okviru *data mining* metodologije i to su nadgledano učenje (*supervised learning*) i nenadgledano učenje (*unsupervised learning*). Kod nadgledanog učenja koriste se algoritmi mašinskog učenja kao što su mašine potpornih vektora (*Support Vector Machines – SVM*) i neuronske mreže, koji na osnovu zadatog skupa „ručno" klasifikovanih dokumenata, nakon procesa „učenja" vrše klasifikaciju stranica na osnovu njihovog sadržaja. U drugom pristupu učenja, tj. nenadgledanom učenju vrši se klasifikacija dokumenata bez prethodnog „znanja", koristeći samo informacije prisutne u samim podacima, a na osnovu postupka k-najbližih suseda (*k-nearest neighbors – KNN*). Ovaj pristup se često koristi u istraživanju podataka, kada ne postoji prethodno definisani korpus podataka ili kada nismo sigurni šta tačno tražimo u podacima. U prilog tome, rezultate pretraživanja možemo posmatrati kao obične dokumente i primenjivati standardne metode mašinskog učenja za njihovo grupisanje (nadgledano – u odnosu na zadatu temu ili nenadgledano – po sličnosti dokumenata).

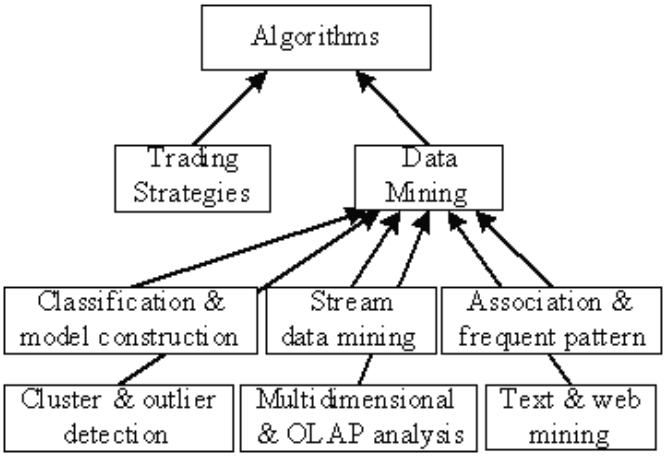


Slika 3: Metod klasterovanja

Otkrivanje znanja pomoću inteligentnih agenata

Znanje, rutinski kreirano i primenjeno u današnjim inteligentnim agentima, skriva se u velikim bazama podataka i može se izvući tehnikama *data mining*-a. Glavni korak u tom pravcu jeste transformacija otkrivenog znanja iz mehanizma zaključivanja ili ponašanja inteligentnih agenata.

10 G. Šimić, Z. Jeremić, E. Kajan, D. Ranđelović, Framework for Delivering e-Government Support, *Acta Polytechnica Hungarica*, Vol. 11, No. 1, 2014, str. 79–96.



Slika 4: Korišćenje algoritama[11] u *data mining*-u

Data mining se često poistovećuje sa pojmom otkrivanje znanja u bazama podataka (*Knowledge Discovery in Databases – KDD*). Otkrivanje znanja u bazama podataka jeste proces identifikacije novih, ispravnih, korisnih i razumljivih šablona i modela iz podataka skladištenih u bazama podataka. *Data mining* je samo jedan korak u tom procesu, čiji je zadatak da pronađe i otkrije šablone i modele. Problem velikog broja atributa u skupovima podataka pojavljuje se od sredine devedesetih godina i danas je vrlo aktuelan. Oblasti sa velikim brojem atributa (nekoliko stotina) danas postoje u genetici: problem odabira gena s određenim koeficijentom ekspresije, i u klasifikaciji teksta: problem otkrivanja neželjenih poruka elektronske pošte (*spams*), problem automatskog sortiranja URL-ova u *Web* direktorijumu. Čak i u skupu podataka sa mnogo manjim brojem atributa, korisno je odabrati koji su atributi važniji za kreiranje boljeg modela u odnosu na druge. Polazeći od činjenice da postoje postupci dubinske analize koji ne mogu odlučiti koji su važniji od nevažnih atributa, proces odabira atributa je od presudnog značaja u onim situacijama kad su razlike od samo nekoliko procenta značajne u klasifikaciji ili predviđanju.

Osnovni zadatak selekcije atributa je redukcija dimenzionalnosti prostora atributa i uklanjanje redundantnih, irelevantnih i zašumljenih podataka, čime se ubrzava rad algoritama učenja, poboljšava kvalitet podataka i povećava tačnost naučenog znanja. Pri odabiru atributa, postoje dva pristupa koji se razlikuju po tome što se želi postići. Prvi pristup je pronalaženje podskupa atributa koji su korisni za izgradnju kvalitetnog modela. Drugi pristup je nalaženje ili rangiranje svih potencijalno važnih atributa. Korištenje podskupa od potencijalno važnih atributa za izgradnju modela pokazuje se kao suboptimalni pristup kod izrade nekih modela. S druge strane, pitanje korisnosti naspram važnosti atributa može biti veoma zanimljiv.

11 L. Cao, G. Weiss, S.P. Yu, A brief introduction to agent mining, *Autonomous Agents and Multi-Agent Systems*, Vol. 25, No. 3, 2012, str. 419–424.

3. Primena tehnika veštačke inteligencije pri analizi podataka u policiji

Napredak u tehnologiji analize velikih količina podataka predstavlja osnovu za razvoj relativno nove oblast poznate kao analiza podataka u policiji. U mnogim policijskim upravama, analiza podataka se svodi na mapiranja zločina za potrebe komandnog osoblja i izrada statistike o samom kriminalu. U drugim slučajevima, analiza podataka u policiji može da predstavlja fokusiranje na analizu raznih policijski izveštaja i informacija o osumnjičenima kao vid pomoći istražiteljima. Analiza podataka u policiji je proces analitičkog istraživanja kriminalnih delatnosti, pojedinaca ili grupa i kao takva čini jednu od važnih aplikacija današnjeg *data mining*-a. Primena *data mining*-a u kriminalističkoj analizi svodi se na otkrivanje obrazaca i šablona, kreiranju prognoze, pronalaženju odnosa i mogućih objašnjenja, mapiranje kriminalnih mreža kao i identifikovanje mogućih osumnjičenih. Tehnike klasteringa baziraju se na pronalaženje odnosa između različitih vrsta kriminala i ranije poznatih zajedničkih karakteristika kriminalnih atributa. Pravila udruženja svode se na pronalaženju pravila u kriminalnim podacima na osnovu kojih se identifikuju obrasci koji pomažu donosiocima odluka o bezbednosti društva da preduzmu adekvatnu akciju prevencije.

3.1. Sistem za detekciju upada – IDS

Multilevel klastering upozorenje i inteligentni klaster modeli upozorenja[12] dobre su tehnike za smanjenje broja upozorenja. Kompleksnost modela, s druge strane, može degradirati karakteristike samog sistema za upozoravanje. Autor Metju (Mathew) i saradnici uložili su veliki trud u tom smislu, predstavljajući tehniku za razumevanje multilevel napada[13] uz praćenje napada baziranog na vizuelizaciji raznovrsnih tokova događaja. Koristili su događaj korelacije koji prati napad kako bi se utvrdio vremenski odnos između raznovrsnih događaja. Navedeni pristup je bio koristan samo da bi se razumele faze u multilevel napadima, ali ne i u predviđanju ponašanja korisnika.

Veoma je važno da se odredi profil i namere napadača kako bi se zaštitila mreža. Zbog toga postoji potreba da se nađe efikasan način identifikacije vrste napadača. Sistemi za detekciju upada (*Intrusion Detection Systems – IDS*) kao što je Snort, besplatan IDS za operativne sisteme Linux i Windows, pomaže u otkrivanju jedne faze upada[14], ali ne i u otkrivanju *multistage* napada (napad sa nekoliko uzastopnih faza) i ponašanje napadača. Razlog tome je veliki broj upozorenja, nedostatak odgovarajućeg modela koji može da detektuje *multistage* napade i nepostojanje metoda koje su u stanju da povežu *multistage* napade sa ponašanjem napadača. Zato su razvijeni drugi sistemi na bazi veštačke inteligencije koji mogu da vrše predikcije (predviđanja) i koji mogu da odrede

12 S. Vaughn, *Multilevel Alert Clustering for Intrusion Detection Sensor Data*, Fuzzy Information Processing Society, USA, 2005.

13 S. Mathew, D. Britt, R. Giomundo, S. Upadhyaya, S. Sudit, *Real-time Multistage Attack Awareness Through Enhanced Intrusion Alert Clustering*, In Situation Management Workshop (SIMA 2005), MILCOM 2005, Atlantic City, NJ, 2005.

14 P. Čisar. Sistem za detekciju upada u mrežnu infrastrukturu, *NBP – Journal of Criminalistics and Law*, Kriminalističko-policijska Akademija, Beograd, Vol. 18(1), 2013, str. 113–128.

profile napadača na osnovu njegovog ponašanja. Multilevel klastering upozorenje i inteligentni klaster modeli upozorenja su veoma korišćene tehnike za smanjenje broja upozorenja.

Ubrzani razvoj internet tehnologija doneo je sve više sofisticiranih upada sa kojima je bezbednost mreže postala veliki izazov današnjice. Napadači mogu imati različite namere, a svaki napad može da ima različite nivoe. Sagledavanje ponašanja napadača u tom prostoru svakako je važno kako bi se razumeli rizici. U studiji slučaja[15] sprovedenoj od strane vlade SAD svi napadači u sajber prostoru svrstani su u devet različitih grupa[16]:

1) *amateri* – ova grupa napadača nema mnogo znanja; oni to čine radi zabave;

2) ***kriminalci*** – nastoje da napadnu sisteme za ostvarivanje novčane dobiti; oni koriste spamove, fišinge kao i špijunske/maliciozne softvere za krađu identiteta i prevare na mreži;

3) ***insajderi*** – u velikom broju slučajeva poseduju detaljno poznavanje sistema žrtve koji im omogućava neograničen pristup i daje mogućnost izazivanja štete sistemu ili krade osetljivih podataka;

4) ***fišeri*** – koriste šeme za krađu identiteta u pokušaju da sakupe informacije u svrhu ostvarivanja novčane dobiti;

5) ***države*** – strane obaveštajne službe koriste IT sredstva za prikupljanje informacija i špijunažu; ovo može biti usmereno na druge države (prijateljske i neprijateljske) ili na nedržavne pretnje;

6) ***hakeri*** – vrše upad u mreži preko neovlašćenog pristupa koji zahteva dobro poznavanje računara i veština programiranja;

7) ***teroristi*** – pokušavaju da unište, onesposobe ili iskoriste ključnu infrastrukturu u cilju ugrožavanja opšte nacionalne bezbednosti;

8) ***bot-mrežni operateri*** – haker-operateri na mreži koji preuzimaju veliki broj računara, a koji se zatim koriste za koordinaciju napada, fišer prevare, spamovanje ili malver (zlonamerne) napade;

9) ***autori špijunskog i zlonamernog softvera*** – pojedinci ili organizacije koji sa zlom namerom sprovode napade na korisnike koristeći i rasturajući špijunski i maliciozni softver.

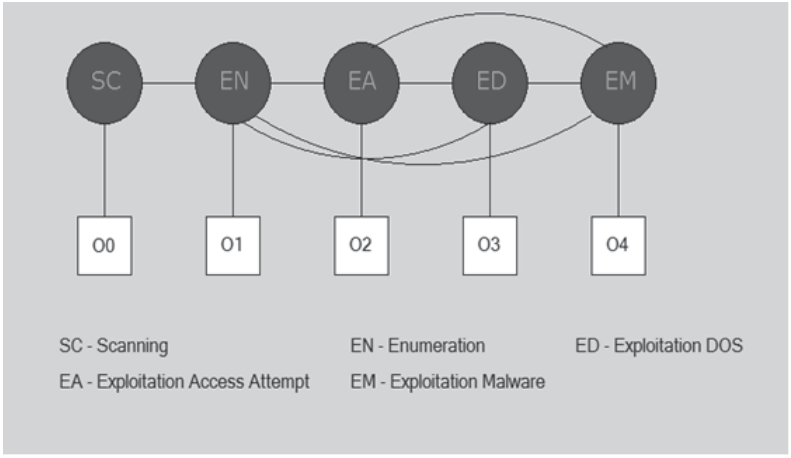
Za predviđanje ponašanja napadača se obično koristi HMM, algoritam mašinskog učenja, kako bi se analiziralo ponašanje napadača po nekim definisanim pravilima za svaku vrstu napadača. U njima je definisano pet faza[17], koje ujedno predstavljaju stanja u HMM: skeniranje, prebrojavanje, pokušaj pristupa, malware pokušaj i servis odbijanja.

15 United States. Government Accountability Office. *Cyberspace: United States Faces Challenges in Addressing Global Cybersecurity and Governance*. Washington D.C. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc301480.

16 S.B. Buckland, F. Schreier, T. H. Winkler, *Demokratsko upravljanje: Izazovi sajber bezbednosti*, FBD, Forum za bezbednost i demokratiju, Beograd, str. 48. http://www.scribd.com/doc/60984092/Cyber-Safety, 2010.

17 R. Katipally, L. Yang, A. Liu, *Attacker Behavior Analysis in Multi-stage Attack Detection*, The Proceedings of Cyber Security and Information Intelligence Research Workshop, ACM Digital Library, Oak Ridge, TN, 2011.

- *Skeniranje*. Napadač pokušava da prikupi informacije o ciljnom sistemu. Vrši se posmatranje ICMP PING komande, jedne od komandi ICMP (*Internet Control Message Protocol*) paketa. Ping je komanda IP protokola koja primaocu nalaže da odgovori na nju i vrati pošiljaocu sadržaj koji je dobio u istom paketu. Koristi se za merenje brzine protoka odziva internet veza.
- *Prebrojavanje*. Napadač pokušava da pronađe ranjivost ciljnog sistema. Vrši se posmatranje portova (pristupna vrata kroz koja informacije dolaze ili odlaze od računara) u servisima za pričanje kao što je CHAT_MSN.
- *Pokušaj pristupa*. Napadač pokušava da dobije pristup resursima ciljnog sistema. Vrši se posmatranje SQL verzije pokušaja prekoračenja. Prekoračenje bafera je anomalija do koje dolazi kada proces u bafer upiše podatke koji su zbirno veći od veličine tog bafera. To se može postići preko SQL naredbi u *web* aplikacijama.
- *Odbijanje usluga.* (*Denial of Service – DoS*). Napadač pokušava da uskrati uslugu drugim korisnicima. Vrši se posmatranje jedne od najslabijih tačaka Windowsa, NetBIOS SMB – DS Trans Maks Param DOS. DoS izaziva prestanak rada servisa ili programa, čime se drugima onemogućava rad sa tim servisima ili programima. DoS napad se najlakše izvršava na transportnom sloju – slanjem velikog broja SYN paketa (*TCP CONNECTION REQUEST*).
- *Malware pokušaj*. Napadač pokuša da izvrši svoj kod na ciljnom sistemu. Vrši se posmatranje asemblerske naredbe „nop" (*no operation*), naredba bez dejstva: SHELLCODE_k86_NOOP koja omogućava pokretanje shell programa pod operativnim sistemom na procesoru serije x86 (Intel/AMD).



Slika 5: Faze napada u IDS-u predstavljene kao stanja HMM

Mapa upozorenja/posmatranja predstavlja jedno od pet skrivenih stanja sistema. Sva upozorenja su klasifikovana u pet različitih grupa u zavisnosti od vrste upozorenja. Na primer, upozorenje na ICMP ping obično se smatra za vrstu skeniranja, dok se upozorenje od *shellcode X86 INC EXC NOOP* smatra za eksploataciju *malware* tipa. Kreirano je 88 pravila po kojima su označena imena stanja za svaku vrstu upozorenja.

Tabela 1: Ponašanje napadača na osnovu pripadajuće grupe

Grupe napadača	Ponašanje
Amater	Skeniranje + prebrojavanje
Insajder, fišer, Spyware/Malware, Botnet (ISBN)	Pokušaj pristupa + odbijanje usluga + malware pokušaj
Kriminalne grupe, teroristi, hakeri, države (CTHN)	Skeniranje + prebrojavanje + pokušaj pristupa + malware pokušaj
Teroristi, hakeri (TH)	Skeniranje + prebrojavanje + odbijanje usluga
Teroristi, hakeri, kriminalne grupe (THC)	Skeniranje + prebrojavanje + pokušaj pristupa + odbijanje usluga + malware pokušaj

Data mining je našao široku primenu u oblasti kriminalistike. Može se primeniti i u ostalim oblastima policije, tamo gde se raspolaže velikim količinama podataka čijom se analizom mogu otkriti određena pravila, zakonitosti i veze. Baze podataka kriminalaca sadrže informacije o samim kriminalcima, prestupnicima, žrtvama kao i vozilima koja su korišćena u zločinu. Među ovim zapisima su grupe krivičnih dela koja se mogu pripisati serijskim kriminalcima, koji su odgovorni za više krivičnih prestupa i obično ispoljavaju određene specifičnosti u svom izvršenju krivičnog dela (silovanje, ubistvo, razbojništvo, itd.), kao i primenu specifičnih metoda za sprovođenje svog zločina.

3.2. Inteligentni sistemi u policiji

U Indiji se već duže vreme koristi inteligentni sistem policije kao alat koji ima neprocenjivu vrednost u promeni trenutnog stanja kriminala. Ovaj sistem predstavlja jednu vrstu inteligentnog alata koji indijska policija koristi za otkrivanje i sprečavanje kriminala[18]. Prvi zadatak u primeni klaster algoritma za otkrivanje šablona kriminala jeste predviđanje veličine stanovništva grada. Proračun kriminala po glavi stanovnika pomaže da se odredi statistika kriminala u određenoj populaciji stanovnika. Međutim, u nekim evidencijama nedostaju jedna ili više vrednosti. Još gora situacija je nedostatak vrednosti „veličina stanovništva grada", što znači da nije bilo statistike po glavi stanovnika kao kompletnog zapisa. U nekim gradovima ne postoje podaci o prijavljenom stanovništvu u bilo kakvom obliku. Da bi se poboljšao proračun „godišnjeg proseka stope kriminala po glavi stanovnika", kao i da bi se obezbedilo otkrivanje svih „onih koji tu ne pripadaju", bilo je potrebno da se popune nedostajuće vrednosti. Osnovni pristup zasnivao se na to da se izvrši klasterizacija veličine stanovništva, formiraju klase iz klastera, a zatim klasifikuju evidencije sa nepoznatom veličinom stanovništva. Opravdanje za korišćenje tehnike klasteringa je to što klase iz klastera imaju

18 M. Gupta, B. Chandraand, M. P. Gupta, Crime Data Mining for Indian Police Information System, *Journal of Crime*, Vol. 2, No. 6, 2006, str. 43–54.

veću verovatnoću da predstavljaju stvarnu veličinu stanovništva u gradovima. Jednina veličine neophodna za klasterizaciju bila je veličine stanovništva svakog zapisa. Ove vrednosti su grupisane pomoću EM algoritma (algoritam je efikasna interativna procedura za estimaciju parametara na osnovu kriterijuma maksimalne verodostojnosti očekivanje – maksimizacija), a deset inicijalnih klastera bilo je izabrano zato što su oni proizveli klastere sa srednjom vrednošću koje predstavljaju kalkulacije po glavi stanovnika približno njihovim stvarnim vrednostima.

Polazeći od datog skupa objekata, grupisanje je proces otkrivanja klase pri čemu se objekti grupišu u klastere koji su unapred nepoznati. U tu svrhu koriste se dve tehnike klasteringa: *K-Means* algoritam i *DBScan* algoritam (*Density-Based Spatial Clustering Application with Noise*). Korišćenje jednog algoritma koji bi bio primenjiv na sve probleme, u praksi je nemoguće ostvariti. Proces usavršavanja inteligentnih alata ostvaruje se hibridnim rešenjima koja spajaju najbolje od nekoliko različitih pristupa i algoritama. Na taj način moguće je stvoriti hibridni genetski algoritam, gde je svaka jedinka lista sastavljena od svih parametara primitivnih funkcija u nekom algoritmu, a postupak optimizacije se svodi na nalaženje najboljeg skupa parametara za svaku primitivnu funkciju u tom grafu. Hibridni genetski algoritam[19] dat u ovom inteligentnom sistemu policije grupiše podatke u m grupa pri čemu su dati:

1) ulaz – vrsta kriminala, broj klastera, broj ponavljanja;

2) izlaz – predikcije/klasteri C0, C1, C2 i C3 i to:

- C0: kriminal je stabilan ili opada; stopa seksualnog uznemiravanja je primarni zločin u pokretu; postoje manji incidenti – ubistvo iz koristoljublja, razbojništvo, pripreme za razbojništvo, silovanje, bračno nasilje i ubistvo iz nehata;
- C1: kriminal raste ili je u toku; neredi, prevare, falsifikati i okrutnost od strane muža i rodbine menjaju primarne stope kriminala; postoje manji incidenti – ubistva, kidnapovanja ili otmice drugih;
- C2: kriminal se generalno povećava; krađe su primarni kriminal sa tendencijom porasta; postoje niži incidenti zločina protiv imovine: provala i krađa;
- C3: malo zločina je u toku; ubistva, silovanja, izazivanje požara u pokretu; postoji manje promene zločina protiv imovine – provale i krađe – ali se koriste za demonstraciju makar nekih karakteristika klastera.

U rezultatima klasterizacije uočen je trend gradskog zločina za svaki tip kriminala u svakoj godini. Dalje, malom modifikacijom klastering semena razna stanja su grupisana kao zone visokog, srednjeg i niskog nivoa kriminala. Na osnovu ovih homogenih grupa, efikasnost policijske jedinice, tj. države može se izmeriti, a metod je dat kao:

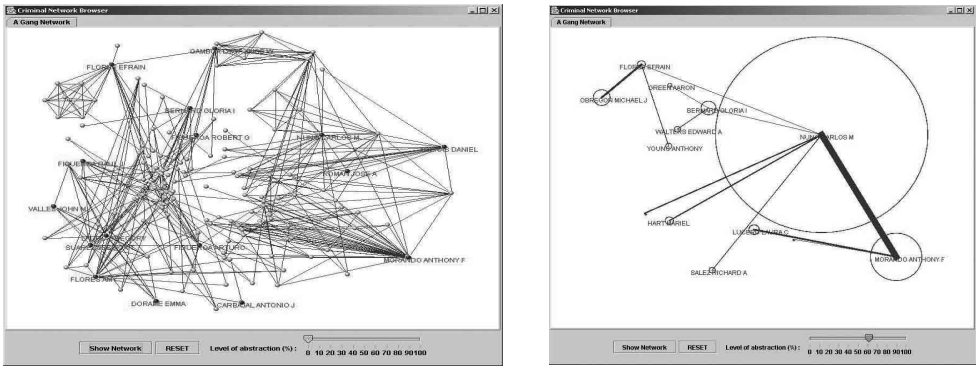
Izlazna funkcija stope kriminala = 1/stopa kriminala (4)

19 A. Malathi, Dr. S. Santhosh Baboo, An Enhanced Algorithm to Predict a Future Crime using Data Mining, *International Journal of Computer Applications*, Vol. 21, No. 1, 2011, str. 1–6.

gde se stopa kriminala dobija deljenjem ukupne gustine kriminala države sa ukupnim brojem stanovnika te države, tako da se stanje policije prikazuje kao učinkovito ako je stopa kriminala niska, odnosno ako je izlazna funkcija stope kriminala visoka.

Osnovni metod podrazumeva grupisanje stanja koji imaju isti trend kriminala, a potom klasifikovanje novog zapisa pomoću informacije „sledeća godina". Ovo je kombinacija koja je sa skromnim podacima u stanju da stvori klasifikator koji će predvideti buduće trendove kriminala. Na osnovu klaster rezultata, klasifikacijski algoritam se primjenjuje da predvidi budući trend kriminala. Klasifikacija se izvodi kako bi se našao odgovarajući klaster za sledeću godinu. To nam omogućuje da napravimo prediktivni model za predviđanje sledeće godine pomoću podataka evidencije ove godine. U ovu svrhu se koristi C4.5 algoritam stabla. Generalno, tehnika stabla se koristi za predviđanje nepoznatih trendova kriminala u sledećoj godini. Eksperimentalni rezultati pokazuju da je ova tehnika predviđanja tačna i brza.

Analizirajući bazu podataka 272 policijske stanice u Tusonu, predstavnici Univerziteta u Arizoni i Hong Kongu, analizirali su izveštaje o 164 zločina[20] počinjena u periodu od 1985. do 2002. godine. Korišćen je prostorni koncept/tehnika, kako bi se utvrdile veze između podgrupa i definisala mreža bandi. Stepen povezanosti između pojedinih podgrupa meren je prema tome koliko su se često njihova imena zajedno pominjala u istim zločinima. Korišćena je metoda klasteringa, kako bi se cela kriminalna mreža na tom području podelila na podgrupe, tj. bande, i *block-modeling* pristup za utvrđivanje veza i komunikacije između istih. Isti pristup je korišćen i za utvrđivanje vođa bandi. Dobijeni su sledeći rezultati prikazani na slici 6. Levi deo slike pokazuje da je korišćenjem DM utvrđeno 16 vođa kriminalnih grupa,čija su imena obeležena crvenim slovima, kao i mesta na kojima su se dešavali zločini. S desne strane slike prikazuju se kriminalne podgrupe, ukupno njih 16. Grupe su nazvane po imenima njihovih lidera. Obim krugova pokazuje broj članova koji pripadaju svakoj grupi, tj. veličinu grupa i područje njihovog delovanja. Debljim linijama prikazano je između kojih podgrupa postoji veći obim komunikacije i čvršće veze.



Slika 6: Analiza kriminalnih mreža

20 H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, M. Chau, Crime data mining: A general framework and some examples, *Computer*, Vol. 37(4), 2004, str. 50–56.

Nakon sprovedenog istraživanja, dobijene rezultate su analizirala tri eksperta iz Policijske stanice u Tusonu, koji su upoređivali ove podatke sa informacijama koje imaju sa terena. Eksperti su potvrdili rezultate dobijene DM. Podaci realno prezentuju stvarno stanje na terenu. Podaci o dve najveće kriminalne podgrupe su takođe bili tačni, kao i jačina veza koja postoji između njih. Te dve podgrupe su smatrane za dve najveće mreže za prodaju narkotika u regionu. Imena vođa grupa su takođe ispravna, a ispostavilo se da su vođe dve najveće podgrupe dobri prijatelji. Eksperti su potvrdili da DM sistem koji je razvijen tokom ovog istraživanja može u velikoj meri da pomogne u definisanju kriminalnih mreža i njihovih struktura, ali i da pomogne u sprečavanju određenih kriminalnih aktivnosti i zločina. Pomoću ovakvih modela može se u velikoj meri ograničiti komunikacija između podgrupa i suziti područje njihovog delovanja.

Zaključak

U današnjem svetu, kriminalci koriste pametnu tehniku kako bi maksimalno iskoristili sve postojeće moderne tehnologije i metode u izvršenju zločina. Ovo im, takođe, olakšava da deluju uzduž i popreko preko cele teritorije države. Visokotehnološki odnosno sajber kriminal ne predstavlja nikakvu novinu kod nas, a još manje u svetu. Kako je danas moguće biti sajber kriminalac iz bilo kog dela sveta, značajno je otežano lociranje prestupnika. Ako se susrećemo sa takvim izazovima, radi kontrole kriminala i održavanja javnog reda, moramo efikasno stvarati baze podataka o zločinima i zločincima u digitalnom obliku, tako da se korišćenje inteligentnih tehnika u informacionim sistemima policije više ne može zanemariti. *Data mining* u računarskoj nauci jeste polje specijalizovano za prikupljanje implicitnih informacija koje se distribuiraju preko uskladištenih zapisa podataka ili postoje kao odnosi među grupama zapisa. Tipični primeri korišćenja veštačke inteligencije postoje baš u oblasti policije kao što su praćenje šema zločina, predviđanje kriminalnog ponašanja pojedinaca, lociranje zločinaca i sl.

Internet trgovina sve više uzima maha, a tako je i sa zloupotrebama platnih kartica i drugim oblicima sajber kriminala. Vrednost sajber kriminala nadmašuje trgovinu narkoticima. Efikasna borba protiv visokotehnološkog kriminala obuhvata kako kratkoročne mere tako i dugoročnu kontrolu. Kada se govori o kratkoročnim merama, one su uglavnom u domenu prevencije i obuhvataju pomenutu analizu rizika – kako iz ugla organizacije tako i iz ugla napadača. Analiza omogućava identifikaciju kritičnih tačkaka i sprovođenje mera za smanjenje rizika, kao i implementaciju sistema otkrivanja i sprečavanja napada. Kriminalne grupe često razvijaju svoje mreže, u okviru kojih se organizuju podgrupe i bande kako bi se izvodile razne vrste kriminalnih aktivnosti. Tehnike DM se mogu uspešno iskoristiti za identifikovanje ovih podgrupa, tj. bandi i utvrđivanje načina komuniciranja i interakcije među njima, kako bi se sprečile mnoge ilegalne aktivnosti i zločini.

Literatura

1. Russell, S. Norvig, P.; *Artificial Intelligence: A Modern Approach*, Prentice-Hall, 2010.
2. Cao, L., Weiss, G., Yu, S. P.; A brief introduction to agent mining, *Autonomous Agents and Multi-Agent Systems*, Vol. 25, No. 3, 2012, str. 419–424.
3. Vaughn, S.; *Multilevel Alert Clustering for Intrusion Detection Sensor Data*, Fuzzy Information Processing Society, USA, 2005.
4. Mathew S., Britt, D., Giomundo, R., Upadhyaya, S. Sudit, S.; *Real-time Multistage Attack Awareness Through Enhanced Intrusion Alert Clustering*, In Situation Management Workshop (SIMA 2005), MILCOM, 2005, Atlantic City, NJ, 2005.
5. Čisar, P.; Sistem za detekciju upada u mrežnu infrastrukturu, *NBP – Journal of Criminalistics and Law*, Kriminalističko-policijska Akademija, Beograd, Vol. 18(1), 2013, str. 113–128.
6. United States Government Accountability Office. *Cyberspace: United States Faces Challenges in Addressing Global Cybersecurity and Governance*. Washington D.C. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc301480.
7. Buckland, S. B., Schreier, F., Winkler, T. H.; *Demokratsko upravljanje: Izazovi sajber bezbednosti*, FBD, Forum za bezbednost i demokratiju, Beograd, 2010, http://www.scribd.com/doc/60984092/Cyber-Safety.
8. Katipally, R., Yang, L. Liu, A.; *Attacker Behavior Analysis in Multi-stage Attack Detection*, The Proceedings of Cyber Security and Information Intelligence Research Workshop, ACM Digital Library, Oak Ridge, TN, 2011.
9. Gupta, M., Chandraand B. Gupta, M. P.; Crime Data Mining for Indian Police Information System, *Journal of Crime*, Vol. 2, No. 6, 2006, str. 43–54.
10. Malathi. A., Santhosh Baboo, Dr. S.; An Enhanced Algorithm to Predict a Future Crime using Data Mining, *International Journal of Computer Applications*, Vol. 21, No. 1, 2011, str. 1–6.
11. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., Chau, M.; Crime data mining: A general framework and some examples, *Computer*, Vol. 37(4), 2004, str. 50–56.
12. Mitchell, T. M.; *Machine Learning*, McGraw-Hill, Inc., New York, USA, 1997.
13. Ilić, V.; Systems based on agent's technology, *InfoM – Journal of Information Technology and Multimedia Systems*, No. 3–4, 2002, Belgrade.
14. Nguyen T. N., Lakhmi C. J. (Eds.); *Intelligent Agents in the Evolution of Web and Applications*, Vol. 167, 2009.
15. Vuletić, D.; *Napadi na računarske sisteme*, Vojnotehničkih glasnik / Military Technical Courier, Vol. LX, No. 1, 2012, str. 235–249.
16. Fosler-Lussier, E.; Markov Models and Hidden Markov Models: A Brief tutorial, Technical Report (TR-98-041), 1998.
17. Rabiner, L. R.; A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, 1989, str. 257–286.

18. Janičić, P., Nikolić, M.; *Veštačka inteligencija*, Matematički fakultet, Beograd, 2010.
19. Šimić, G., Jeremić, Z., Kajan. E., Ranđelović, D.; Framework for Delivering e-Government Support, *Acta Polytechnica Hungarica*, Vol. 11, No. 1, 2014, str. 79–96.

ARTIFICIAL INTELLIGENCE IN PROCESS OF COLLECTING AND ANALYZING DATA WITHIN POLICE WORKS

Kristijan Kuk
Academy of Criminalistic and Police Studies, Belgrade

Summary: Complex real problems increasingly require intelligent systems that combine knowledge, techniques and methodologies from various sources. Intelligent systems based on artificial intelligence techniques that are associated with the behavior of people can perform the processes of learning, reasoning and solving all kinds of problems. Such systems, which automatically can perform tasks set by the user or other software, today thankfully called intelligent agents. Independent, intelligent agents on the Internet can be very successful to perform some search work on behalf of and for the needs of different users. For efficient collection, manipulation and management of data, such software can be very interesting from the standpoint of intelligent data analysis in many areas the police. Analysis of the data collected by an intelligent agent (a software robot-bot) can be successfully utilized, among many jobs in the police, and in the field of crime and in particular manifestation of cyber-crime, traffic safety, emergencies, etc. To make the collection and analysis of data from criminal activities on the Internet effective, it is necessary to examine the existing artificial intelligence techniques to be used for the conclusion of the intelligent agents. On the other hand, using of methods of artificial intelligence in finding data along with intelligent data analysis (data mining) should be used, which has found wide use in the area of business, economics, mechanics, medicine, genetics, transport etc.

Keywords: artificial intelligence, intelligent agent, software robot-bot, data mining, predictive policing.