Pregled najizazovnijih problema u data mining istraživanju

Mladen Vukašinović
Postdiplomski – Magistarski studij
Univerzitet Mediteran - Fakultet za Informacione Tehnologije
Podgorica, Crna Gora
mladen©ac.me

Sadržaj— **Ovaj rad predstavlja pregled istraživanja koje je sprovedeno da bi se rangirali najvažniji problemi u Data Mining istraživanju. Ovdje su sumirani indentifikovani najvažniji problemi sa kojima se data mining istraživanje susreće. Pri odabiru kriterijuma o tome šta se smatra važnim i vrijednim temama za buduća istraživanja u data miningu, su korišćena mišljenja najaktivnijih istraživača u oblastima data mininga i učenju mašina.**

Ključne riječi- data mining; overview of research; most important problems;

I. UVOD

Data mining [1] je nova disciplina koja ima za cilj da filtrira podatke u bazama podataka, da ih sumira, i pronalazi obrazce. Data mining ne treba posmatrati kao jednostavnu-vremensku vježbu. Ogroman broj podataka kolekcija može se analizirati i ispitati na neograničen broj načina. Kako vrijeme odmiče, nove vrste objekata i dezena mogu privući pažnju, i mogu biti vrijedni u traženju potrebnih podataka.

Data mining je, zbog dobrih razloga, nedavno privukao veliku pažnju: to je nova tehnologija, rješava nove probleme, sa velikim potencijalom za komercijalna i naučna otkrića. Međutim, ne treba očekivati da pruži odgovore na sva pitanja. Kao i sva otkrića procesa, uspješan data mining je sposobnost sretnog slučajnog otkrića. Dok pronalaženje podataka obezbjeđuje korisne alate, to ne znači da će neizbježno dovesti do važnih, interesantnih, ili vrijednih rezultata. Moramo da pazimo od pretjeranog preuveličavanja vjerovatnog ishoda. Ali, potencijal je tu.

II. RAZVOJ OBJEDINJAVANJA TEORIJA DATA MINING

Nekoliko ispitanika smatra da je trenutno stanje data mining istraživanja previše "ad hoc." [2] Mnoge tehnike su dizajnirane za individualne probleme, kao što su klasifikacije ili grupisanje, ali nema ujedinjenja teorija. Međutim, teorijski okvir koji objedinjuje različite data mining zadatke, uključujući grupisanje, klasifikacije, udruženje pravila, itd, kao i različite data mining pristupe (kao što su statistike, mašinsko učenje, baza podataka, itd), da bi pomogli na terenu i obezbjedili osnovu za buduća istraživanja.

Takođe, postoji mogućnost i potreba za data mining istraživače da riješe neke dugogodišnje probleme u statističkim istraživanjima, kao što su vjekovima stari problemi izbjegavanja lažne korelacije. Ovo se ponekad dovodi u vezu sa problemom istraživanja za "Duboka znanja", koji je skriven i uzrok mnogih primjedbi.

III. SKALIRANJE ZA VISOKO DIMENZIONALNE PODATAKE I VELIKE BRZINE TOKOVA PODATAKA

Jedan izazov je kako da dizajnirate klasifikatora da rukuje ultra visokom klasifikacijom dimenzionalnog problema. . Postoji jaka potreba sada da se izgrade korisni klasifikatori sa stotinama milionima ili milijardama karakteristika, za aplikacije kao što su tekstualni podaci i ljekovi za bezbjednost analiza.

Ovakvi problemi često počinju sa desetinama hiljada osobina, ali i sa interakcijama između funkcija, tako da se broj podrazumijevanih funkcija dobija ogromnom brzinom. Jedan od važnih problema je tok data mininga u izuzetno velikim bazama podataka (Na primjer 100 TB). Satelitski i računarskih mreža podaci mogu lako da budu ovih veličina. Međutim, pronalaženje podataka današnjim tehnologijama je još uvijek suviše sporo da bi se nosilo sa podacima ovog razmjera.

Pored toga, data mining treba da bude kontinuiran, onlajn proces, a ne povremeni jednokratani proces. Organizacije koje mogu da urade ovo će imati odlučujuću prednost u odnosu na one koje to ne čine. Ogromni podaci predstavljaju novi izazov za istraživanje podataka data mininga. Jedan poseban primjer je iz veoma brzih mreža gdje se nadamo da ćemo naći informacije za razne svrhe, uključujući utvrđivanje anomalnih događaja koji možda ukazuju na napade bilo koje vrste.

Tehnički problem je kako izračunati modele preko striminga podataka, koji se prilagođavaju promjenama u okruženju iz koje su podaci izvučeni. To je problem "koncept Drift" ili "odstupanje okoline". Ovaj problem je naročito težak u kontekstu velikih striming podataka. Kako možemo izračunati jedan model koji je tačan i veoma efikasan?

Na primjer, ne možemo pretpostaviti da imaju veliku računarsku snagu i sredstva za skladištenje mnogo podataka, odnosno da upoređuju podatake više puta. Dakle, inkrementalni istraživački i efikasani model ažuriranja za održavanje tačnog modeliranja trenutnog protoka su veoma veliki problemi.

Tokovi podataka mogu takođe dolaziti iz senzorske mreže i RFID aplikacije. U budućnosti, RFIDs (Radio frequency identification) će biti ogroman prostor, i analiza ovih podataka je od ključnog značaja za svaki uspjeh.

IV. DATA MINING SEKVENCI I VREMENSKOG TOKA PODATAKA

Sekvencijalni i vremenski data mining [3] je i dalje važan problem. Uprkos napredku u drugim srodnim oblastima, kako da efikasno klastere, klasifikuje i predvidi trendove ovih podataka je i dalje važna otvorena tema. Poseban izazov je problem šuma u podacima sa vremenskim tokom. To je važno otvoreno pitanje za raspravu. Mnogi vremenski tokovi podataka koji treba da se obrade su kontaminirani šumom, što otežava tačne kratkoročne i dugoročne prognoze. Primjeri ovih aplikacija uključuju predviđanja finansijskih vremenskih serija i seizmičke vremenske serije. Iako se tehnika obrade signala, kao što su talasne analize i filtriranje, može primjeniti za uklanjanje šuma, često zaostaje uvođenje filtriranja podataka. Takvo zaostajanje smanjuje tačnost predviđanja, jer prediktori moraju prevazići zaostajanje prije nego što mogu da predvide šta će se desiti. Postojeći data mining metodi takođe imaju poteškoća u rukovanju podataka sa šumom i učenje smislenih informacije iz tih podataka.

Neka od ključnih pitanja koje treba riješiti u dizajnu praktičnih data mininga za podatke sa šumom su:

- Pretraživanja koja koriste informacije: Upotreba pogrešnih, suvišnih, ili suviše malih kriterijuma pretrage; eventualno nedosljedne informacije iz raznih izvora; semantička analiza (meta) informacija; asimilacija informacija u inpute predskazivača.
- Učenje data mininga za izmjenu izbora kriterijuma informacija: Dioba pristrasnosti u povratne informacije; razvoju pravila za pretraživanja koja prikupljaju informacije; razvoj pravila za informacije koje mogu da asimiluju podatke.
- Predviđač da predvidi buduće trendove: Korišćenje kvalitativnih informacija; multiobjektivna optimizacija u otvorenom obliku.

V. DATA MINING KOMPLEKSA SAZNANJA IZ KOMPLEKSIH PODATAKA

Jedan važan tip [4] kompleksnih znanja je u obliku grafikona. Nedavna istraživanja su dotakla temu otkrivanja grafikona i strukturiranih obrazaca iz velikih baza podataka, ali jasno je da treba mnogo više toga da se uradi.

Drugi oblik kompleksnosti je od podataka koji nisu IID (Nezavisni i identično raspodijeljeni). Ovaj problem se može javiti kada se istražuju podaci iz više izvora. U većini oblasti, predmeti interesovanja nisu nezavisni jedan od drugog, i nisu jednog tipa. Potrebni su nam data mining sistemi koji mogu da stabilno istražuju strukture odnosa između objekata, kao što su povezane veb stranice, socijalne mreže, metaboličke mreže u ćeliji, itd. Ipak, još jedan važan problem je kako da se istraže nerelacioni podaci. Velika većina podataka većine organizacija je u tekstualnoj formi, a ne u bazi podataka i više podataka složenih formata, uključujući slike, multimedije i veb podatke. Stoga, postoji potreba za proučavanjem metoda data mininga koje prevazilaze klasifikaciju i grupisanje.

Nekoliko zanimljivih pitanja obuhvataju kako da se bolje automatski obavi sumiranje teksta i kako da se prepoznaju kretanja objekata i ljudi sa veb i bežičnih logova podataka, kako bi se otkrilo korisno prostorno i vremensko znanje. Sada postoji jaka potreba za integrisanje zaključivanja data mininga i znanja. To je važno za buduće teme. Konkretno, jedna važna oblast je da ugrade pozadine znanja u data mining. Najveći jaz između onoga što data mining sistemi mogu da urade danas i ono što smo željeli da urade je da oni ne mogu da donose odluke na osnovu informacija koje imaju iz stvarnog svijeta nego jedino mogu da dobijene podatke proslijede korisnicima. Obavljanje ovih zaključaka, i automatizovanje cijele data mining petlje, zahtijeva predstavljanje i korišćenje saznanja o svijetu u okviru sistema. Jedan od važnih primjena integracije je da se ubrizgavaju informacije i poslovna saznanja u procesu otkrivanja znanja.

U vezi sa data mining kopleksnim saznanjima, tema istraživanja interesantih saznanja i dalje je važna. U prošlosti, nekoliko istraživača su rešavali taj problem iz različitih uglova, ali mi još uvijek nemamo veoma dobro razumijevanje onoga što je interesanto za krajnjeg korisnika.

VI. DATA MINING U MREŽI PODEŠAVANJA VI-I ZAJEDNICE I DRUŠTVENE MREŽE

Današnji svijet je povezan kroz mnoge vrste veza. Ove veze su: veb stranice, blogovi, elektronska pošta i druge. Mnogi ispitanici smatraju istraživanje zajednice kao i društvenih mreža veoma važnom temom. Strukture zajednice su važne osobine društvene mreže. Identifikacija problema sama po sebi predstavlja jedan veliki izazov.

Prvo, veoma je važno da imaju pravu karakterizaciju pojma "Zajednica" koji treba da se otkrije.

Drugo, entiteti-čvorovi koji su uključeni, su distribuirani u realnu-životnu aplikaciju, a samim tim i poželjno distribuirano sredstvo se identifikuje.

Treće, snimak na bazi podataka možda neće moći da prikaže pravu sliku, ono što je najvažnije leži u lokalnim vezama (npr. priroda i učestalost lokalnih interakcija) između entiteta-čvorova.

Pod ovim okolnostima, naš izazov je da se shvati (1) statičke mreže struktura (npr. topologija i klastera) i (2) dinamičko ponašanje (kao što su faktori rasta, robustnost, i funkcionalne efikasnosti). Slični izazovi postoje u bio-informatici, kao što smo trenutno skrenuli pažnju na dinamičke studije regulatornih mreža.

Pitanja u vezi sa ovim problemom je ono zašto su lokalni algoritmi-protokoli neophodni u cilju otkrivanja (ili obliku) zajednica u odozdo-nagore načinu (kao u realnom svijetu).

Konkretno pitanje je sledeće. E-mail razmjena u okviru organizacije ili u sopstvenom poštanskom sandučetu tokom dugog vremenskog perioda može da se istraži i da se pokaže kako različite mreže u uobičajenim praksama ili druženjima

počinju da se pojavljuju. Kako možemo istražiti i dobiti korisno znanje od njih?

VI-II ISTRAŽIVANJA U I ZA RAČUNARSKE MREŽE-VELIKIH BRZINA I VELIKIH PROTOKA

Istraživanje problema mreža predstavlja ključni izazov. Mrežni linkovi se ubrzavaju i provajderi sada dodjeljuju brzine likova od jednog Giga Etherneta do 10 Giga Etherneta. Da bi mogli da otkriju anomalije (npr. nagle saobraćajne zastoje zbog DoS (Denial of Service) napada ili katastrofalnih događaja), provajderi će morati da budu u stanju da pokupe IP pakete na visokim link brzinama i da analiziraju velike količine (Nekoliko stotina GB) podataka svaki dan. One će ovdje morati da imaju visoko skalabilna rješenja. Dobri algoritmi su, dakle, potrebni da bi se otkrilo da li DoS napadi nepostoje. Takođe, nekada za napad koji je otkriven, kako da se napravi razlika između legitimnog saobraćaja i napada na saobraćaj, tako da je moguće da se odustane od paketa napada? Mi trebamo tehnike za:

- Otkrivanje DoS napada,
- Trag unazad da se sazna ko su napadači i
- Odbacivanje paketa koji pripadaju napadu na saobraćaj.

VII. DISTRIBUIRANI DATA MINING I ISTRAŽIVANJE VIŠETRUKIH PODATAKA

Problem distribuiranog data mininga je veoma važan u problemima mreža. U distribuiranom okruženju (kao što su senzor ili IP mreže), jedna se distribuirana sonda postavlja na strateškim lokacijama u okviru mreže. Problem ovdje je da budemo u mogućnosti da dobijemo podatke iz testne sonde, i da otkrijemo uopšteni obrazac za dobijanje podataka iz svih ostalih. Mogu postojati različiti modeli distribuiranog data mininga u ovom slučaju, ali jedan može da se koristi za prikupljanje podataka sa distribuiranih sajtova, i drugi model koji bi podjednako tretirao sve sajtove.

Ovdje je očigledno bio cilj da se smanji prikupljanje istih podataka sa različitih lokacija, kao i da se smanji prekoračenje u komunikacijama.

U distribuiranom istraživanju podataka, jedan je problem kako istražiti podatke iz više heterogenih izvora podataka: iz više baza podataka i multi-relacionih izvora.

Druga važna oblast su „strani upadi“ u data mining. U sve većem broju domena - za spam elektronske pošte, za borbu protiv terorizma, za detekciju upada u računarske sisteme, klikovi za spam, pretraživač spama, nadzor, otkrivanje prevara , prevare pri kupovini, dijeljenju fajlova, itd. - data mining sistema „neprijateljska lica“ namjerno mogu da manipulišu sa dobijenim podacima (npr. da proizvode lažne negative). Treba da razvijemo sisteme koji će to eksplicitno da uzmu u obzir , kombinovanjem data mininga i teorije.

VIII. DATA MINING ZA PROBLEME BIOLOŠKE I ŽIVOTNE SREDINE

Mnogi istraživači koji smo ispitali vjeruju da data mining bioloških podataka nastavlja da bude izuzetno značajan problem, kako za data mining istraživanja tako i za biomedicinske nauke.

Primjer istraživanja je pitanje kako da se primjeni data mining u dizajniranju vakcine HIV-a. U molekularnoj biologiji, više složenih data mining zadataka postoje, koji se ne mogu rješavati standardnim data mining algoritmima. Ovi problemi uključuju više različitih aspekata, kao što su DNK, hemijske osobine, 3D strukture i funkcionalna svojstva.

Takođe, postoji potreba da se ide dalje od bio-data mininga. Data mining istraživači bi trebalo da razmotre informacije, ekološke i zaštite životne sredine [5]. Jedna od najvećih zabrinutosti danas, koja će zahtijevati značajne napore data mininga, je pitanje o tome kako mi najbolje možemo da razumijemo i samim tim iskoristimo naše prirodno okruženje i resurse - pošto je svijet danas veoma "bogat resursima" !

Data mining će moći da napravi veliki uticaj u oblasti integrisane fuzije podataka i istraživanjama u aplikacijama ekološko-životne sredine, posebno kada se uključuju distribuirani-decentralizovani izvori podataka, na primjer autonomne mobilne senzorske mreže za praćenje klimatskih promjena i-ili vegetacijske promjene.

Na primjer, kako se data mining tehnologije mogu koristiti za proučavanje i saznanja faktora u posmatranom dupliranju broja uraganskih pojava tokom proteklih decenija, kao što je nedavno prijavljeno u naučnim listovima? Većina data mininga kojim se bavimo danas su brzo razvijena, na primjer, one iz akcija tržišta ili u gradskom saobraćaju. Postoji mnogo zanimljivih saznanja koja tek treba da bude otkrivena, kao što su pravilnosti u dinamičkim promjena i-ili koje se interakcije uzimaju u obzir. U tom smislu, jedan od izazova je danas kako se nositi sa problemom dinamičko vremenskih obrazaca, ponašanja, identifikacije i predviđanja u: (1) sistemima velikih razmjera (npr. Sistemi globalnih klimatskih promjena i potencijalne epidemije "ptičjeg gripa", itd.) i (2) ljudske centralne sisteme (npr. korisnički prilagođenu interakciju između čovjeka i računara ili P2P transakcije).

Vezano za prethodna pitanja o važnosti ovih aplikacija, postoji potreba da se fokusiramo na "ubice aplikacija" data mininga. Do sada tri važne i izazovne aplikacije za data mining su se pojavile: bioinformatika, personalizacija i bezbjednost aplikacija. Međutim, više istraživanja je potrebno kako bi se proširilo korišćenje ovih aplikacija, kao i njihov spisak.

IX. DATA MINING U VEZI SA PROBLEMOM PROCESA

Važne teme postoje vezano sa poboljšanjem alata i procesa data mininga kroz automatizaciju, kao što je predložilo nekoliko istraživača. Specifična pitanja uključuju kako da se automatizuje sastav data mininga i izgrade metodologije u data mining sistemima koje bi pomogle korisnicima da izbjegnu mnoge greške pri pronalaženju podataka. Ako bi se

automatizovali različiti procesi rada data mininga, bilo bi moguće da se smanji ljudski rad najviše što je moguće.

Još jedno važno pitanje je kako da se automatizuju čišćenja postojećih podataka. Mi danas možemo veoma brzo da gradimo modele i obrasce za pronalaženje podataka, ali 90 posto troškova je u pred-obradi (integraciji podataka, čišćenju postojećih, itd.). Smanjenje ovih troškova će imati mnogo veću isplativost nego dalje smanjenje troškova modela izgradnje i obrazaca-nalaza.

Drugi problem je kako da se izvrši sistematsko dokumentovanje čišćenja podataka. Još jedan problem je kako da se kombinuju vizuelne, interaktivne i automatske data mining tehnike zajedno. Primjećuje se da u mnogim aplikacijama, data mining ciljevi i zadaci nemogu biti u potpunosti određeni, naročito u istraživačkoj analizi podataka. Vizuelizacija pomaže da se sazna više o podacima i definišu-usavrše zadaci data mininga.

Takođe, postoji potreba za razvojem teorije i za interaktivna istraživanja velikih-kompleksnih baza podataka. Važno pitanje koje se postavlja je: Šta su kompozicioni pristupi za istraživanja upita sa multi-koracima ? Ogovori na ova pitanja tek treba da se istraže.

X. BEZBJEDNOST, PRIVATNOST I INTEGRITET PODATAKA

Nekoliko istraživača smatra zaštitu privatnosti u data miningu kao veoma važnu temu [5]. To jest, kako da se obezbijedi privatnost korisnika, dok se njihovi podaci istražuju. U vezi sa ovom temom je data mining radi zaštite bezbjednosti i privatnosti. Jedan ispitanik kaže da ako ne riješe pitanje privatnosti podataka, data mining će postati pogrdan termin široj javnosti.

Neki ispitanici smatraju problem procjene znanja integriteta veoma važnim. Citiram njihova zapažanja: "Data mining algoritmi se često primjenjuju na podatke koji su namjerno promjenjeni od prvobitne verzije, u cilju da se dezinformišu primaoci takvih podataka ili da se ugrozi privatnost i bezbjednost sa takvim prijetnjama. Takve modifikacije mogu da promjene, u nepoznatoj mjeri, znanje sadržano u originalnim podacima. Kao rezultat toga, jedan od izazova sa kojima se suočavaju istraživači je razvoj mjera, ne samo da se procjeni integritet znanja prikupljanjem podataka, već i mjera za procjenu integriteta znanja pojedinačnih obrazaca. Pored toga, problem procjene znanja integriteta predstavlja nekoliko izazova".

U vezi sa pitanjem znanja ocjene integriteta, dva najznačajnija izazova su: (1) razvijanje efikasnih algoritama za upoređivanje znanja sadržaja od dvije (prije i poslije) verzije podataka, i (2) razvoj algoritama za procjenu uticaja koji određene modifikacije podataka imaju na statistički značaj pojedinih obrazaca dobijenih od šire klase data mining algoritama.

Prvi izazov zahtjeva razvoj efikasnih algoritama i struktura podataka da bi se procijenio integritet znanja prikupljenih podataka. Drugi izazov je da se razvije algoritam za mjerenje uticaja koji modifikacija vrijednosti podataka ima na statističku značajnost otkrivenog obrazca, iako to može biti nesprovodljivo za razvijanje globalnih mjera za sve data mining algoritme.

XI. BAVLJENJE NESTATIČKIM, NEURAVNOTEŽENIM I NA TROŠKOVE OSJETLJIVIM PODACIMA

Važno pitanje je da li naučeni modeli posjeduju vrijeme, jer podaci nisu statični i oni se stalno mijenjaju u mnogim domenima. Istorijske akcije u uzorkovanju i izradi modela nisu optimalne, ali nisu bile ni nasumično izabrane. Ovo daje sledeće izazove fenomenu procesa za prikupljanje podataka. Pretpostavimo da koristimo podatke prikupljene u 2000 naučenih modela. Zatim ovaj model primjenimo da bismo izabrali od 2001-nog pojedinca. Nakon toga, mi koristimo podatke o pojedincima izabranih od 2001 da nauče novi model, a zatim primjenimo ovo za model u 2002. Ako se ovaj proces nastavi, onda će svaki put novi model morati da uči, njen trening je skup kreiran korišćenjem različitih sklonosti selekcija. Dakle, izazov ovog problema je kako da ispravimo različite sklonosti što je više moguće. Još jedan problem u vezi je kako da se nosi sa neuravnoteženim i na troškove osjetljivim podacima, to je veliki izazov u istraživanju.

U tipičnom realnom svijetu skupa podataka, postoji najmanje 100000 primjera i 100000000 funkcija, bez ijedne dobro definisane ciljne klase. Zanimljivi slučajevi imaju učestalost manju od 0,01%. Postoji mnogo informacija o troškovima i koristima, ali ne i ukupan model dobitka i gubitka. Postoje različite cijene za različite matrice primjene. Međutim, najpovoljniji unosi matrica su nepoznati. Primjer za ovaj skup podataka su podaci baza direktnog marketinga obrazovne fondacije. Osim toga, troškovi različitih ishoda su zavisili od primjera, na primjer, lažni negativan trošak direktnog marketinga je direktno proporcionalan iznosu potencijalne donacije. Tradicionalni metodi za dobijanje ovih troškova su se oslanjali na metode uzoraka. Međutim, metode uzoraka mogu lako da daju pristrasne rezultate.

XII. ZAKLJUČAK

Od svog početka u kasnim 1980-im, data mining je postigao ogroman uspjeh. Mnogi novi problemi su se pojavili u istraživanju pronalaženja podataka. Međutim, i dalje postoji nedostatak blagovremene razmjene važnih tema u zajednici kao cjelini.

Prethodni rad sistematizuje pregled najvažnijih problema u data mining tehnologiji. Moram napomenuti da je istraživanje sprovedeno na velikom broju radova iz ove oblasti, nije bilo moguće uzeti u obzir sve radove iz ove oblasti zbog ogromnog broja radova, tako da je moguće da ako se uspije uraditi analiza nad svim postojećim radovima onda bi moglo doći do promjene u izboru glavnih problema na koje nailaze istraživači iz ove oblasti. Za kriterijum odabira uzeta su u obzir i mišljenja eksperata iz ove oblasti.

Za sada bi trebalo uzeti u obzir navedene probleme kao glavne smjernice, na kojima treba da se radi da bi se otklonili problemi sa kojima se istraživači susrijeću pri pronalaženju korisnih podataka, informacija.

LITERATURA

[1] Data Mining by Bill Palace, Spring 1996.

[2] Data Mining - Theory, Methodology, Techniques,and Applications by Graham J. Williams Simeon J. Simoff, Springer-Verlag Berlin Heidelberg 2006

[3] Data Mining: Mining stream time series and sequence data by Datamining Tools Inc on Aug 18, 2010.

[4] Mining Complex Data by Zbigniew W. Ras, Shusaku Tsumoto and Djamel Zighed, 2007 Third International Workshop, Springer-Verlag Berlin Heidelberg 2008

[5] Data Mining :Concepts and Techniques, Second Edition by Jiawei Han, Micheline Kamber, 2006 by Elsevier Inc

ABSTRACT

This paper presents an overview of research conducted to rank the most important problems in data mining research. Here is a summarized ten of identified major problems that data mining research meets. In the selection criteria of what is considered valuable and important topics for future research in Data Mining, opinions were used the most active researchers in the areas of data mining and machine learning.

OVERVIEW OF MOST IMPORTANT PROBLEMS IN DATA MINING RESEARCH

Mladen Vukašinović