



Piše: Prof. dr Ivona Brajević

Primena metaheuristika u kontekstu Big Data

U današnje vreme stvara se i prikuplja mnogo više podataka nego što smo ih imali u bilo kom trenutku u prošlosti. Termin Big Data odnosi se na velike skupove podataka koji se ne mogu obraditi i analizirati na tradicionalan način, korišćenjem konvencionalnih procesa i alata. Ove skupove podataka karakterišu velika količina podataka, velike brzine obrade i pristupa i široka raznovrsnost formata. Svi ovi podaci potiču iz različitih izvora u zavisnosti od industrije i organizacije, kao što su objave na sajtovima društvenih mreža, digitalne slike i video zapisi, senzori koji se koriste za sakupljanje informacija, podaci koji se generišu za proizvode pri kupovini.

Razumevanje i korišćenje koncepta Big Data je od esencijalnog značaja za uspešno unapređenje poslovnih procesa, kao i samog opstanka poslovanja uz pomoć raspoloživih digitalnih alata i tehnologija. Izazovi koji se javljaju u ovom konceptu uključuju projektovanje i realizaciju infrastrukture i servisa za skladištenje ovih podataka, njihovu pretragu i analizu. Proces analize velike količine raznovrsnih podataka izvršava njihovu obradu u cilju dobijanja znanja. Savremene tehnologije pored toga što podržavaju njihovo skladištenje, pružaju mogućnost da se ti podaci razumeju i da njihova vrednost bude upotrebljena.

Složena analiza ovih podataka predstavlja rudarenje podataka, odnosno otkrivanje obrazaca i odnosa od interesa koji su skriveni u velikoj količini neobrađenih podataka. Tradicionalne metode rudarenja podataka nemaju adekvatnu primenu kod velikih količina podataka. Rudarenje ovih podataka zahteva visoko skalabilne metode, efikasnije korake predobrade, napredno okruženje za paralelno izračunavanje i efektivnu korisničku interakciju. Iz ovih razloga značajno se povećava potreba za kreiranjem visoko skalabilnih paralelnih algoritama mašinskog učenja i novih metoda rudarenja podataka.

Primena metaheuristika u kontekstu Big Data značajna je za analizu podataka pri rešavanju problema rudarenja podataka. Savremene heuristike ili metaheuristike predstavljaju algoritme koji su uglavnom inspirisani prirodnim procesima. Njihovom razvoju se poslednjih godina poklanja sve

više pažnje, dok su najistaknutije od njih, populacione metaheuristike, zasnovane na analogiji ponašanja individualne ili kolektivne inteligencije živog sveta pri potrazi za informacijama, izvorima hrane i načinu opstanka u celini. Mnogi optimizacioni problemi koje nije moguće rešiti klasičnim pristupom, rešavaju se uspešno algoritmičnim inspirisanim pojavama u prirodi. Kako je primena metaheuristika nezavisna od problema koji rešava, može se reći da ne postoji oblast ili problem iz određenog domena gde se ne bi mogao primeniti metaheuristički pristup.

Metaheuristike se smatraju fleksibilnijim optimizacionim tehnikama u odnosu na egzaktne algoritme. Metaheuristički okviri su definisani uopšteno, tako da metaheuristički algoritmi mogu da se prilagode potrebama rešavanja velikog broja različitih optimizacionih problema. Osnovna karakteristika metaheuristika je da prilikom rešavanja optimizacionih problema postižu zadovoljavajuće rezultate u razumnom vremenskom periodu, pri čemu ne zahtevaju ili mogu zahtevati samo nekoliko pretpostavki o datom problemu.

Specifično, za rešavanje teških optimizacionih problema velikih dimenzija, metaheuristike često postižu bolji odnos između kvaliteta rezultata i vremena utrošenog za njihovo izračunavanje u odnosu na egzaktne algoritme. Ovi algoritmi se mogu paralelizovati. U kontekstu gde se podaci redovno ažuriraju i gde odgovor mora da bude kreiran u realnom vremenu, metaheurističke metode veoma brzo mogu da daju

kvalitetno rešenje. Za klasične metode istovremeni rad sa različitim tipovima podataka je ograničen, dok metaheuristike predlažu različite načine kodiranja koji su u mogućnosti da razmatraju nekoliko tipova podataka istovremeno. Ovo daje mogućnost da se udruženo analiziraju podaci koji dolaze sa različitih izvora. Metaheuristike su metode optimizacije zasnovane na funkciji cilja, tako da omogućavaju procenu kvaliteta izdvojenog znanja.

Odnosi metaheuristika i koncepta Big Data su čvrsto povezani sa korakom analize podataka koji se sastoji od ekstrahovanja znanja iz dostupnih podataka. Zadaci koji se javljaju u rudarenju podataka uključuju klasterizaciju, klasifikaciju, izbor karakteristika i problem pravila udruživanja. Svi ovi problemi mogu da se formulišu kao problemi kombinatorne optimizacije. Njihovo rešavanje u kontekstu Big Data primenom egzaktnih metoda je ograničeno, tako da metaheuristike predstavljaju prihvaćen alternativni pristup.

Klasterizacija je zadatak grupisanja objekata tako da za svaki objekat važi da je sličniji objektima iz svoje grupe ili klastera, nego objektima iz drugih grupa. Sličnost objekata se određuje primenom neke od mera za računanje udaljenosti dva objekta. Mnoge klasične metode klasterizacije nije moguće primeniti za klasterizaciju Big Data jer zahtevaju potpuno izračunavanje udaljenosti između objekata. Iz ovog razloga metaheuristički algoritmi mogu predstavljati pogodnu tehniku za klasterizaciju podataka ekstremnog obima.



Nadgledana klasifikacija predstavlja još jedan važan zadatak rudarenja podataka koji kao ulaz dobija kolekciju objekata od kojih svaki pripada određenom broju klasa i koji je opisan svojim vrednostima za konačan skup atributa. Cilj je da se konstruiše klasifikator koji može da precizno predvidi klasu kojoj pripada novi objekat. Klasifikacija ima brojne primene, recimo, utvrđivanje da li je određena transakcija kreditnom karticom prevara, identifikacija kupaca koji mogu da budu zainteresovani za određeni proizvod ili predviđanje korisnih karakteristika aktera u društvenoj mreži. Prednosti primena metaheuristika za rešavanje ovog problema u kontekstu Big Data su mogućnost rada sa potencijalno velikim skupovima podataka i sposobnost adaptacije kroz upotrebu određene reprezentacije za raznovrsne podatke, što može da omogući istovremenu obradu različitih tipova atributa.

Izbor karakteristika je zadatak koji ima za cilj izdvajanje optimalnog relevantnog skupa karakteristika ili atributa koji su neophodni za klasifikaciju. Odgovarajući izbor karakteristika može da poboljša efikasnost modela zaključivanja. Rešavanje ovog problema upotrebom različitih metaheuristika zastupljeno je u literaturi. Primećeno je da rešavanje problema izbora karakteristika može da bude vremenski zahtevno kada je obim podataka

Metaheuristike se smatraju fleksibilnijim optimizacionim tehnikama u odnosu na egzaktne algoritme. Metaheuristički okviri su definisani uopšteno, tako da metaheuristički algoritmi mogu da se prilagode potrebama rešavanja velikog broja različitih optimizacionih problema.

veoma veliki. Iz ovog razloga pretragu je potrebno unaprediti inkorporacijom novih mehanizama u postojeće metaheuristike.

Pretraga za pravilima udruživanja je često korišćen pristup za otkrivanje veza od interesa među kolonama ili atributima velikih baza podataka. Inicijalna formulacija ovog problema ima za cilj da za dati skup proizvoda i veliku kolekciju evidencija prodaje koje obuhvataju datum transakcije i spisak kupljenih proizvoda u transakciji pronađe vezu između proizvoda koji su sadržani u različitim transakcijama. Od ovih prvih primena, izučavani su mnogi drugi problemi uključujući neprekidnu proizvodnju, detekciju upada i probleme iz bioinformatike. Veliki broj metoda za rešavanje ovog problema zasnovan je na metaheurističkim algoritimima. U kontekstu Big Data rešavanje problema pretrage za pravilima udruživanja je otežano, jer broj atributa i broj generisanih pravila može biti veoma veliki.

Sam pristup kontekstu Big Data jeste veliki izazov. Ovde se radi o izuzetno širokoj količini podataka iz različitih izvora i samim

tim izuzetno velikim mogućnostima za povećanje znanja i donošenje boljih odluka u različitim oblastima. Te mogućnosti su na raspolaganju naučnicima sa svih područja, kao i onima koji donose odluke i rukovode. Što se bolje iskoriste podaci kojima se raspolaže, toliko će se biti uspešan u tekućem vremenu.

Metaheuristike se odlikuju dobrom sposobnošću da prihvate izazov pribavljanja relevantnih podataka u kontekstu Big Data. Njihova fleksibilnost čini ih lako primenljivim za raznovrsne probleme s različitim tipovima podataka.

Primena metaheuristika u kontekstu Big Data utiče na razvoj metaheurističkog dizajna, jer će se i dalje generisati sve više podataka, sve brže i na sve različitim izvorima. Kreiranje još fleksibilnijih metaheurističkih pristupa za dobijanje boljih rezultata predstavlja ključni faktor uspeha. □

Prof. dr Ivona Brajević
Fakultet za primenjeni menadžment,
ekonomiju i finansije
www.mef.edu.rs