

## Primena fuzzy logike u metodama statističke analize

### Šta je fuzzy logika?

Kada je genijalni engleski matematičar George Boole sredinom XIX veka u svom delu »Zakoni mišljenja«, pokazao da zakoni formalne logike, koji su vekovima izučavani prema Aristotelu, mogu biti predmet matematičkog računa, započeo je razvoj nove oblasti matematike. Da bi ova nova oblast zaživela, bilo je potrebno sačekati još malo vremena. Tek kada je nemački matematičar George Cantor, krajem XIX veka stvorio teoriju skupova, otvorena je nova epoha u savremenoj matematici.

Prvih sto godina matematička logika se postepeno razvijala, stvarajući rezultate koji su doveli u pitanje mnoge osnovne postavke matematike. Međutim, jedan od osnovnih problema  $\{0,1\}$  vrednosne logike, zasnovane na Boolovoj algebri, je njena nemogućnost da reši realne probleme koji imaju vrednosti na domenu  $[0,1]$ . Naime, u realnom životu ništa nije crno ili belo, tačno ili pogrešno, a klasična matematička logika zna samo za ove dve vrednosti 0 ili 1. Da bi se ovaj problem rešio, prvo su uvedeni koncepti trovrednosne logike  $\{0,,1\}$ , ali ni oni nisu mogli da dosledno reprezentuju suštinu realnih odnosa.

Prelaskom na princip da se vrednosti uzimaju iz domena  $[0,1]$ , koji mi nazivamo fuzzy ili kontinualnom logikom, ovaj problem je naizgled rešen, ali se pojavio novi. Naime, iako je predloženo mnogo generalizacija Boolove logike na  $[0,1]$  domenu, bez obzira kako se osnovne logičke operacije izaberu, pojedine tautologije ne važe.

Gde smo mi u odnosu na svet?

Dobra vest za sve lokal-patriote: u ovoj oblasti, kao i u košarci, svet zaostaje za nama! Naime, nova sintaksno strukturirana semantički konveksna logika ili S3C logika, koja predstavlja rezultat rada naših ljudi, potpuno rešava gore navedene probleme. Ono od čega se polazi, jeste posmatranje logičkog izraza kao integralne celine dva osnovna dela:

- > osnovnih logičkih funkcija
- > strukture same logičke formule

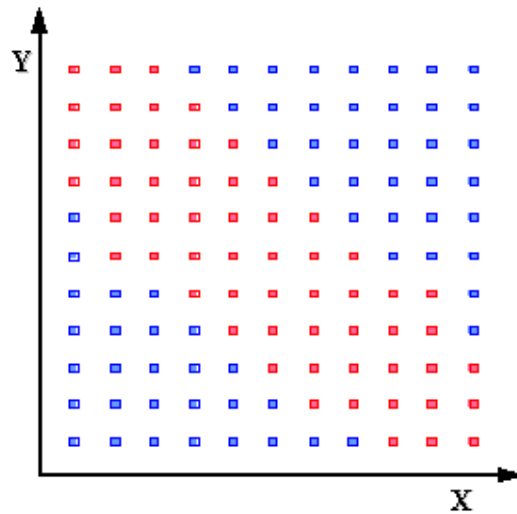
U skladu sa ovim, n-arna logička funkcija  $[0,1]$  domena može se predstaviti kao linearno konveksna kombinacija sopstvene strukture i osnovne logičke funkcije kao težinskog koeficijenta koji množi osnovne komponente strukturnog vektora. Ovo nam obezbeđuje da sama vrednost logičke formule zavisi i od njene strukture. Sve ovo nam, najjednostavnije rečeno, omogućava, da prvi put u istoriji tačno utvrdimo logičku zavisnost između varijabli koji opisuju određene objekte.

Pošto preciznije objašnjenje ove teorije značajno prevazilazi obim ovog izlaganja mi ćemo se ovde zaustaviti, a zainteresovanom čitaocu preporučujemo pionirski rad iz ove oblasti: Radojević, G. Dragan : [0,1] – valued logic : A natural generalisation of boolean logic, YUJOR 10(2000)

### Ima li logike u statistici?

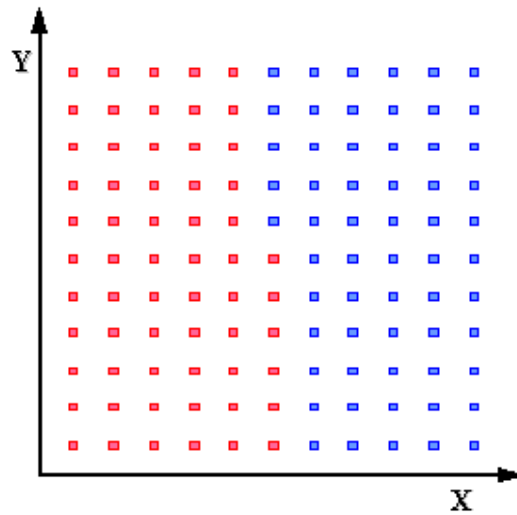
Različite metode multivarijacione analize, u svom radu nastoje da identifikuju, određene funkcionalne zavisnosti između varijabli koji opisuju posmatrane entitete. Pri tome se uglavnom zahteva funkcionalna zavisnost linearnog tipa, jer se sa nelinearnim funkcijama gotovo ne može raditi. Gde je problem? Problem je u tome što se linearnost teško ostvaruje i u laboratorijskim uslovima, a da ne govorimo o realnom životu. Zašto mi onda koristimo ove linearne modele? Odgovor je jednostavan : ne postoje bolji. Da budemo precizniji, do sada nije postojao bolji alat!

Da bi smo lakše demonstrirali primenu S3C logike u statističkoj analizi, pogledajmo sledeći primer.



Slika 1. – Izvorni skup podataka

Na grafiku (Slika 1) uočavamo dve klase opservacija: plave koje poseduju određeno svojstvo i crvene koje to svojstvo nemaju. Predpostavimo da nam je cilj da na posmatranom skupu podataka izvršimo diskriminacionu analizu, koja predstavlja jednu od najčešće korišćenih metoda multivarijacione analize. Pošto se u osnovi klasične diskriminacione analize nalazi linearna algebra, jedino što ova metoda može da uradi jeste da kroz n-dimenzioni prostor (u našem slučaju prostor je dvodimenzionalan), povuče odgovarajuću hiper-ravan (u našem slučaju to je prava).

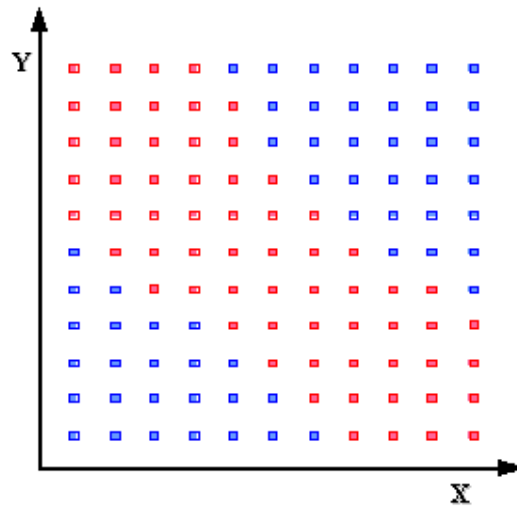


Slika 2. – Rezultati klasične diskriminacione analize

Na grafikonu (Slika 2.) vidi se kako je klasična diskriminaciona analiza podelila polazni skup podataka. Kao što smo očekivali diskriminaciona funkcija oblika:

(i predstavljaju određene realne konstante) koja se jasno uočava na posmatranom grafikonu, nije uspela da identifikuje klase polaznog skupa podataka. Ovako dobijeni rezultati su potpuno neupotrebljivi i da se radi o stvarnom istraživanju, u ovom trenutku bi ono verovatno bilo prekinuto.

Međutim, nepostojanje linearne funkcionalne zavisnosti između  $X$  i  $Y$ , u opštem slučaju ne znači da između njih ne postoji određena logička povezanost. Da bi smo ispitali mogućnost postojanja logičkih odnosa između promenljivih  $X$  i  $Y$ , potrebno je da izvorni skup varijabli proširimo vektorima njihovim bazičnim logičkim kombinacijama. Ove bazne kombinacije omogućiće nam da analizom obuhvatimo svih 16 mogućih logičkih kombinacija polaznog skupa podataka. Posebno je interesatno da čovek u procesu mentalnog zaključivanja, takođe koristi svih 16 logičkih kombinacija. Ukoliko sada na ovaj izvedeni skup podataka primenimo diskriminacionu analizu, dobićemo sledeći rezultat



Slika 3. – Rezultati diskriminacione analize fuzzy proširenog skupa podataka

Kao što se vidi na grafiku (Slika 3.) prezentovana metoda je gotovo u potpunosti identifikovala klase izvornog skupa podataka. Uočimo da smo u radu opet koristili klasičnu diskriminacionu analizu, ali ovog puta nad proširenim skupom podataka. Na taj način smo sačuvali sve prednosti koje nam u radu pruža linearna algebra, a pri tome smo uspjeli da identifikujemo nelinearnu zavisnost koja očigledno postoji između X i Y. U konkretnom slučaju logička zavisnost između izvornih varijabli je  $((X \text{ xor } Y))$ , gde xor predstavlja ekskluzivnu disjunkciju.

Na isti način, proširivanjem originalnog skupa podataka njihovim fuzzy logičkim kombinacijama, moguće je poboljšati sve ostale metode multivarijacione analize. Pogled u budućnost

Posljednjih godina svedoci smo široke primene različitih metode analize podataka u skoro svim naučnim oblastima. Dva su osnovna razloga za tako nešto. Prvi se odnosi na ubrzani razvoj kompjuterske tehnologije, koja uz snažnu logistiku sofisticiranih softvera, kao što su SPSS, MatLab... omogućava napredak naučnog rada u ovoj oblasti. Drugi razlog proističe iz sve veće zainteresovanosti različitih privredni subjekata i srodnih institucija, za rezultate do kojih se dolazi analizom podataka. Ovaj drugi razlog je naročito interesantan, jer u ovu oblast naučnog delovanja, unosi velike količine novca, koji omogućava komforan rad multidisciplinarnih istraživačkih timova, koje, pored statističara, čine i : ekonomisti, socijolozi, psiholozi, informatičari...

Kada je reč o primeni izložene metode, potrebno je naglasiti da rezultati koji se dobijaju nikada nisu lošiji od rezultata do kojih se dolazi primenom klasičnih metoda. Razlog za ovo leži u činjenici da je novi pristup prirodno uopštenje klasičnog pristupa. Međutim, novi pristup pored očiglednih prednosti krije u sebi i određene nedostatke. Ovi nedostaci, pre svega se odnose na povećanje dimenzija problema, usled proširivanja izvornog skupa podataka njihovim logičkim kombinacijama. Pored toga izložena teorija još uvek nije dobila svoje formalno matematičko priznanje, iako se na tom poslu užurbano radi. Zbog



svoga ovoga, nezahvalno je procenjivati, kada će ovaj pristup rešavanja problema zaživeti u praksi.

Osnovna ideja ovog izlaganja jeste da se kod čitaoca izazove određeni nivo interesovanja, jer prateći radove iz ove oblasti čitalac će imati priliku da prisustvuje rađanju jednog sasvim novog koncepta analize podataka.