Fejsbuk Nalazač Poklona

Facebook Gift Finder

Ivan Rakić

Elektrotehnički Fakultet u Beogradu

**Sadržaj –** Glavni zadatak web aplikacije Facebook Gift Finder je predlaganje poklona za prijatelje sa Facebok-a, na osnovu informacija prikupljenih sa njihovih profila. Ovaj rad opisuje realizaciju aplikacije koristeći CHAID stabla.

**Abstract –** Facebook gift finder application's main task is suggestion of gifts for Facebook friends, based on information gathered from their profiles. This paper describes realization and usage of the application using CHAID trees.

1. uvod

Odabir poklona predstavlja više od proste razmene fizičkih objekata i važan je oblik komunikacije između ljudi. Izborom poklona šaljemo poruku osobi kojoj je poklon namenjen. Neki pokloni potenciraju distancu, drugi intimnost, treći podsećaju na prijateljstvo i zajedničke oblasti interesovanja.

Izborom pravog poklona pokazujemo razumevanje i poštovanje osobe i njenih interesovanja i zato su kupovina i odabir poklona važni aspekti socijalne interakcije.

Postojeće web aplikacije za predlaganje poklona najčešće su realizovane na sledeća dva načina:

* korisnik mora da popuni niz pitanja vezanih za osobu kojoj je poklon namenjen, i aplikacija na osnovu odgovora predloži poklone.
* korisnik odabira kategoriju poklona iz liste kategorija i aplikacija mu predloži najpopularnije poklone prema podacima prikupljenim od ranijih korisnika.

Glavni nedostatak prvog pristupa je u tome što korisnik mora da popuni upitnik. Što više pitanja ima u upitniku, to će aplikacija dati bolje predloge, ali će korisniku oduzeti više vremena. Osim toga, korisnici se ne osećaju lagodno da daju informacije o osobi kojoj žele da poklone nešto, ili neke informacije naprosto ne znaju.

Nedostatak drugog postojećeg pristupa je u tome što donosi odluku samo na osnovu popularnosti poklona iz kategorije koju je korisnik odabrao, bez ikakvih informacija o osobi kojoj je poklon namenjen.

Fejsbuk profili sadrže većinu informacija koje su neophodne za odabir poklona: interesovanja, hobiji, mesta i događaji kojima je osoba prisustvovala, obrazovanje. Pri tome korisnik ne mora da popunjava upitnike niti daje bilo kakve dodatne informacije, a poklon će biti odabran prema ukusu i interesovanjima ciljane osobe.


Slika 1. Primeri informacija sa Facebook-a koje se mogu iskoristiti pri pronalaženju poklona

Centralni deo aplikacije predstavlja CHAID algoritam, a koristi se za klasifikaciju osobe kojoj je poklon namenjen na osnovu ulaznih podataka sa Facebook profila.

1. CHAID stabla

CHAID je skraćenica od *Chi*-squared Automatic Interaction Detector. CHAID stabla su stabla sa višestrukim granjanima, posebno popularna u marketinškim istraživanjima u kontekstu segmentacije tržišta.

Za svaki čvor u stablu, CHAID određuje za svaki potencijalni prediktor(ulazni podatak koji se koristi za određivanje podele) najbolju n-arnu podelu, i potom koristeći Bonferinijevu funkciju korekcije odabira prediktor sa najpogodnijom podelom.


Slika 2. Primer CHAID stabla

CHAID spada u nadgledane algoritme za Prikupljanje Skrivenog Znjanja(Data Minning), pa je za formiranje CHAID stabla neophodan skup poznatih izlaznih podataka za određene ulazne podatke. Ove podatke moguće je prikupiti sa postojećih aplikacija za prikupljanje i kupovinu poklona.

Faze CHAID algoritma su:

1. Priprema ulaznih podataka: ulazne podatke koji su kontinualni podeliti u kategorije.
2. Mešanje kategorija: Za svaku ulaznu promenljivu se utvrđuju parovi kategorija koji su najviše korelisani sa ciljnom promenljivom.
3. Bira se ulazna promenljiva koja će rezultovati najznačajniju podelu.

Za određivanje korelisanosti ulaznih promenljivih i izlazne promenljive koristi se Poasonova chi kvadratna funkcija:

$X^{2 }= \sum\_{j=1}^{J}\sum\_{i=1}^{I}\frac{(n\_{ij}-m\_{ij})2}{m\_{ij}}$ (1)

pri čemu je:

$n\_{ij}=\sum\_{}^{}f\_{n}I(x\_{n}=i, y\_{n}=j)$ (2)

frekvencija posmatrane ćelije tabele, a *mij* procenjena očekivana frekvencija pojavljivanja.

1. Proces izrade aplikacije

Izrada aplikacije se sastoji od sledećih etapa:

* Određivanje ciljanih izlaznih podataka i njihovih kategorija koji će se koristiti pri izboru poklona.
* Određivanje relevantnih ulaznih podataka i njihova kategorizacija.
* Formiranje CHAID stabla za svaki izlazni podatak.
* Formiranje baze podataka poklona i povezivanje poklona sa izlaznim podacima.
* Izrada korisničkog interfejsa

Ulazni podaci dobijeni sa Fejsbuk profila se koriste da bi se ciljani prijatelj kategorizovao za svaku izlaznu stavku(npr. koliko voli sport, koliko je posvećen hobijima, da li voli nakit), zatim se na osnovu izlazne stavke uparuju sa informacijama o poklonima u bazi i formira se lista predloga poklona, koja se prikazuje korisniku.

* 1. **Određivanje ciljanih izlaznih podataka**

Prikupljanjem podataka sa već postojećih aplikacija za predlog poklona kao i sprovođenjem upita nad odabranim grupama mogu se utvrditi kategorije korisnika u koje su relevantne pri izboru poklona i na osnovu njih odrediti izlazni podaci.

* 1. **Određivanje relevantnih ulaznih podataka**

Pošto se utvrde izlazni podaci, mogu se utvrditi koji ulazni podaci su bitni za njihovo određivanje, kao i način prikupljanja i provera istinitosti ulaznih podataka. Za svaki ulazni podatak mora se izvršiti kategorizacija.

* 1. **Formiranje CHAID stabla za svaki izlazni podatak**

Po određivanju ulaznih i ciljnih izlaznih podataka, formira se CHAID stablo podelom čvorova po ulaznim podacima do unapred utvrđene tačnosti. Tačnost se utvrđuje postavljanjem vrednosti koju Poasoniva chi kvadratna funkcija ne sme da pređe.

Ukoliko izlazi CHAID algoritma ne daju dovoljno dobre rezultate, potrebno je izvršiti ponovnu pripremu ulaznih podataka koji daju najslabije podele u čvorova u CHAID algoritmu.

* 1. **Formiranje baze podataka**

Potrebno je projektovati bazu podataka koja pored osnovnih informacija o poklonima, kao što su naziv, cena, lokacija, opis, sadrži i listu tagova za svaki poklon, pri čemu su tagovi kategorije izlaznih podataka CHAID algoritma.

Npr. poklon sa nazivom sportski ranac ima tagove :

* Bavljenje sportom: Srednja i velika zainteresovanost.
* Provođenje vremena u prirodi: Srednja zainteresovanost.
	1. **Izrada korisničkog interfejsa.**

Ovakva aplikacija ne zahteva suviše složen korisnički interfejs i najpraktičnije rešenje je integracija ove aplikacije sa Facebook-om.

Ulazni podaci koje korisnik unosi se svode na izbor Facebook prijatelja iz liste prijatelja korisnika kome se bira poklon.

Najveći deo korisničkog interfejsa čini prikaz izlaznih podataka, koji predstavljaju listu poklona za odabranog prijatelja. Pogodan način prikaza poklona mora biti takav da korisnik brzo može da uoči izgled poklona i lokaciju na kojoj može taj poklon da kupi.

1. Opis rada aplikacije

Rad aplikacije se sastoji od sledećih celina:

* Preprocesiranje i utvrđivanje pouzdanosti ulaznih podataka.
* Izvršavanje CHAID algoritma i dobijanje izlaznih podataka.
* Pronalaženje poklona u bazi koji najviše odgovaraju dobijenim izlaznim podacima i prikaz rezultata.
	1. **Preprocesiranje i utvrđivanje pouzdanosti ulaznih podataka**

Informacije na Fejsbuku nisu potpuno pouzdane pa dodatni nivo apstrakcije između ulaznih podataka i izbora poklona koji predstavljaju rezultati CHAID algoritma delimično pruža otpornost od uticaja netačnih informacija na korisnikovom profilu.

Za određene bitnije informacije se, ipak mora dodatno proveriti njihova tačnost. Tako, na primer, od trenutnog mesta stanovanja korisnika zavisi koje će mu lokacije za kupovinu poklona aplikacija predlagati. Kako korisnik može uneti koji god grad ili mesto želi za trenutno mesto stanovanja, ovu informaciju treba potvrditi recimo proverom mesta i događaja na kojima je korisnik bio.

* 1. **Izvršavanje CHAID algoritma i dobijanje izlaznih podataka.**

Na osnovu ulaznih podataka propuštenih kroz CHAID algoritam dobijaju se kategorije izlaznih podataka.

CHAID algoritam se sprovodi i za korisnika koji kupuje poklon i za korisnika kome je poklon namenjen. Izlazni podaci korisnika koji kupuje poklon se koriste da bi se utvrdile oblasti zajedničkog interesovanja sa korisnikom kome je poklon namenjen i poklonima iz tih oblasti dala prednost u odnosu na ostale.

Potrebno je uzeti u obzir i sociološke i ekonomske aspekte korisnika koji kupuje poklon.

* 1. **Pronalaženje poklona u bazi**

Na osnovu izlaznih podataka pronalaze se pokloni u bazi
i formira se lista potencijalnih poklona.

Zatim se formiranu lista filtrira na osnovu utvrđenih zajedničkih oblasti interesovanja i na osnovu odnosa između korisnika koji kupuje poklon i korisnika kome je poklon namenjen.

1. Zaključak

Aplikacija za pronalaženje poklona može bez većih problema da se integriše sa Fejsbukom. Kako Fejsbuk već vodi evidenciju o datumima rođenja Fejsbuk korisnika i

podseća korisnike na rođendane njihovih prijatelja, opcija za predlog poklona prijateljima bi mogla dobro da se uklopi u takvom kontekstu.

Glavni problem prilikom izrade ove aplikacije bi bio formiranje baze podataka za poklone.

Jedno od rešenja je da Fejsbuk pravi svoju sopstvenu bazu

podataka poklona. Ovo rešenje ima nekoliko nedostataka , od kojih su glavni glomaznost takve baze i pitanje načina ažuriranja baze.

Drugo rešenje je da se Fejsbuk poveže sa sajtovima prodavnica i šoping centara i podatke dovlači od njih. Ovo rešenje takođe ima nedostatak: svaki sajt od koga Fejsbuk vuče podatke morao bi da u bazu poklona doda informacije vezane za tagove poklona koji su neizostavni deo ove aplikacije.

Najbolje rešenje je da prodavnice i brendovi koji žele da budu vidljivi iz aplikacije, naprave Fejsbuk stranice na kojima će sami dodavati i ažurirati artikle. Pravljenjem ovakvih stranica prodavnice bi dobijale dodatnu reklamu i kupce zbog aplikacije, pa bi im pravljenje stranice na Fejsbuku bilo u interesu.

Ukoliko aplikacija za pronalaženje poklona ne bi bila integrisana sa Fejsbukom, već samo koristila podatke sa Fejsbuka za pronalaženje poklona, sadržala bi i sopstvenu bazu poklona.

literatura

[1]Ritschard, G., CHAID and Earlier Supervised Tree Methods, Dept of Econometrics, University of Geneva, Switzerland, 2010.

[2]Bigss, D., Ville, B., and Suen, E. (1991). A Method of Choosing Multiway Partitions for

Classification and Decision Trees. Journal of Applied Statistics, 18, 1, 49-62.

[3]Ramaswami, M. and Bhaskaran, R., A CHAID Based Performance Prediction Model in Educational Data Mining, Department of Computer Science(DDE), Madurai Kamaraj University Madurai, Tamilnadu 625 021, India, 2010.