

UTILIZAREA FUZZY LOGIC ÎN ANALIZA AUTOMATĂ A DATELOR DIN GOOGLE ANALYTICS

Sergiu, CHILAT

Universitatea Tehnică a Moldovei

Abstract: *Analiza datelor despre utilizatorii aplicațiilor web a devenit o sarcină de bază a comerțului electronic, aceasta fiind necesară pentru controlul și optimizarea sitului, cu scopul de a obține un profit maxim. Două dintre cele mai populare motoare de căutare Google și Yandex, oferă utilizatorilor săi instrumente specializate: Google Analytics și Yandex.Metrica. Acestea permit acumularea datelor despre utilizatorii sitului, cum ar fi: amplasarea geografică, preferințe și interese, comportamentul pe site etc. Aceste date pot fi utilizate pentru optimizarea sitului, însă nu există un modul care ar putea automatiza procesul de luare a deciziilor, aceasta fiind sarcina administratorului SEO (Search Engine Optimization). În acest articol vor fi cercetate posibilitățile logicii Fuzzy de a automatiza procesul de luare a deciziilor și substituirea administratorului în unele situații.*

Cuvinte cheie: *fuzzy logic, mulțimi fuzzy, web analytics, Google Analytics, SEO, automatizare, reglare automată.*

INTRODUCERE

Creșterea exponențială a numărului utilizatorilor Internetului a condus inevitabil la reorientarea pieței, comerțul electronic devenind o componentă importantă în tot mai multe domenii. Datorită avantajelor pe care le oferă, acesta atrage tot mai mulți utilizatori, numărul siturilor de tip E-commerce fiind într-o creștere permanentă și stabilă. Ca urmare, concurența este de asemenea într-o continuă creștere, iar motoarele de căutare cum ar fi Google sau Yandex, dau o mai mare prioritate acelor situri care oferă un conținut util și relevant pentru utilizatori.

Pentru a prezenta utilizatorilor un conținut util și relevant, este nevoie de a cunoaște care sunt interesele și preferințele lor. În acest scop sunt utilizate instrumente specializate, care permit colectarea datelor despre utilizatori. Două instrumente de acest tip sunt Google Analytics și Yandex.Metrica, oferite de Google și Yandex, care sunt unele dintre cele mai populare motoare de căutare. Fiecare dintre aceste instrumente asigură posibilitatea acumulării datelor despre utilizatorii sitului: numărul de vizitatori, numărul vizitatorilor unici, numărul de pagini vizualizate de utilizator etc. În plus, toate aceste date sunt grupate în dependență de locația utilizatorului, sursele traficului, cuvintele cheie utilizate la căutarea informației, tipul browserului, sistemul de operare, etc. Analiza acestor date și luarea deciziilor, în dependență de anumiți factori, este făcută de administratorul SEO, deoarece nici Google Analytics nici Yandex.Metrica nu oferă posibilitatea de a genera în mod automat prognoze sau indicații. Aceste decizii nu sunt de fiecare dată cele optimale, fiind prezent factorul uman, iar un rol important îl are și experiența. Procesul de luare a deciziilor este unul foarte anevoios, iar uneori acest lucru este practic imposibil, ajungând-se la compromisuri și decizii nu tocmai favorabile. De multe ori, administratorul SEO execută unele și aceleași operații de rutină în mod repetat, ceea ce duce la descreșterea randamentului.

Logica fuzzy vine ca o rezolvare a acestei probleme. Aici se operează cu termenul de "variabilă lingvistică". Pentru o variabilă lingvistică, se specifică numărul necesar de termeni, și fiecareia dintre ei i se atribuie o valoare. Pentru unul din termeni se va seta valoarea egală cu unitatea, iar pentru toate celelalte, valori din intervalul $[0..1]$, astfel, se poate ajunge la excluderea restricțiilor logicii lui Aristotel.

Pentru a exclude factorul uman și a economisi timpul administratorului SEO, ar fi binevenit un modul, care ar automatiza multe dintre procesele de rutină, însă realizarea unui astfel de modul, utilizând metodele clasice este foarte complicată și în unele cazuri imposibilă. Este nevoie de ceva deosebit, ceva care ar permite analiza datelor și luarea deciziilor în mod automat și ar fi dotat cu "gândire umană". În acest scop ar putea fi utilizată logica Fuzzy, care permite modelarea sistemelor cu logică neclară, ea fiind foarte apropiată de logica umană.

În acest articol se va face o analiză a instrumentului Google Analytics și vor fi analizate posibilitățile logicii Fuzzy în procesul de luare a deciziilor în locul administratorului SEO.

1. ANALIZA DATELOR OFERITE DE GOOGLE ANALYTICS

Google Analytics (prescurtat GA) este o aplicație, oferită de compania Google, care este utilizată pentru colectarea informației despre utilizatorii sitului și gruparea ei în categorii, pentru o analiză ulterioară de către administratorului SEO.

Google Analytics pune la dispoziție multe instrumente și o varietate mare de date, grupate după sens și conținut, însă nu există posibilitatea prelucrării automate a acestor date. În prezentul articol vor fi analizate posibilitățile Google Analytics, informațiile despre utilizatori pe care le pune la dispoziție și posibilitățile logicii Fuzzy în procesarea automată a acestor informații și luarea deciziilor în locul administratorului SEO.

Google Analytics pune la dispoziție mai multe instrumente utilizate la colectarea informațiilor despre utilizatori. Vor fi analizate următoarele compartimente ale Google Analytics, pentru care ar putea fi utilizată logica Fuzzy:

- Vârsta utilizatorilor
- Pagini populare

1.1. Vârsta utilizatorilor

În acest compartiment, GA grupează datele obținute în 6 grupe (fig. 1), pe care le vom nota cu GV_1 - GV_6 :

- GV_1 - de la 18 până la 24 ani;
- GV_2 - de la 25 până la 34 ani;
- GV_3 - de la 35 până la 44 ani;
- GV_4 - de la 45 până la 54 ani;
- GV_5 - de la 55 la 64 ani;
- GV_6 - peste 65 ani.

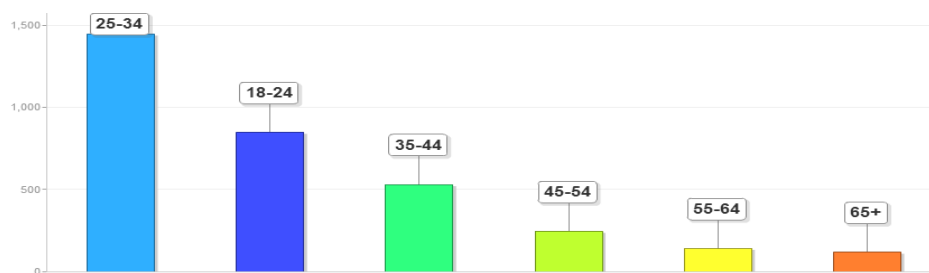


Fig. 1. Gruparea datelor după vârsta utilizatorilor în Google Analytics

Acest compartiment este util pentru a determina interesele utilizatorilor în dependență de vârsta pe care o au. Astfel, conținutul pe site ar putea fi împărțit în 3 grupe:

- GV_T - pentru persoane tinere;
- GV_M - pentru persoane de vârstă medie;
- GV_B - pentru persoane bătrâne.

Deoarece nu există o limită de trecere clară între aceste categorii de vârstă, gruparea în 6 grupe (GV_1 - GV_6) pe care o pune la dispoziție Google Analytics, nu poate fi utilizată direct, deci este nevoie de aplicarea unui algoritm care ar determina din care din cele 3 grupe pe care le-am definit face parte utilizatorul. Aici ar putea fi aplicată logica Fuzzy.

Vom cerceta un exemplu concret, un utilizator cu vârsta de 35 ani. Conform grupării făcute de GA, această persoană se află la limita inferioară a grupei GV_3 . Apare o dilemă, în care dintre cele 3 grupe definite, GV_T , GV_M sau GV_B se află persoana respectivă?

Am putea considera că grupele GV_1 și GV_2 determină grupa persoanelor tinere – GV_T , grupele GV_3 și GV_4 determină grupa persoanelor de vârstă medie – GV_M , iar GV_5 și GV_6 determină grupa persoanelor bătrâne – GV_B .

Conform separării stricte pe care o face Google Analytics, persoana cu vârsta de 35 ani aparține grupei persoanelor de vârstă medie - GV_M .

Vom considera o a doua persoană cu vârsta de 34 ani. Conform grupării făcute de Google Analytics, această persoană aparține grupei GV_2 , și se află la limita superioară a acestei grupe.

Conform acestei grupări, persoana cu vârsta de 34 ani este tânără, iar cea de 35 de ani este de vârstă medie, deci au interese diferite, ceea ce nu este prea corect, deoarece un conținut pentru o persoană de 34 ani, poate fi la fel de util ca și pentru o persoană de 35 ani.

Aici am putea aplica logica Fuzzy, pentru determinarea gupei din care face parte o persoană și prezentarea unui anumit conținut în dependență de grupa din care face parte. În acest scop, vom utiliza mulțimi fuzzy pentru repartizarea persoanelor în 3 grupe: tineri, de vârstă medie și bătrâni(fig. 2).

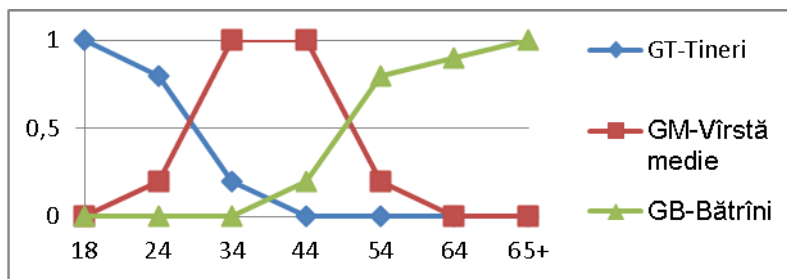


Fig. 2. Reprezentarea Fuzzy a distribuției vârstei

După cum se poate observa din fig. 1, o persoană este "foarte" tânără la vârsta de 18 ani(valoarea maximală 1), "mai puțin" tânără la vârsta de 24 ani(valoarea 0,8), "puțin" tânără(valoarea 0,2) la vârsta de 34 ani, iar la vârsta de 44 ani nu mai este tânără(valoarea 0).

La fel și pentru persoanele bătrâne, o persoană nu este bătrână pînă la vârsta de 24 ani(valoarea minimală 0), "puțin" bătrână la vârsta de 34 ani(valoarea 0,2), "aproape" bătrână(valoarea 0,8) la vârsta de 54 ani, iar la vârsta de 65 ani este bătrână(valoarea maximală 1).

Pentru persoanele de vârstă medie se observă o mică creștere(0,2) a funcției de apartenență de la 18 la 24 ani, fiind observată o creștere semnificativă în intervalul 24-34 ani(pînă la 0,8), iar în intervalul 34-44 este atinsă valoarea maximală(1). După 45 ani, valoarea funcției de apartenență descrește, ajungînd la 0 pentru vârsta de 64 ani.

Astfel, aplicînd mulțimile fuzzy definite și un algoritm bazat pe logică fuzzy, vom obține următoarele rezultate:

Dacă un conținut de pe site este destinat persoanelor de vârstă medie, acesta va fi foarte util(va avea utilitatea maximă 1) pentru utilizatorii cu vârsta din intervalul 34-44, pentru utilizatorii cu vârste cuprinse între 24-34 și 44-54 va avea o utilitate medie(între 0,2 și 0,8) iar pentru utilizatorii cu vârsta pînă la 24 ani și după 54 ani, acest conținut va avea o utilitate mică, deci nu va prezenta interes.

Utilizînd datele prezentate în fig. 1, și mulțimile fuzzy definite, se poate observa că numărul maxim de utilizatori au vârsta cuprinsă între 25-34 ani și apoi între 18-24 ani, deci, algoritmul va genera în mod automat două rezultate "Conținutul este util pentru persoanele de vîrsă tânără și medie" și "Conținutul nu este util pentru persoanele în vîrstă".

Utilizarea acestui algoritm de automatizare este utilă în cazul siturilor cu un număr mare de pagini, astfel economisindu-se foarte mult timp.

1.2Pagini populare

În acest compartiment, Google Analytics, prezintă informații despre numărul de vizitatori pentru fiecare pagină(fig. 3).

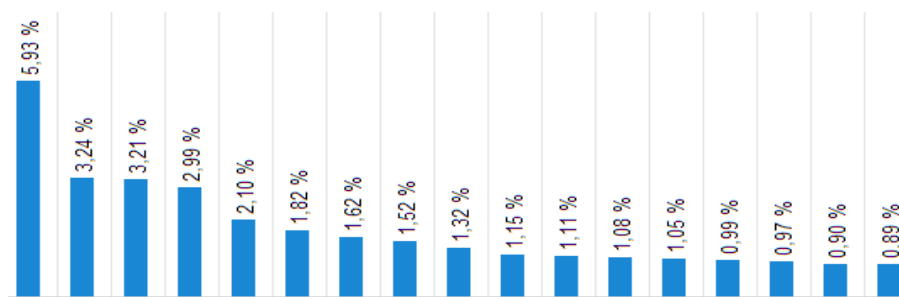


Fig. 3. Reprezentarea grafică a numărului de utilizator pentru fiecare pagină

De multe ori este necesar de a determina care sunt cele mai mult vizitate pagini și paginile cu cele mai puține vizualizări. Acest lucru este făcut de administratorul SEO, iar atunci când numărul paginilor este mare, acest proces devine unul foarte costisitor. O automatizare a analizei acestor date, ar rezolva o serie de probleme, cu atât mai mult, ca numărul de vizualizări poate fi grupat după zile, săptămâni sau luni, în dependență de scopul utilizării acestor date. Astfel volumul de lucru pentru administrator crește foarte mult.

Pentru a putea aplica logica fuzzy, datele vor fi grupate în 4 mulțimi fuzzy (fig. 4), în dependență de numărul de utilizatori:

- G_H – cu un număr mare de vizualizări (peste 5% din total);
- G_M – cu număr mediu de vizualizări (între 1 și 5%);
- G_L – cu număr mic de vizualizări (pînă la 1%);

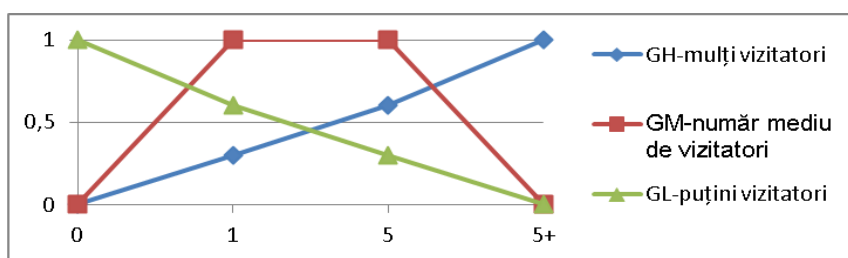


Fig. 4. Reprezentarea Fuzzy a distribuției vizualizărilor

Atunci când se face analiza unei pagini concrete, administratorul poate decide că o pagină care are un număr de vizualizări de 4.8% să fie atribuită categoriei G_H , cu toate că limita minimă este de 5%. Algoritmul bazat pe logică fuzzy ar trebui să genereze în mod automat aceleași decizii. Aici, modificare limitei depinde de anumiți parametri cum ar fi: numărul total de pagini, numărul de pagini care trec de limita de 5%, raportul dintre numărul de pagini din fiecare grupă.

Concluzii

Logica fuzzy operează cu noțiunea de ”variabilă lingvistică”, ceea ce îi permite să opereze cu date inexacte și generarea de soluții în baza informațiilor incomplete.

Analizând posibilitățile Google Analytics, se poate face următoarea concluzie:

Pentru excluderea factorului uman și economisirea timpului, analiza datelor poate fi făcută de o aplicație specializată, care ar putea genera soluții în mod automat, iar aceste soluții să fie prezentate în limbaj uman, deci să nu mai fie nevoie de o prelucrare adăugătoare pentru a fi înțelese de om. Crearea acestei aplicații, utilizând logica clasică nu ar fi posibilă, iar la rezolvarea problemei poate fi aplicată cu succes logica fuzzy.

În baza exemplului cercetat, se poate observa că datele, care inițial erau împărțite în mulțimi clasice, au fost regrupate în mulțimi fuzzy, după care a putut fi aplicată logica fuzzy pentru prelucrarea datelor și generarea soluțiilor. Crearea mulțimilor fuzzy este un prim pas în implementarea algoritmului bazat pe logică fuzzy, acesta fiind una din cele mai importante etape, unde este nevoie de o generare corectă a acestor mulțimi.

Logica fuzzy și algoritmi cu logică fuzzy pot fi utilizați în rezolvarea problemelor nedeterminate, acolo unde este nevoie de o ”gîndire” asociativă și variabilă, asemenea gîndirii umane. Deoarece analiza datelor despre utilizatorii unui site este un proces continuu și neliniar, prelucrarea în timp real a datelor în mod automat oferă o serie de avantaje, cum ar fi: economie de timp, creșterea vitezei de procesare, excluderea greșelilor pe care le poate face programatorul SEO.

Bibliografie

1. Zadeh, L.A., Fuzzy sets, Information and Control, vol. 8, 1965, pp. 338–353.
2. Smith. C. A., Automated continuous process control. John Willey and Sons, New York, 2002.
3. Preitl, Ș., Precup, R.-E., Introducere în conducerea fuzzy a proceselor, Ed. Tehnică, București, 1997.
4. Google Analytics Feature List. Sursă electronică (citată 01.11.2013). http://www.google.ru/intl/en_uk/analytics/features/index.html
5. Возможности счетчика Google Analytics. Sursă electronică (citată 02.11.2013). <http://www.sembook.ru/book/instrumentariy-dlya-prodvizheniya/vozmozhnosti-schetchika-google-analytics/>