

Utilizarea algoritmilor NeuroFuzzy la generarea paginilor web personalizate

Sergiu CHILAT

Technical University of Moldova

chilatsergiu@gmail.com

Abstract — în acest articol este prezentată o metodă de generare a conținutului personalizat pe web-site. Implementarea metodei va permite generarea în mod automat a conținutului personalizat, ceea ce va duce la mărirea conversiei și a vânzărilor produselor expuse pe site. Automatizarea va fi făcută în baza unui algoritm neurofuzzy și a unei rețele neuronale, care, având la intrare un set de date despre utilizator la ieșire va determina tipul acestui utilizator, și va permite crearea automată a bazei de reguli în baza cărora va fi prezentat conținutul.

Index Terms — logică fuzzy, rețele neuronale, neurofuzzy, conținut personalizat .

I. INTRODUCERE

Problemele de luare a deciziilor în situații complexe ocupă la moment un loc special în domeniul tehnologiilor informaționale. Modelele matematice sunt utilizate pe larg la descrierea și analiza sistemelor complexe din așa domenii ca: economic, social, industrial etc. Teoria optimizărilor a creat un set de metode, care, cu utilizarea calculatoarelor, facilitează luarea de decizii atunci când parametrii sunt cunoscuți și fixați, sau când parametrii primesc valori aleatorii, dar se supun unor legi sau funcții care generează aceste valori. Există, totuși, anumite probleme care nu pot fi supuse descrierii formale, deoarece parametrii au valori inexacte sau incomplete, iar apartenența la o clasă nu este discretă ci continuă.

Un caz particular al acestui tip de probleme este și prezentarea eficientă a conținutului pe un web-site. Aici, metodele tradiționale de expunere a informației nu dau un randament maxim. Din această cauză, multe web-situri utilizează mecanisme de prezentare a conținutului personalizat. Spre exemplu, gigantul Amazon, afișează pentru fiecare utilizator al său o pagină principală diferită, iar conform unor studii efectuate de Adobe și Econsultancy, 52% dintre oamenii care vînd produse pe internet, consideră personalizarea conținutului drept o strategie de bază [1]. În acest scop, în baza informațiilor obținute de la vizitatorul sitului, pot fi utilizate diferite strategii: geo-targeting, grupurile de vîrste, cuvintele cheie utilizate la căutare, etc.

În acest articol se va propune o metodă nouă, bazată pe logica fuzzy și rețele neuronale. Această metodă va consta în:

- păstrarea în baza de date a sitului a istoriei vizitelor și acțiunilor vizitatorului;
- analiza istoriei, utilizînd algoritmi speciali creați;
- generarea bazei de reguli;
- afișarea unui conținut personalizat, conform bazei de reguli generate;
- analiza parametrilor de conversie și adaptarea automată a sistemului.

II. PROBLEMA GENERĂRII PAGINILOR CU CONȚINUT PERSONALIZAT

De cele mai multe ori, la prezentarea conținutului pe un web-site sunt utilizate anumite criterii bine determinate, cum ar fi: prezentarea informației sortat după dată, dintr-o anumită categorie, în dependență de filtrele setate de utilizator, etc.

Astfel, toți utilizatorii văd unul și același conținut, dar, ținînd cont de faptul că fiecare are interesele proprii, așa o prezentare este ineficientă și are multe dezavantaje, printre cele mai principale fiind:

- prezentarea informației cu o utilitate mică pentru utilizator, înaintea celei cu o utilitate mai mare;
- prezentarea aceluiași conținut pentru toți utilizatorii, astfel, micșorîndu-se nivelul de conversie;
- pentru a găsi informația necesară, utilizatorul trebuie să utilizeze elementele de navigare și de căutare puse la dispoziție de site;
- utilitatea elementelor de navigare depinde de experiența utilizatorului și de programatorul care a proiectat interfața.

În ultimul timp, tot mai multe web-situri abordează această problemă foarte serios, deoarece unul din scopurile principale ale unui site, este acela de a prezenta un conținut util pentru un număr cît mai mare de vizitatori. Aici, se pot defini mai multe metode și strategii, fiecare din ele avînd o anumită eficacitate.

Una din cele mai populare și eficiente metode este, prezentarea conținutului personalizat. Această metodă are o importanță mare, îndeosebi în cazul siturilor comerciale (internet magazine, portaluri e-commerce, etc.).

Nu în zadar, unii din cei mai mari agenți e-commerce și motoare de căutare, acordă o mare prioritate prezentării conținutului personalizat. Printre aceștia se pot menționa:

- www.amazon.com – conform studiilor efectuate [1], ca urmare a aplicării metodei respective, s-a obținut o creștere a numărului de click-uri cu 149%, iar numărul de vizitatori a crescut cu 0.43%;
- www.yandex.ru – ca urmare a lansării platformei “Kaliningrad”, care realizează căutări personalizate,

calitatea rezultatelor căutării a crescut cu 75-80% [2] astfel vizitatorul economisind 14% din timp la căutarea informației necesare [2].

III. PREZENTAREA METODEI DE REZOLVARE A PROBLEMEI

Metoda propusă va putea fi utilizată în web-siturile de tip e-commerce, pentru care ar putea prezenta un mare interes, deoarece poate duce la creșterea vânzărilor cu până la 30-35%.

În acest articol se va propune o metodă nouă, bazată pe logica fuzzy și algoritmi neurofuzzy. Această metodă va consta în:

- păstrarea în baza de date a informației despre vizitatorul sitului, a istoriei vizitelor și acțiunilor sale;
- analiza istoriei vizitelor și acțiunilor, utilizând algoritmi speciali creați;
- crearea rețelei neuronale;
- instruirea rețelei;
- elaborarea bazei de reguli (variabile lingvistice);
- generarea unui conținut personalizat, în conformitate cu rezultatele generate de rețeaua neuronală și regulile elaborate;
- analiza parametrilor de conversie și adaptarea automată a sistemului.

Păstrarea în baza de date a sitului a istoriei vizitelor și acțiunilor vizitatorului

Sistemul elaborat în baza metodei propuse, va identifica vizitatorul și îi va atribui un identificator unic, care va fi apoi utilizat pentru ducerea evidenței acțiunilor efectuate la vizitele ulterioare.

Despre fiecare utilizator va fi păstrată următoarea informație:

- numele și datele de contact introduse;
- vârsta și sexul;
- poziționarea geografică;
- platforma (sistemul de operare, browserul, etc.);
- adresa URL de pe care a venit pe vizitatorul pe site;
- cuvintele cheie utilizate în motorul de căutare.

De asemenea, vor fi fixate următoarele acțiuni ale vizitatorului:

- deschiderea paginii;
- timpul petrecut pe fiecare pagină deschisă;
- lista de pagini vizitate în cadrul unei sesiuni;
- adăugarea sau ștergerea unui produs în/din coșul virtual;
- finalizarea comenzii.

Metoda propusă constă din 2 compartimente:

1. analiza comportamentului și determinarea tipului cumpărătorului;
2. analiza informației despre utilizator și determinarea produselor de care are nevoie.

IV. ANALIZA COMPORTAMENTULUI ȘI DETERMINAREA TIPULUI CUMPĂRĂTORULUI

La această etapă, sistemul va acumula datele despre comenzile efectuate și produsele cumpărate de utilizator,

iar în baza unei rețele neuronale, va determina tipul utilizatorului, pentru a determina ce fel de produse să i se propună (scumpe, ieftine etc.).

Analiza istoriei, utilizând algoritmi speciali creați

Sistemul va utiliza datele colectate despre utilizator și va face o analiză a acestora. Vor fi analizați următorii parametri:

- numărul total de pagini vizitate;
- numărul de pagini vizitate la o singură vizită (media);
- timpul total petrecut pe site;
- timpul petrecut pe site la o singură vizită;
- numărul de produse comandate;
- numărul de produse cumpărate;
- suma totală a produselor comandate;
- suma totală a produselor cumpărate;
- raportul dintre produsele comandate și cele cumpărate.

Crearea rețelei neuronale

Problema: determinarea tipului clientului în dependență de numărul produselor vizualizate, comandate și procurate.

La etapa inițială în sistem vor fi introduse: setul de formare și datele de testare a rețelei.

Ulterior, se va iniția procesul de învățare a rețelei prin metoda backpropagation (fig 1), care va continua atâta timp cât eroarea generată pentru datele de intrare va scădea. Oprirea procesului de învățare a rețelei se va opri, la momentul în care eroarea obținută pentru datele de testare va începe să crească.

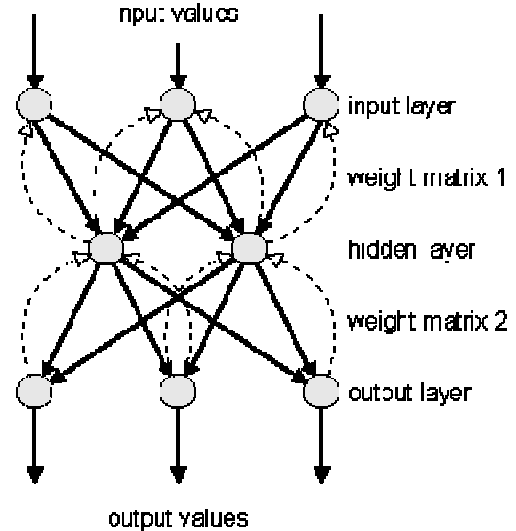


Fig. 1. Învățarea rețelei prin metoda backpropagation.

La intrarea rețelei neuronale se vor transmite următorii parametri:

- numărul total al produselor vizualizate;
- numărul produselor comandate;
- numărul produselor cumpărate;
- suma produselor comandate;
- suma produselor cumpărate;

Rețeaua va avea 10 ieșiri:

- vizualizează multe produse și le cumpără pe cele

- scumpe;
- vizualizează multe produse și le cumpără pe cele ieftine;
- vizualizează puține produse și le cumpără pe cele scumpe;
- vizualizează puține produse și le cumpără pe cele ieftine;
- vizualizează multe produse și nu le comandă;
- vizualizează puține produse și nu le comandă;
- vizualizează multe produse și a cumpărat multe;
- vizualizează multe produse și a cumpărat puține;
- cumpără toate produsele comandate;
- doar comandă, dar nu cumpără nimic.

Elaborarea bazei de reguli

Baza de reguli va avea următorul format:

DACĂ condiție ATUNCI acțiune

astfel, baza de reguli[3, 4] va avea următorul conținut:

DACĂ

vizualizează multe și le cumpără pe cele scumpe
ATUNCI

prezentăm mai întâi produsele mai scumpe;

DACĂ

vizualizează multe și le cumpără pe cele ieftine
ATUNCI

prezentăm mai întâi produsele mai ieftine;

DACĂ

cumpără toate produsele comandate
ATUNCI

oferim o reducere de preț;

DACĂ

doar comandă dar nu cumpără
ATUNCI

nu are posibilitatea de a comanda;

Baza de reguli prezentată este doar un model, dar ar putea fi completată cu reguli noi, pentru a mări exactitatea soluțiilor generate și a adapta sistemul la diferite situații.

Afișarea unui conținut personalizat, în conformitate cu rezultatele generate de rețeaua neuronală și regulile elaborate

În baza datelor obținute la ieșirea rețelei și a strategiilor elaborate(baza de reguli), se va lua decizia ce tip de conținut este mai bine de prezentat unui vizitator concret.

O așa abordare va genera un randament maxim, deoarece, clienților care procură produse scumpe, acest tip de produse li se va prezenta în primul rând. Exact la fel și în cazul celor care procură produse ieftine. De asemenea, cei care procură mult, vor beneficia de reduceri, iar cei care doar comandă dar nu procură nimic, nu vor mai putea face comenzi.

Analiza parametrilor de conversie și adaptarea

automată a sistemului

Încă o funcție de bază este cea de analiză a rezultatelor obținute în urma aplicării metodei și adaptarea automată a bazei de reguli și a rețelei neuronale utilizate. Acest lucru se va face în baza analizei sumelor vânzărilor pe ultima perioadă în comparație cu alte perioade anterioare.

Dacă vânzările descresc, regulile generate de rețeaua neuronală nu sunt calitative și se intervine pentru a o adapta, iar în cazul când acestea se mențin la același nivel sau cresc, nu este nevoie de modificat nimic.

V. ANALIZA INFORMAȚIEI DESPRE UTILIZATOR ȘI DETERMINAREA PRODUSELOR DE CARE ARE NEVOIE

Despre fiecare utilizator va fi păstrată următoarea informație:

- numele și datele de contact introduse;
- vârsta;
- sexul;
- poziționarea geografică;
- platforma(sistemul de operare, browserul, etc.);
- adresa URL de pe care a venit pe vizitatorul pe site;
- cuvintele cheie utilizate în motorul de căutare.

Al doilea compartiment va prezenta produsele în dependență de vârsta, sexul și interesele utilizatorului.

De exemplu, produsele sunt împărțite în următoarele categorii:

- pentru tineri;
- pentru persoanele de vârstă medie;
- pentru persoanele de vârstă înaintată.

Aici nu există o delimitare foarte clară, de aceea datele acumulate despre vârsta utilizatorilor nu pot fi utilizate direct.

De exemplu, dacă un potențial cumpărător are vârsta de 32 ani, nu putem spune cu certitudine că el face parte din grupa persoanelor tinere, sau din grupa persoanelor de vârstă medie. Acest utilizator poate procura atât un produs pentru tineri cât și unul pentru persoane de vârstă medie, deoarece nu există o limită clară de trecere între aceste categorii de vârste.

Chiar dacă ar fi făcută o delimitare clară, aceasta nu va da un rezultat satisfăcător. Presupunem că grupele de vârste au următoarele limite bine definite:

- tineri – până la 35 ani;
- de vârstă medie – de la 35 până la 45 ani;
- de vârstă înaintată – mai mult de 45 ani.

Aplicând aceste limite, cumpărătorul de 32 ani, aparține primei grupe(tineri). Dacă pe site există destule produse din această categorie, nu există nici o problemă. Însă, în cazul în care nu există produse din această categorie, dar există produse din categoria a 2-a(de vârstă medie) produsele din această grupă nu vor fi propuse cumpărătorului, ceea ce nu este chiar corect, deci, apare o problemă.

Această problemă ar putea rezolvată aplicând algoritmi cu logică fuzzy.

Inițial, se va realiza distribuția fuzzy a vârstelor cumpărătorilor(fig. 2).

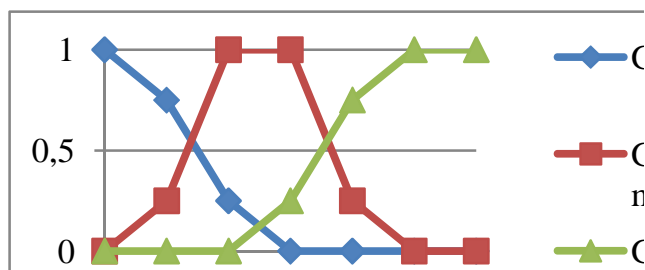


Fig. 2. Diagrama Fuzzy a distribuției vârstei.

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

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

Pentru persoanele de vîrstă medie se observă o mică creștere (0,2) a funcției de apartenență de la 18 la 24 ani, și 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 descreește, ajungînd la 0 pentru vârsta de 64 ani.

Astfel, se va crea baza de reguli [3, 4], care va transforma datele de intrare în date de ieșire în felul următor:

IF age = young THEN content = for young

IF age = medium THEN content = for medium

IF age = old THEN content = for old

În așa mod, la intrare sistemul primește vârsta numerică a utilizatorului, utilizează baza de reguli creată și determină categoria(grupul) din care i se va prezenta conținutul.

Deci, cumpărătorului cu vârsta de 32 ani i se va prezenta produsul cu funcția de apartenență la grupa "pentru tineri" cu valoare maximă. În cazul în care funcția nu va da rezultate, va fi prezentat conținutul cu la care funcția de apartenență la grupa "de vîrstă medie" va returna valoarea maximă. Astfel, probabilitatea că cumpărătorului i de va propune un produs care ar satisface interesele sale de vîrstă - crește, astfel crescînd și numărul produselor vîndute.

VI. CONCLUZII

Metoda propusă permite selectarea automată a conținutului optim pentru a fi prezentat utilizatorului sitului. Avantajul principal este că, fiecărui utilizator i se vor prezenta acele produse pe care cel mai probabil le va procura. De asemenea, metoda propusă pune la dispoziția administratorului un mecanism simplu de creare a regulilor de prezentare a conținutului. Administratorul va scrie o regulă de prezentare în formatul

DACĂ

cumpără toate produsele comandate

ATUNCI

oferim o reducere de preț;

iar algoritmul elaborat va determina automat tipul utilizatorului și îi va prezenta un anumit conținut.

Prezentarea conținutului personalizat are o importanță mare, îndeosebi în cazul siturilor comerciale (internet magazine, portaluri e-commerce. etc.).

Metoda propusă poate maximiza profitul, ca urmare a creșterii numărului de produse vîndute.

În cazul în care numărul de pagini (produse) expuse pe site este de ordinul sutelor sau chiar miilor, o căutare eficientă a produsului necesar poate deveni o reală problemă pentru utilizator, fiind posibilă și varianta cea mai pesimistă: produsul necesar există pe site, însă utilizatorul nu îl poate găsi. Aplicînd metoda descrisă în articol, această problemă poate fi rezolvată, deoarece, din baza de date vor fi selectate cele mai relevante produse, acestea fiind propuse utilizatorului în primul rînd, fără ca acesta să mai fie nevoit să utilizeze elementele de navigare și căutare a sitului.

REFERINȚE

- [1] Как использовать персонализированный контент и поведенческий таргетинг для поднятия конверсии, Available: <http://uniofweb.ru/analytics/1024/>
- [2] Новая платформа поиска Яндекса с персональными результатами: «Калининград», Available: <http://habrahabr.ru/company/yandex/blog/162245/>
- [3] L. A. Zadeh. *Fuzzy sets, Information and Control*, vol. 8 (1965), pp. 338-353.
- [4] L. A. Zadeh. *Computing With Words. Principal Concepts and Ideas* (2012), pp. 21-24.
- [5] C. A. Smith. *Automated continuous process control* (2002), pp. 115-122.