Utilizarea seturilor de date de foarte mare volum în studiul sistemelor colaborative

ION IVAN, PhD
ACADEMY OF ECONOMIC STUDIES BUCHAREST
Email: ionivan@ase.ro

Cristian CIUREA
ACADEMY OF ECONOMIC STUDIES BUCHAREST
Email: cristianciurea@yahoo.com

Abstract. În acest articol se prezintă cerințele de studiu pentru sisteme claborative, se dimensionează structura şi volumul de date necesare analizei sistemelor colaborative, se definesc proceduri de colectare şi validare a datelor, se enumeră algoritmi de construire a subcolectivităților omogene, se efectuează calcule cu seturi foarte mari de date şi se interpretează rezultatele.

1. DINAMICA SISTEMELOR COLABORATIVE

Sistemele colaborative reprezintă un domeniu interdisciplinar aflat la intersecția economiei, informaticii, managementului şi sociologiei. Colaborarea implică organizații care au o misiune comună şi care se unesc pentru a forma o noua structură. Un sistem colaborativ informatic este asemenea unei firme de distribuție al cărei obiectiv este de a vinde cantități din ce în ce mai mari din produsele sale.

Sistemele sunt formate din componente şi interacțiuni între acestea. Când sistemele colaborative sunt folosite în mod voluntar, unul dintre cei mai importanți conducători către succes este modul în care utilizatorii simt experiența lor cu sistemul, şi anume: dacă le place, dacă sistemul le oferă ceea ce aşteaptă de la el, dacă sunt capabili să comunice natural şi liber cu ceilalți participanți şi dacă doresc să-l recomande celorlalți.

Sistemele colaborative se ***clasifică*** după următoarele criterii:

a) *nivel de complexitate;* iar după acest criteriu se identifică:

- *sisteme colaborative cu nivel redus de complexitate,* au puține componente şi numărul de relații este restrâns;
- *sisteme colaborative cu nivel mediu de complexitate*, au număr redus de componente, dar nu au număr mare de fluxuri sau sisteme cu număr mare de fluxuri şi care au număr mare de componente;
- *sisteme colaborative* mari sau *cu grad mare de complexitate;*
- *sisteme colaborative deosebit de complexe*, au foarte multe componente şi foarte multe fluxuri: bănci, poliție, lanț intern de hoteluri, linii aeriene de transport. Sistemul bancar se află printre sistemele colaborative cu nivel foarte mare de complexitate, deoarece este format dintr-un număr mare de componente şi este caracterizat de o diversitate mare a legăturilor dintre acestea.

b) *tip de aplicare;* criteriu ce grupează sistemele în*:*

- *sisteme colaborative în educație;*
- *sisteme colaborative ale apărării;*
- *sisteme colaborative productive.*

c) *mod de organizare;* criteriu care le împarte în:

- *sisteme liniare,* în care subsistemele intracționează unele cu altele, în ambele sensuri;

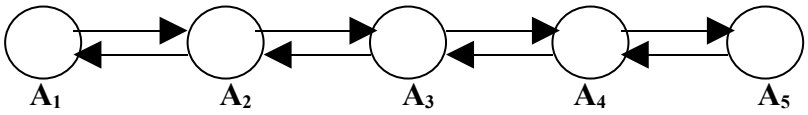

Fig. 1. Sistem colaborativ liniar

între activitățile $\mathbf{A_1}$ și $\mathbf{A_2}$ se schimbă mesajul $\mathbf{M_1}$, între $\mathbf{A_2}$ și $\mathbf{A_3}$ se schimbă mesajul $\mathbf{M_2}$, între $\mathbf{A_3}$ și $\mathbf{A_4}$ se schimbă mesajul $\mathbf{M_3}$, iar între $\mathbf{A_4}$ și $\mathbf{A_5}$ se schimbă mesajul $\mathbf{M_4}$;

- *sisteme arborescente;* organizate pe niveluri, ca în figura 2:

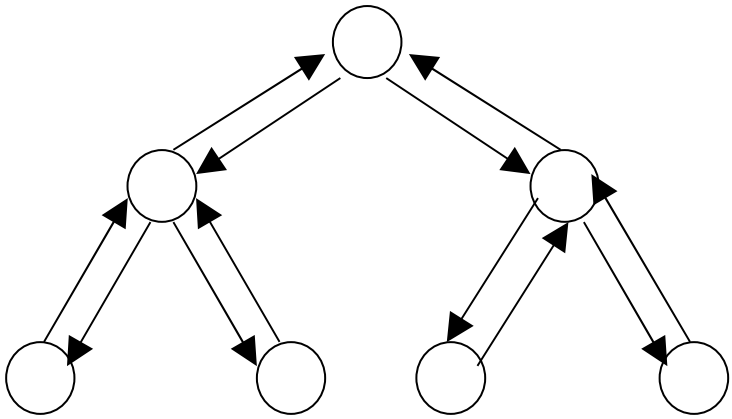

Fig. 2. Sistem colaborativ arborescent

într-un sistem arborescent, mesajele circulă între activități în sens ierarhic, adică un mesaj de pe nivelul doi va ajunge la nivelul zero numai dacă trece și pe la nivelul unu, iar un mesaj al activității de bază, reprezentată prin rădăcina arborelui, va fi propagat doar către activitățile aflate pe nivelul imediat inferior. De pe acest nivel, mesajul respectiv va fi transmis activităților reprezentate de nodurile copil ale nodurilor de pe nivelul unu;

- *sisteme rețea;* în care componentele comunică între ele indiferent de nivelul pe care se află;

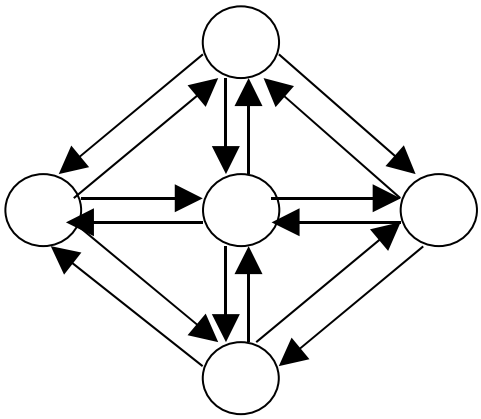

Fig. 3. Sistem colaborativ de tip rețea

Dinamica sistemelor colaborative vizează modificări legate de calitatea, structura, funcțiile, dimensiunea, procedurile şi standardele acestora. Dinamica sistemelor colaborative se studiază folosind analiza matematică, prevăzând comportamentul pe termen lung al fiecăror sisteme mari, câştigând o privire în interiorul designului sistemului: care parametri determină comportamentul de grup şi cum sunt afectate caracteristicile sistemului. Se dezvoltă o clasă de modele matematice care descriu dinamica colectivă a sistemului colaborativ şi care ilustrează abordarea prin aplicarea la mai multe studii de caz, care include atât agenți software, cât şi roboți. Pentru fiecare sistem, se transformă un set de ecuații care descriu cum se schimbă sistemul în timp şi analizează soluțiile lor. În final se elucidează ce spun aceste soluții despre comportamentul sistemului colaborativ.

2. DECIZII ÎN SISTEME COLABORATIVE

Sistemul decizional este foarte strâns legat de sistemul informațional. Legătura dintre sistemul decizional şi cel informațional se prezintă astfel:

- în primul rând, la intrare, sistemul informațional furnizează sistemului decizional datele de care acesta are nevoie şi pe care le prelucrează dându-le forma adecvată pentru a fi folosite în cadrul procesului decizional;
- în al doilea rând, datele, odată prelucrate şi luată decizia, aceasta este preluată de către sistemul informațional şi dirijată spre compartimentele unde ea trebuie sa ajungă pentru a fi pusă în practică;
- în al treilea rând, sistemul decizional necesită chiar în interiorul său un circuit intermediar de informații în condițiile în care sistemul colaborativ este condus prin intermediul managementului participativ.

Deciziile referitoare la sistemele colaborative se împart în:

- *decizii curente,* care se referă la deciziile zilnice, contribuie la realizarea obiectivelor individuale şi predomină la nivelul managementului mediu şi inferior. Intrările şi ieşirile unor astfel de decizii vizează problemele de zi cu zi care au loc într-un sistem colaborativ. Costurile implicate de luarea unei decizii curente greşite sunt mici, iar riscurile sunt, de asemenea, reduse;
- *decizii pe termen scurt,* care au în vedere deciziile pe perioade de maxim câteva luni şi care contribuie la realizarea obiectivelor de grup şi predomină la nivelul managementului mediu şi inferior. Costurile implicate de luarea unor astfel de decizii sunt mai ridicate decât în cazul deciziilor curente;
- *decizii pe termen mediu,* care au în vedere deciziile pe perioade cuprinse între câteva luni şi un an, contribuie la realizarea obiectivelor de grup şi predomină la nivelul managementului mediu şi superior. Costurile şi riscurile asumate în cazul unei decizii pe termen mediu eronate sunt destul de mari;
- *decizii pe termen lung,* care se referă la deciziile pe perioade cuprinse între 1 şi 3 ani, contribuie la realizarea obiectivelor de grup şi predomină la nivelul managementului superior. O decizie greşită de acest gen presupune riscuri şi costuri foarte mari şi are consecințe catastrofale asupra sistemului colaborativ implicat;
- *decizii strategice,* care se referă la o perioadă cuprinsă între 3 şi 5 ani şi care contribuie nemijlocit la realizarea obiectivelor fundamentale sau derivate şi vizează activitățile de ansamblu ale sistemului sau principalele sale componente.

Decizia economică reprezintă linia de acțiune aleasă în mod conştient în cadrul procesului de conducere al unui sistem dintr-un număr oarecare de posibilități în scopul atingerii unor obiective în condiții de eficiență maximă. Decizia pune în lucru resursele, stabileşte şi atinge obiectivele sistemului. Necesită o activitate de pregătire şi elaborare la care participă un număr mare de persoane şi compartimente din cadrul sistemului.

Mediul ambiant decizional constă în ansamblul elementelor eterogene si exogene unui sistem colaborativ, care alcătuiesc situaţia decizională caracterizată prin manifestarea unor influenţe directe şi indirecte semnificative asupra conţinutului şi rezultatelor deciziei. În mediul ambiant decizional se constată o evoluţie contradictorie: pe de o parte se înregistrează o serie de transformări de natură să ofere premise mai bune pentru un proces decizional eficient, iar pe de altă mediul ambiant decizional tinde să devină din ce în ce mai complex datorită adâncirii diviziunii sociale, reducerea ciclului de viaţă al produselor şi accelerarea ritmului de uzură morală. Pe plan decizional, aceste elemente se traduc într-un număr sporit de variabile şi condiţii limită şi în implicarea interdependenţelor dintre acestea.

3. PROCEDURI DE COLECTARE A DATELOR

Încă de la apariţia primelor calculatoare perfect operaţionale în anii 1950, oamenii au înţeles puterea acestora de a colecta, manipula, clasifica, stoca şi regăsi date într-un mod mult mai rapid, mai flexibil şi mai eficient decât puterea umană. Volumul şi complexitatea informaţiilor depăşesc capacitatea de prelucrare a unei persoane. Acesta este motivul pentru care sistemele informaţionale permit stocarea datelor pentru o analiză ulterioară şi oferă metodele de procesare şi filtrare pentru a extrage numai datele relevante. Mai mult, cu ajutorul algoritmilor matematici şi statistici ce implementează metode cunoscute pentru înţelegerea datelor se obţin informaţii cu înţelesuri noi.

Calculatoarele au devenit indispensabile pentru colectarea şi manipularea informaţiilor, atunci când informaţiile se constituie în volume deosebit de mari.

Odată cu specificarea proiectului logic al unui sistem colaborativ, se recurge şi la proiectarea procedurilor de culegere, validare, transmitere, stocare şi prelucrare a datelor de intrare. Intrările într-un sistem colaborativ se realizează prin preluarea datelor:

- din documente primare, introduse de operatori;
- transmiterea prin toate modalităţile, inclusiv prin satelit;
- prin scanare;
- sub formă de impulsuri de la instalaţiile şi maşinile din procesul tehnologic.

Procesarea documentelor presupune extragerea informaţiilor conţinute în câmpurile dată si convertirea lor ca succesiuni de biţi, format ce permite stocarea lor într-o bază de date. Procesarea formularelor se consideră completă în momentul în care toate informaţiile din documente au fost extrase, verificate şi salvate într-o bază de date. Colectarea de date de pe formulare se face fie implicând un număr apreciabil de persoane care să le culeagă şi să le introducă într-un produs program, folosind tastatura computerelor, fie prin introducerea automată a datelor în sistem.

Printre beneficiile rezultate din folosirea unei soluţii de culegere automată a datelor din formulare tipărite pe suport de hârtie se enumeră:

- creşterea eficienţei şi a preciziei în preluarea şi procesarea informaţiilor;
- diminuarea numărului de personal utilizat într-o astfel de activitate şi implicit, minimizarea erorilor umane;
- accesibilitate sporită a informaţiilor colectate şi usurinţă în utilizarea lor;
- stocarea imaginii documentelor pe suport magnetic conduce la reducerea substanţială a spaţiilor alocate clasicelor arhive, precum şi a costurilor de administrare aferente;
- scurtarea timpului de procesare a informaţiilor având în vedere realizarea mult doritului STP (Straight Through Processing);
- îmbunătăţirea calităţii serviciilor ofertate de către beneficiari clientilor lor.

O descriere succintă a modului de lucru dintr-o aplicaţie de culegere automată a datelor este următoarea:

- setul de formulare este scanat folosindu-se un scanner de viteză;
- majoritatea datelor este recunoscută în mod automat;
- un număr mic de caractere pe care programul nu este sigur ca le-a recunoscut face ca imaginea documentului scanat să fie trecută la un operator uman, pentru verificare şi corectare;
- datele verificate sunt salvate într-o bază de date.

Arhivele de documente impugn dezvoltarea de algoritmi flexibili de regăsire a componentelor sau a grupelor de componente pentru a corespunde necesarului de fundamentare a deciziilor colaborative.

Confidenţialitatea este asigurată prin faptul ca informaţia este livrată doar persoanei autorizate şi numai la o locaţie autorizată, deoarece controlul accesului şi autentificarea sunt foarte importante pentru desfăşurarea proceselor privind transferul mesajelor. Accesul la baza de date se face numai prin utilizarea unei proceduri de acces. Procedura de acces este ierarhizată pe nivele şi parole de acces parametrizate. Această procedură este pusă la dispoziţia utilizatorului în vederea personalizării.

De calitatea datelor de intrare într-un sistem colaborativ depinde calitatea informaţiilor oferite de sistem factorilor de decizie sau utilizatorilor, în general. Este de dorit a preîntâmpina, pe cât posibil, introducerea în sistem a datelor eronate. Procedurile de validare a datelor se grupează astfel:

- proceduri tradiţionale, care constau în verificarea ochiometrică a datelor din documentele primare;
- proceduri automate de control şi validare a datelor, prin care validarea se realizează prin programele de validare elaborate de programatori şi proceduri automate de control şi corecţie automată a datelor, prin care se identifică şi corectează erorile în mod automat, fără intervenţia factorului uman.

4. INDICATORI AI DINAMICII SISTEMELOR COLABORATIVE

Aceşti indicatori caracterizează subseturi de date omogene şi se împart în următoarele categorii:

- *indicatori care operează cu momente de timp;* într-un sistem colaborativ reprezentat de o *bancă* ce are o multitudine de sucursale şi agenţii amplasate în numeroase ţări, este interesant a se descoperi intervalele orare din cursul unei zile sau zilele dintr-o săptămână în care au loc cele mai multe retrageri de numerar; pentru a afla intervalul orar, format din două ore, în care au loc cele mai multe retrageri de numerar dintr-o zi, se măsoară, din două în două ore, volumul ridicărilor de numerar. În decursul unei zile există douăsprezece măsurători. Intervalul orar cu cele mai numeroase retrageri de numerar se determină astfel:

  $IHM = max\ \{IH_1, ..., IH_{12}\}$, unde:

  IHM - intervalul orar dintr-o zi, cu cele mai numeroase retrageri de numerar;
  IH_i - volumul retragerilor de numerar din intervalul *i* de timp;
  pentru a determina ziua cu cele mai numeroase retrageri dintr-o săptămână, se procedează asemănător, cu deosebirea că maximul se calculează dintre şapte valori, reprezentând volumul retragerilor zilnice;

- *indicatori care operează cu niveluri ale unei variabile;* în sistemul colaborativ *bancă,* indicatori care operează cu niveluri ale unei variabile sunt volumul de depuneri, respectiv volumul de retrageri zilnice;

- *indicatori care operează cu momente de timp şi niveluri ale unei variabile.*
  Aceşti indicatori se constituie într-o metrică a sistemelor colaborative.

Evoluţia proceselor economice se reprezintă prin modele dinamice continue: prin ecuaţii diferenţiale sau prin sisteme de ecuaţii diferenţiale, după cum se evidenţiază un singur indicator principal sau un ansamblu de indicatori corelaţi prin ecuaţiile modelului, atât între ei, cât şi cu variabilele factoriale care condiţionează procesul. Modele dinamice continue liniare ale sistemelor cibernetice se întâlnesc frecvent în cercetarea dinamicii proceselor economice şi se reprezintă prin ecuaţii diferenţiale liniare.

În *[QUIN99]* se prezintă implementări ale algoritmilor de clasificare a datelor, folosind metoda inductivă top-down pentru construirea arborilor de decizie. Aceştia se construiesc pe baza testării fiecărui nod al arborelui, începând cu nodul rădăcină, pentru fiecare înregistrare. Fiecare nod reprezintă numele unui atribut. Se introduce instanța într-o clasă existentă, pe baza caracteristicilor comune, evaluându-se atributul corespunzător nodului la care s-a ajuns. În funcție de valoare, instanța urmează o ramură. Când nu mai există noduri de evaluat, instanța este clasificată. Dacă o anumită clasă nu mai diferă într-un mod evident de alta în urma introducerii a tot mai multe înregistrări, cele două se vor uni, proces recunoscut sub denumirea de proces *pruning*.

5. SOFTWARE PENTRU MANAGEMENTUL SETURILOR DE DATE DE FOARTE MARI DIMENSIUNI

Păstrarea unor volume uriaşe de date la nivelul unei firme şi regăsirea acestor date a condus în anii 1990 la dezvoltarea de noi tehnologii, construind:

- depozite de date;
- concentrări de date;
- extrageri de date.

Inteligenței artificiale îi revine un rol special în a dezvolta şi implementa tehnologii pentru crearea şi manipularea seturilor de date extrem de mari şi cu nivel de complexitate foarte ridicat *[BODE01]*.

Corespunzător au fost dezvoltate o serie de produse software pentru gestiunea volumelor foarte mari de date, şi anume:

- Oracle Datawarehouse, utilizat pentru organizarea datelor în depozite de date după modelul multidimensional;
- Oracle OLAP (**O**n **L**ine **A**nalytical **P**rocessing), pentru regăsirea datelor din depozitele de date, folosind tehnologia OLAP;
- Oracle Discoverer, pentru obținerea unor situații de ieşire din depozitele de date, inclusiv obținerea de rapoarte dinamice;
- Oracle Miner, folosit pentru statistica avansată, descoperirea, cunoaşterea şi extragerea informațiilor din date.

Există produsul WEKA, the Waikato Environment for Knowledge Analysis, pus la dispoziție gratuit de catedra de specialitate a Universității Waikato din Hamilton, Noua Zeelandă, produs ce permite aplicarea tehnicilor de învățare automată asupra problemelor practice şi integrează diverse unelte pentru învățarea automată ce sunt utilizate într-un mediu de lucru uzual, caracterizat de o interfață omogenă. Utilizatorii folosesc gama largă de tehnici de învățare automată pentru extragerea unor informații utile din baze de date foarte mari. Trebuie precizat faptul că WEKA se utilizează în orice domeniu de interes, având astfel un avantaj major asupra celorlalte aplicații de *data mining*, mai ales asupra celor comerciale care sunt destinate unui singur domeniu de activitate. WEKA conține *unelte* pentru preprocesarea datelor, iar pentru clasificarea acestora se utilizează arbori de decizie, regresie, clusterizare, reguli de asociere si vizualizare. Aplicația este dezvoltată în Java, iar codul sursă este deschis, eliberat sub licență GNU General Public License. Acesta este un mare avantaj al sistemului WEKA spre deosebire de alte aplicații, deoarece permite modificarea sistemului de către utilizatori în modul în care aceştia au nevoie de el, eventual cu dezvoltarea de noi tehnici de învățare automată şi implementarea de algoritmi proprii. De asemenea, la fel de important este faptul că sistemul se utilizează pe mai multe platforme: Unix, Linux şi Microsoft Windows.

Ultima *versiune* pusă la dispoziția utilizatorilor este WEKA 3.4.3 şi se instalează atât pe platforma Windows cât şi pe alte platforme, cum ar fi Linux sau Unix. Trebuie menționat faptul că pentru MacOS X nu este disponibilă, deocamdată, decât versiunea WEKA 3.4.2. Pentru a rula WEKA trebuie să existe instalată pe sistem maşina virtuală Java 1.4. O versiune anterioară a WEKA este

WEKA 3.0, ce se bazează pe lucrul în linie de comandă. La *lansarea WEKA* apare fereastra *GUI Chooser* care permite utilizatorilor să opteze pentru lucrul în linie de comandă *CLI* sau pentru deschiderea lucrului în interfața grafică, *Explorer*. WEKA Explorer pune la dispoziție în interfața grafică pachetele sistemului, şi anume:

- *preprocessing,* în cadrul căruia se deschid seturile de date atât sub forma fişierelor ARFF, cât şi dintr-o bază de date anume; de asemenea, sistemul permite o filtrare nesupravegheată a datelor cu unul din filtrele puse la dispoziție;
- *classify,* ce permite alegerea şi rularea oricărui algoritm de clasificare din cele şase categorii de algoritmi definite;
- *cluster,* în cadrul căruia se alege şi se rulează metoda de clusterizare a datelor;
- *associate,* ce permite setarea unei reguli de asociere a datelor şi aplicarea acesteia;
- *select attributes* este un alt pachet WEKA şi permite configurarea şi aplicarea oricărei combinații de atribute din cele ce definesc setul de date pentru a depista care sunt cele mai relevante atribute din set;
- *visualize* permite vizualizarea setului current de date în una sau două dimensiuni, iar dacă atributele au valori continue este utilizat un spectru de nuanțe ale aceleiaşi culori pentru reprezentarea valorilor, pe când pentru atribute discrete fiecare valoare este reprezentată cu altă culoare. Suplimentar acestor pachete de instrumente pentru lucrul cu seturi de date, WEKA conține şi un clasificator pe bază de arbori de decizie WEKA CLASSIFIERS TREES USERCLASSIFIER şi o interfață grafică pentru realizarea de rețele neuronale WEKA CLASSIFIERS FUNCTIONS NEURAL NEURALNETWORK.

Setul de date utilizat în mediul de programare WEKA trebuie să fie în format ARFF pentru a fi prelucrat. Datele provin de cele mai multe ori dintr-o tabelă Excel sau dintr-o bază de date şi trebuiesc convertite în formatul ARFF, cel mai larg răspândit pentru baze de date în fişiere text. Folosirea acestui format în paralel cu suportul direct pentru baze de date este un alt avantaj al WEKA. Pe lângă aceste elemente favorabile ce caracterizează sistemul WEKA, există şi câteva dezavantaje, şi anume faptul că necesită învățarea utilizării interfeței, înțelegerea algoritmilor şi a modului de interpretare a rezultatelor numerice şi grafice. În plus, WEKA foloseşte termeni statistici în loc să folosească termeni corespunzători datelor de intrare (de exemplu, din aplicațiile economice) aşa cum fac alte produse software specializate pe mediul de afaceri şi mult mai intuitive pentru un manager sau economist.

Pentru evaluarea sistemelor colaborative se implementează un produs software care permite obținerea nivelurilor pentru metrica asociată.

Calitatea datelor este asigurată în situația analizei şi prelucrării unui volum foarte mare de date. Asupra acestor date se execută operații precum sortări, regrupări, concatenări sau eliminări de valori aberante. Ca şi în cazul unui set de date redus ca volum, în cazul unui volum de date foarte mare se calculează anumiți indicatori şi se evidențiază anumite caracteristici ale acestuia pe baza următoarelor intervale de valori, obținute empiric:

- *[0; 0,78]* este intervalul care exprimă caracterul rău al indicatorului calculat;
- *(0,78; 0,92]* exprimă caracterul bun al indicatorului;
- *(0,92; 1]* exprimă caracterul foarte bun al respectivului indicator.

Se presupune că pentru un set de date de 300 de valori se calculează următorii indicatori ale căror valori obținute sunt prezentate în tabelul 1:

Indicator 1	*Indicator 2*	*Indicator 3*	*Indicator 4*	*Indicator 5*	*Indicator 6*	*Indicator 7*	*Indicator 8*
C_1	G_1	H_1	P_1	A_1	F_1	S_1	B_1
C_2	G_2	H_2	P_2	A_2	F_2	S_2	B_2
C_3	G_3	H_3	P_3	A_3	F_3	S_3	B_3
...	...	...	...	...	...	...	...
C_{300}	G_{300}	H_{300}	P_{300}	A_{300}	F_{300}	S_{300}	B_{300}

Tabelul 1. Analiza unui set de date

În cazul în care valorile indicatorilor din tabelul 1 nu se încadrează în intervalul [0; 1] se va proceda la normalizarea indicatorilor, conform relației de calcul:

$$VN_i = \frac{V_i - V_{\min}}{V_{\max} - V_{\min}} \in [0; 1], \text{ unde:}$$

VN_i - valoarea normalizată a indicatorului i;
V_i - valoarea indicatorului i;
V_{max} - valoarea maximă a indicatorului i;
V_{min} - valoarea minimă a indicatorului i.

Valorile normalizate ale indicatorilor se sortează descrescător. Se alege o distanță h şi se determină un element I_{min}. Se calculează nivelurile $I_{min} + h$, $I_{min} + 2h$, $I_{min} + 3h$, $I_{min} + 4h = I_{max}$ şi se construiesc seturile de date ce aparțin intervalelor:
$[I_{min}; I_{min} + h)$,
$[I_{min} + h; I_{min} + 2h)$,
$[I_{min} + 2h; I_{min} + 3h)$,
$[I_{min} + 3h; I_{min} + 4h)$.

Seturile sunt mai omogene şi permit aplicarea indicatorilor definiți în cadrul metricii, obținând niveluri reprezentative.

6. CONCLUZII

Creşterea volumului de informații, precum şi perfecționarea produselor software de exploatare a acestuia au condus la o nouă calitate a folosirii datelor prin analize care relevă conducerii organizației informații greu sau chiar imposibil de obtinut pe alte căi. Se obțin astfel informații privind preferințele clienților, profilul lor sau distribuția. Astfel se furnizează conducerii date, cum ar fi: în ce regiune a țării se vinde mai bine un anumit produs, care sunt preferințele unui anumit segment de piață.

Este evident că astfel de informații nu se obțin decât folosind anumite prelucrări, cum ar fi analiza multidimensională, anumite metode statistice de prognoză şi alte metode matematice aplicate unui volum foarte mare de date. Aceste metode matematice reclamă folosirea unui software specializat deosebit de complex.

Din perspectivã economicã, globalizarea comerțului, ascuțirea dramaticã a concurenței, scurtarea spectaculoasã a ciclurilor de viațã a produselor datoritã dinamicii tehnologiei, impunerea unor cerințe calitative extrem de ridicate, precum şi alte asemenea evoluții, au evidențiat şi mai mult valoarea strategicã a informației. Manipularea operativã a informației a impus la rîndul ei noi modele manageriale, mai suple şi mai funcționale. Nevoia de a rãspunde în timp optim cerințelor pietei a condus la descentralizare şi la reducerea numãrului nivelelor decizionale, consacrând aşa-numitele *ierarhii plate,* care se bazeazã pe delegarea puterii decizionale operative cãtre eşalonul managerial secund. Practic, clasicul funcționar este pe cale de a fi înlocuit de *lucrãtorul cu informații*. Atât la nivelul conducerii strategice, cât şi la nivelul managementului operativ, nevoia de informație purã, corectã şi semnificativã a devenit vitalã pentru sistemele colaborative.

BIBLIOGRAFIE

[ARBA05] Raluca ARBA – *Collaborative Electronic Marketplace*, in International Workshop „Collaborative Support Systems in Business and Education”, Cluj-Napoca, october 2005,

pag. 11.

[BODE01] Constanta BODEA – *Baze de date: tendințe în evoluția tehnologiilor şi aplicațiilor*, Editura Inforec, Bucureşti, 2001.

[DOBR05] Ovidiu DOBRICAN – *An Example of Collaborative System*, in International Workshop „Collaborative Support Systems in Business and Education", Cluj-Napoca, october 2005, pag. 48.

[FAYY96] Fayyad U., Piatetsky-Saphiro G., Smyth P., Uthurusamy R. - *Advances in knowledge discovery and data mining,* Cambridge, MA, 1996, London.

[IVAN07] Ion IVAN, Cătălin BOJA, Cristian CIUREA – *Metrici ale sistemelor colaborative*, lucrare în curs de apariție, predată Editurii ASE.

[QUIN99] Ross QUINLAN - *Data Mining from an AI Perspective*, Proceedings of the 15th International Conference on Data Engineering, 23-26 March 1999, Sydney, Austrialia.

[ROBB51] Robbins H., Monro S. - *A stochastic approximation method,* Ann. Math. Stat. 22, 1951.