

Aplicarea tehnicilor de Data Mining pentru creșterea calității terapiilor logopedice personalizate

Mirela Danubianu, Tiberiu Socaciu

Rezumat—In ultimii ani s-a acordat o atenție sporită tulburărilor de limbaj care afectează un procent semnificativ al copiilor preșcolari sau din primii ani de școală, făcându-se investiții consistente pentru dotarea cabinetelor logopedice cu tehnică de calcul. Utilizarea aplicațiilor informatice în scopul asistării și urmării terapiei logopedice a permis ca la dispoziția cercetătorilor să se găsească un volum considerabil de date. Volumul crescut de date disponibile nu a condus imediat la un volum similar de informații care să constituie suport al unor decizii eficiente de terapie, deoarece metodele clasice de prelucrare a datelor nu sunt aplicabile în astfel de cazuri.

Cuvinte cheie—data mining, terapie logopedică personalizată, tulburări de limbaj

I. INTRODUCERE

ÎN perioada actuală, din ce în ce mai multă atenție este acordată copiilor cu tulburări de limbaj iar investițiile în terapia logopedică sunt în creștere. Urmare a studiilor efectuate s-a ajuns la concluzia că sunt copii care inițial pot prezenta probleme complexe difuze care însă evoluează în dificultăți mai specifice odată cu creșterea. Pare să existe o asociere puternică între problemele timpurii de limbaj și dificultățile ulterioare de exprimare. Dificultatea pentru cercetători și terapeuți constă în identificarea acelor copii care au tulburări ce indică o gamă largă de probleme care nu se vor rezolva în mod spontan sau care pot conduce la diferite deficiențe ulterioare. Pe baza cercetărilor tiparelor de dezvoltare a vorbirii și limbajului acești copii ar putea fi identificați astfel încât terapia să fie dirijată mult mai eficient.

Dezvoltarea tehnologiilor informației și utilizarea aplicațiilor informatice în scopul asistării și urmării terapiei logopedice a permis ca la dispoziția cercetătorilor să se găsească un volum considerabil de date. Acest lucru a fost

posibil pe de o parte datorită dezvoltării tehnologiei bazelor de date, pe de altă parte datorită evoluției mediilor de stocare, care au capacitatea de a păstra cantități impresionante de date. Volumul crescut de date disponibile nu a condus imediat la un volum similar de informații care să constituie suport al unor decizii eficiente de terapie, deoarece metodele clasice de prelucrare a datelor nu sunt aplicabile în astfel de cazuri.

II. DATA MINING

A. Ce este Data Mining

Data mining presupune aplicarea unor analize asupra volumelor mari de date și descoperirea algoritmilor care, în limite acceptabile ale eficienței de calcul produc o enumerare particulară a tiparelor din aceste date, poate facilita descoperirea, în date aparent fără legătură, relații prin care se pot anticipa eventuale probleme sau care ar putea rezolva problemele studiate. Deci, data mining este definită ca operația de extragere a informațiilor interesante și anterior necunoscute și reprezintă o singură fază în procesul complex de descoperire a cunoștințelor în bazele de date (KDD - Knowledge Discovery in Databases). Acest proces constă în următoarea succesiune de pași:[8]:

- *înțelegerea domeniului aplicației* și a cunoștințelor anterioare relevante și identificarea scopului procesului KDD din punctul de vedere al utilizatorului.
- *crearea unei mulțimi de date țintă* care presupune selectarea unei mulțimi de date sau focalizarea asupra unui subset de eșantioane de date pe care va fi realizată descoperirea.
- *curățarea și pregătirea datelor pentru prelucrare*, etapă care constă în operații precum înlăturarea zgomotelor colectarea informațiilor necesare pentru a modela sau a calcula zgomotele, decizii asupra strategiilor de manipulare a datelor incomplete din câmpuri, contabilizarea secvențierii în timp a informațiilor și a schimbărilor efectuate
- *reducerea și proiecția datelor* constă în depistarea caracteristicilor utile pentru reprezentarea datelor în concordanță cu scopul propus. Prin utilizarea de metode

M. Danubianu – Universitatea ‘Ștefan cel Mare’ Suceava, e-mail: mdanub@eed.usv.ro.

T. Socaciu - Universitatea ‘Ștefan cel Mare’ Suceava, e-mail: tibisocaciu@yahoo.com

care reduc dimensiunile sau de metode de transformare, se reduce numărul de variabile luate în considerare sau se găsesc reprezentările invariante ale datelor

- *explorarea datelor (data mining)* care constă în căutarea tiparelor interesante într-o formă de reprezentare particulară sau într-un set de astfel de reprezentări: reguli sau arbori de clasificare, regresie, clustering, reguli de asociere și altele. În concordanță cu scopul procesului de descoperire a cunoștințelor din date se va folosi una din metodele particulare de explorare a datelor. Un rol important îl poate juca utilizatorul care poate corecta performanțele pașilor anterior parcurși,
- *interpretarea tiparelor* extrase, posibilă întoarcere la oricare din pașii anteriori pentru iterații viitoare. Acest pas poate implica de asemenea vizualizarea tiparelor/modelelor extrase.
- *consolidarea cunoștințelor* extrase și încorporarea acestora într-un alt sistem pentru acțiuni viitoare sau simpla documentare și raportare a acestora către părțile interesate. În această etapă se încearcă rezolvarea potențialelor conflicte cu cunoștințele acceptate anterior.

Procesul de descoperire a cunoștințelor din date poate implica iterații semnificative și poate conține bucle între oricare doi pași. Cele mai multe lucrări care fac referire la procesul descoperirii cunoștințelor din date și-au focalizat atenția asupra pasului de explorare a datelor. Totuși, ceilalți pași sunt la fel de importanți în practică, pentru succesul aplicațiilor de descoperire a cunoștințelor din date.

B. Metode și tehnici de Data Mining

Deoarece explorarea datelor impune numeroase aspecte provocatoare pentru cercetare, aplicarea directă a metodelor și tehnicilor dezvoltate de domenii înrudite precum învățarea automată, statistica sau bazele de date nu pot rezolva problemele apărute. În acest scop este necesară efectuarea de studii dedicate pentru a găsi noi metode sau tehnici integrate pentru o explorare eficientă. Ca urmare a acestui fapt, data mining s-a constituit ea însăși într-un domeniu aparte de cercetare.

Rezultatele obținute în urma aplicării algoritmilor de data mining pot furniza răspunsuri la două mari categorii de probleme: predicție respectiv descrierea. Deși limitele dintre predicție și descriere nu sunt foarte tranșante – în sensul că anumite modele predictive pot fi descriptive până în măsura în care acestea pot fi inteligibile și vice-versa- distincția este utilă pentru a înțelege obiectivul global al descoperirii. În acest context, este esențial de luat în considerare importanța relativă între predicție și descriere pentru o aplicație particulară de explorare a datelor. Sunt numeroase situațiile când, în contextul descoperirii cunoștințelor din date, descrierea tinde să fie mai importantă decât predicția, situație care este în contrast cu aplicațiile de învățare automată sau de recunoaștere a formelor unde predicția este adesea scopul principal. Obiectivele predicției respectiv ale descrierii sunt realizate prin intermediul următoarelor metode primare de explorare a datelor:

- *clasificarea*, care presupune găsirea unei funcții care include un articol de date într-una din mai multe clase predefinite.
- *regresia* este utilizată la prezicerea unei valori a unei variabile continue bazată pe valorile altor variabile, presupunând un model de dependență liniar sau neliniar. Regresia logică este utilizată pentru precizarea valorii unei variabile binare. Este un instrument de clasificare care este utilizat la precizarea valorii unei variabile continue, cum ar fi: "probabilitatea că un individ cu probleme de pronunție va fi corectat în urma terapiei".
- *gruparea (clustering-ul)* identifică o mulțime finită de categorii sau clustere pentru a descrie datele. Strâns legată de acesta este metoda estimării densității de probabilitate care constă în tehnici de estimare din date a funcției de probabilitate asociată pentru toate variabilele/câmpurile unei baze de date.
- *rezumarea* găsește o descriere compactă pentru o submulțime de date.
- *regulile de asociere* presupun găsirea unui model care descrie dependențele semnificative dintre variabile.
- *detectarea schimbărilor și a deviației* descoperă cele mai semnificative schimbări produse în date în intervalul dintre două măsurări consecutive.

În Figura 1. sunt prezentate schematic metodele de explorare a datelor și tipul problemelor de rezolvat pentru care acestea sunt potrivite.

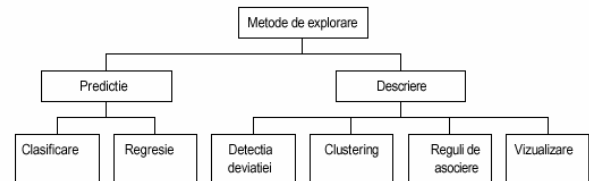


Fig. 1. Metode de explorare a datelor

Explorarea datelor poate fi gândită ca o formă specifică de învățare automată, în care mediul este observat prin intermediul unei baze de date

III. TULBURĂRI DE LIMBAJ

Limbaajul este cel mai folosit mijloc de comunicare interumană, îndeplinind în raport cu individul numeroase funcții. În primul rând îi permite acestuia să trăiască în comunitate și să se adapteze diferitelor situații. Este un mijloc prin care individul are posibilitatea de a se valoriza sau de a-și regla o conduită alterată. De mare importanță sunt: funcția de comunicare a limbajului, care face posibilă transferarea unui conținut de la o persoană la alta și funcția cognitivă de integrare, conceptualizare și de elaborare a gândirii.

Fiecare individ este supus acțiunii concertate a numeroși factori. Mulți dintre acești factori, nocivi fiind, pot influența negativ dezvoltarea normală a individului. În particular, există câteva categorii care pot afecta negativ dezvoltarea limbajului și care pot determina diferite disfuncții.

Este indicată o analiză a acestor factori atât în scopul de a preveni efectele lor nedorite cât și pentru a folosi metodele adecvate de stabilire a diagnosticului. În plus, această analiză

poate fi un punct de plecare în terapia personalizată.

Cauzele care determină tulburări de limbaj pot acționa în perioada intrauterină de dezvoltare a fătului, în timpul nașterii sau după nașterea individului și se pot încadra într-o multitudine de categorii: de la cauze organice de natură centrală sau periferică, la cauze funcționale, neurologice sau psiho-sociale.

La rândul lor, ținând cont de diferite criterii, tulburările de limbaj pot fi împărțite în: tulburări de pronunție - articulație (dislalie, rinolalie etc.), tulburări de ritm și fluență a vorbirii (bâlbâială), tulburări de voce sau tulburări de dezvoltare a limbajului. O parte din aceste tulburări, în special cele de pronunție-articulație, au o frecvență foarte crescută la copii. În cazul unor forme ușoare sau dacă se aplică la timp o terapie adecvată ele dispar odată cu înaintarea în vârstă.

Activitatea de intervenție logopedică își propune, printre altele, câteva obiective specifice, cum ar fi depistarea, evaluarea complexă și identificarea tulburărilor de limbaj și de comunicare la copiii preșcolari și la școlarii mici precum și orientarea terapiei logopedice spre corectare, recuperare compensare adaptare și integrare socială. Acest din urmă deziderat presupune aplicarea unei terapii personalizate fiecărui logopat sau grup de logopați cu caracteristici similare, terapie adaptată gravității afecțiunii și îndreptată spre eliminarea cauzelor care au generat-o.

IV. POSIBILE APLICAȚII ȘI LIMITE ALE TEHNICILOR DE DATA MINING ÎN DOMENIUL LOGOPEDIEI

Legat de activitatea logopedică, sarcinile executate de data mining pot fi grupate în următoarele categorii:

- *clasificarea* plasează logopații în segmente predefinite. În acest mod este posibilă monitorizarea dimensiunii și a structurii diferitelor grupuri.
- *clusteringul* grupează logopații pe baza similitudinii diferitelor caracteristici dar nu se bazează pe definirea apriori a grupurilor. Este o sarcină importantă deoarece ajută logopezii să înțeleagă cine le sunt pacienții.
- *regulile de asociere* urmăresc detecția conexiunilor între date, pe baza asocierilor sau a descoperirii secvențelor

Din cele menționate se poate desprinde ideea că data mining poate fi un instrument folositor. Dar trebuie menționată și următoarea limitare a acesteia:

- *data mining analizează numai datele colectate de la logopații care au ieșit din programul de terapie.* Aplicațiile de data mining generează informații prin analiza tiparelor de date obținute din sistemul de programe aferente asistării și urmăririi terapiei logopedice. Tiparele astfel detectate pot ajuta la predicția evoluției logopaților aflați în mod curent în proces de terapie sau la proiectarea unei scheme de terapie adecvată pentru aceștia. Cu toate acestea tehnologia data mining nu poate furniza informații despre deficiente, persoane sau comportamente care nu se regăsesc în bazele care furnizează datele de analizat.

V. STUDIU DE CAZ

Ideea încercării de a îmbunătăți calitatea terapiei logopedice prin aplicarea unor tehnici de data mining a avut la bază proiectul TERAPERS dezvoltat în cadrul Centrului de cercetare în Calculatoare din Universitatea “Ștefan cel Mare” din Suceava. Acest proiect și-a propus realizarea unui sistem care, printre altele, să asiste profesorii logopezi în activitatea de terapie a dislaliei și să urmărească modul în care logopații răspund diferitelor programe de terapie personalizată.

Figura 2 prezintă fluxul principalelor activități derulate cu sprijinul sistemului TERAPERS. Toate aceste activități se concretizează într-un volum consistent de date stocate într-o bază de date relațională.

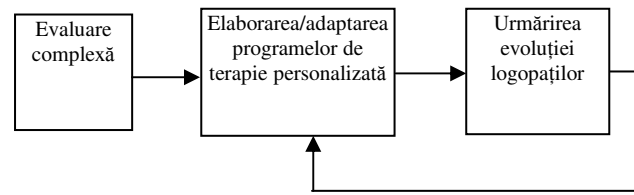


Fig. 2 Sarcinile sistemului TERAPERS.

Adaptarea programelor de terapie presupune o examinare complexă a copiilor, precum și înregistrarea unor date cu caracter relevant referitoare la anamneza personală și familială.

Examinarea complexă a modului în care copii articulează fonemele în diferite construcții permite stabilirea unui diagnostic și încadrarea într-o anumită categorie de gravitate. Datele de anamneză colectate pot furniza informații interesante relativ la diferitele cauze care pot influența negativ dezvoltarea normală a limbajului.

Programelor de terapie personalizată aplicate furnizează date precum număr de ședințe/săptămână, exercițiile aferente fiecărei faze de terapie, modificarea programului inițial conform evoluție logopatului. În plus, raportul descărcat de pe dispozitivul mobil colectează date referitoare la efortul depus de copil în activitatea independentă: exercițiile făcute, numărul de repetări pentru fiecare și rezultatele obținute.

Acțiunea de urmărire a evoluției copilului se materializează în date care indică momentul evaluării și starea copilului în momentul respectiv.

Pornind de la analiza datelor disponibile în baza de date se pot face următoarele observații.

Se poate utiliza clasificarea care pornește de la informația conținută în mulțimea variabilelor predictor, cum ar fi date de anamneză (personală sau familială) sau referitoare la stilul de viață, pentru a asocia logopații cu diferitele segmente.

Clustering-ul poate descoperi un subgrup dintr-un segment predeterminat, cu comportament omogen față de diferitele metode de terapie care poate fi țintit eficient printr-o terapie specifică.

O sarcină importantă a asocierii ar putea fi angajată într-un efort de a determina de ce un anumit program de terapie specifică a avut succes pe un segment de logopați, iar pe altul a fost ineficientă.

VI. CONCLUZII ȘI DEZVOLTĂRI VIITOARE

Tehnologia data mining poate fi un instrument util pentru domeniul terapiei logopedice deoarece este în măsură să furnizeze informația care permite implementarea unor programe de terapie personalizate și optimizate, adaptate caracteristicilor fiecărui copil. Acest lucru conduce la scăderea duratei terapiei, la creșterea posibilităților de a obține rezultate superioare iar în final la scăderea costurilor terapiei.

REFERINȚE BIBLIOGRAFICE

- [1] Danubianu M. (2003) Determinarea metodei optime de explorare a datelor pentru un sistem inteligent de evaluare a activității în turism, volumul „Tehnologii Informaționale”, Suceava, pag.156-163, ISBN 793-666-059-1.
- [2] Danubianu M.(2005) Determinarea claselor pe baza tehnicii SBA, volumul seminarului “Procesare distribuită”, pag. 86-90, ISBN 973-666-177-6.
- [3] Fleisher, C. S., & Blenkhom, D. (2003) Controversies in competitive intelligence: The enduring issues. Westport, CT: Praeger.
- [4] Danubianu M. (2006) Using data mining techniques for decision support systems –Proceedings of the International Conference on Signal/Image Processing and Pattern Recognition -"UkrObraz-2006",pag. 19-22, August 28-31, 2006, Kiev, Ucraina.
- [5] Pentiuc., Ghe-St, MORARIU, N. Morariu, M, Pentiuc. L (2001) Intelligent System For Impact Prognosis Of The Economic Decisions At District Level, Advances in Electrical and Computer Engineering ISSN 1582-7445 - Volume 1(18), Number 1(15), 2001.
- [6] Zhao, L. J., & Zhu, D. (2003). Workflow resource selection from UDDI repositories with mobile agents. Proceedings of Web2003, USA.
- [7] Zhu, D., Premkumar, G., Zhang, X., & Chu, C. (2001). Data mining for network intrusion detection: A comparison of alternative methods. Decision Sciences, 32(4), 635-660.
- [8] Wirth, R. and Hipp, (2000) J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pages 29-39, Manchester, UK.
- [9] Danubianu M. , Pentiuc S.G., Schipor O., Nestor M. , Ungurean I., Distributed Intelligent System for Personalized Therapy of Speech Disorders, Proceedings of ICCGI09, Atena.