RELEVANŢA MODELELOR GRAFICE ŞI REŢELELOR BAYESIENE

Inga LISNIC[1*],
Sergiu SCROB[1]

[1]*Universitatea Tehnică a Moldovei, Facultatea Calaculatoare Informatică şi Microelectronică, Departamentul Ingineria Software şi Automatică, Doctorandă, Chişinău, Republica Moldova*

*Autorul corespondent: Inga Lisnic, inga.lisnic@ati.utm.md

Rezumat. *Modelele grafice reprezintă o strînsă legătură între teoria probabilității și teoria grafurilor. Ele oferă un instrument natural pentru a face față a două probleme care apar pe parcursul matematicii și ingineriei aplicate - incertitudinea și complexitatea - și în special joacă un rol din ce în ce mai important în proiectare și analiza algoritmilor de învățare a mașinilor. Fundamentale pentru ideea unui model grafic este noțiunea de modularitate - un sistem complex este construit prin combinarea unor părți mai simple.Teoria probabilității furnizează lipiciul prin care piesele sunt combinate, asigurând ca sistemul în general este consecvent și oferă modalități de interfațare a modelelor la date. Latura teoretică a modelelor grafice oferă atât o interfață atractivă intuitiv prin care oamenii pot modela seturi de variabile care interacționează puternic, cât și o structură de date care se acordă în mod natural designului de algoritmi eficienți cu scop general. În cadrul acestui articol, vom discuta doar despre modele grafice direcționate, adică rețelele Bayesiene.*

Cuvinte cheie*: Rețele Bayesiene, nod, graf, probabilitate, Bayes,*

Introducere

În lumea înconjurătoare, fenomenele deterministe ocupă doar o mică parte. Imensa majoritate a fenomenelor din natură și societate sunt stocastice (aleatoare). Studiul acestora nu poate fi făcut pe cale deterministă și, de aceea, știința hazardului a apărut ca o necesitate. **Teoria probabilităților** studiază legile după care evoluează fenomenele aleatoare [1]. Aplicarea matematicii la studierea fenomenelor aleatoare se bazează pe faptul că, prin repetarea de mai multe ori a unui experiment, în condiții practic identice, frecvența relativă a apariției unui anumit rezultat (raportul dintre numărul experimentelor în care apare rezultatul și numărul tuturor experimentelor efectuate) este aproximativ același, oscilând în jurul unui număr constant. Dacă acest lucru se întâmplă, atunci unui eveniment dat îi putem asocia un număr, anume probabilitatea sa. Această legătură între structura unui câmp de evenimente și număr este o reflectare în matematică a transferului calității în cantitate. Problema convertirii în număr a unui câmp de evenimente revine de a defini o funcție numerică pe această structură, care să fie o măsură a posibilităților de realizare a evenimentelor. Realizarea unui eveniment fiind probabilă, această funcție se numește probabilitate.

Formula probabilităţii totale. Formula lui Bayes

Teoremă. Dacă A şi H_1, H_2, ..., H_n,... sunt evenimente aleatoare legate de acelaşi câmp de probabilitate (Ω,F, P) şi satisfac condiţiile:[1]

a) evenimentul A implică producerea cel puţin a unuia din evenimentele şi H_1, H_2, .., H_n;
b) evenimentele şi H_1, H_2, ..., H_n sunt incompatibile două câte două;
c) $P(H_i)>0$,
atunci au loc ***formula probabilităţii totale*** [4]

$$P(A) = P(H_1)P(A|H_1) + \cdots + P(H_n)P(A|H_n) \quad (1)$$

şi ***formula lui Bayes*** [4]

$$P(H_J|A) = \frac{P(H_J)P(A|H_J)}{P(H_1)P(A|H_1)+\cdots+P(H_n)P(A|H_n)+\cdots} \quad (2)$$

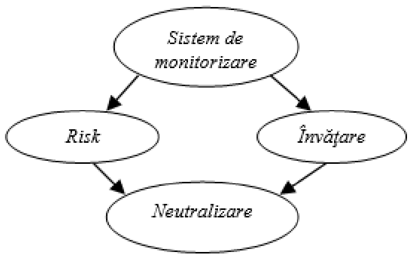
Rețelele Bayesiene

Rețelele Bayesiene reprezintă *modele grafice* [6] pentru evidențierea relațiilor probabilistice ale unor variabile în medii incerte. Astfel că fiecare nod al grafului este constituit de o variabilă aleatoare, în timp ce arcele dintre noduri constituie dependențe probabilistice dintre variabilele aleatoare corespunzătoare. Aceste dependențe condiționale din cadrul grafului sunt deseori estimate prin folosirea unor metode computaționale de natură statistică. Rețelele Bayesiene corespund unei structuri grafice diferite, denumită și *graf aciclic direcționat* și sunt deseori utilizate în domenii precum statistica, inteligența artificială sau învățarea automată. Această popularitate se datorează atât fundamentării matematice solide cât și posibilității de înțelegere intuitivă a acestora. Rețelele Bayesiene obțin o performanță ridicată în cazul în care seturile de date sunt incomplete, eliminând problema predicțiilor inconsistente datorate lipsei corelațiilor dintre variabilele de intrare. De asemenea, acestea permit învățarea relațiilor cauzale care ne sunt utile în momentul în care dorim să înțelegem domeniul problemei, ajutându-ne de asemenea și la obținerea predicțiilor în condiții instabile.

Cu ajutorul tehnicilor de statistică Bayesiene, rețelele Bayesiene contribuie la formarea unui domeniu vast de cunoștințe și informații necesare oricărei analize de natură probabilistică, iar cu ajutorul celorlalte metode Bayesiene și a altor modele ne oferă o abordare eficientă pentru evitarea aglomerării datelor. [8] Pentru a putea scoate în evidență caracteristicile rețelelor Bayesiene, prezentăm figura de jos.

P(SM=T)	P(SM=F)
0.8	0.2

C	P(Risk=T|SM)	P(Risk=F|SM)
T	0.9	0.1
F	0.01	0.99



P(Învățare=T)	P(Învățare=F)
0.02	0.98

SM	Risk	P(Neutralizare=T|SM,Învățare)	P(Neutralizare=F|SM,Învățare)
T	F	0.9	0.1
T	F	0.2	0.8
T	F	0.9	0.1
T	F	0.01	0.99

Figura 1.Reprezentarea rețelei bayesiene

Ne este prezentată o structură care scoate în evidență posibilitatea că un anumit risk exprimă probabilitatea unor întâmplări negative și impactul acestora, acest eveniment fiind reprezentat de variabila *Neutralizare.* Acțiunea respectivă poate fi datorată unui algoritm reprezentat de variabila *Învățare*, precum și unor surse reprezentate de variabila *Sistem de monitorizare.* În ultimul rând, pentru ca acest scenariu să fie relevant, este necesară prezența unui risk care să sufere de această acțiune, evenimentul fiind reprezentat de variabila *Risk.* Toate aceste variabile sunt binare, adică pot fi ori adevărate, ori false.

Conform conceptului de independență abordat de rețelele Bayesiene, în cadrul figurii de mai sus pot fi observate mai multe astfel de stări. De exemplu, variabilele *Sistem de monitorizare* și *Învățare* sunt *marginal independente*, însă dacă luăm în considerare și variabila *Neutralizare*, atunci acestea vor deveni *condițional dependente.* Acest principiu conferă o factorizare compactă a distribuției comune a probabilităților, iar numărul de parametrii ai modelului este redus semnificativ, datorită distribuției multinomiale în acest caz, de la $2^4 - 1 = 31$ la *10.*

Deducția în rețelele Bayesiene

Rețelele Bayesiene pot fi utilizate în deducția valorilor pentru anumite variabile țintă, în funcție de valorile observate ale altor variabile. Având în vedere faptul că se lucrează cu variabile aleatoare, de cele mai multe ori, asignarea unei singure valori ca rezultat nu va fi corectă. Scopul nostru este acela de a identifica distribuția probabilităților pentru funcțiile țintă și poate fi realizat cu ușurință, în cazul în care cunoaștem toate valorile variabilelor rămase în cadrul rețelei. Vizualizând această problemă într-o manieră mai generală, putem spune că dorim să găsim distribuția probabilităților pentru anumite variabile, luând în considerare doar un subset din variabilele rămase.

De-a lungul timpului, au fost folosite mai multe modalități de inferență, două dintre cele mai cunoscute fiind următoarele [4]:

- *suport predictiv* pentru nodul X_i, bazat pe evidența nodurilor conectate la acesta, prin intermediul nodurilor părinte (se mai numește și inferență de jos în sus).
- *suport diagnostic* pentru nodul X_i, bazat pe evidența nodurilor conectate la acesta prin intermediul nodurilor copii (se mai numește și inferență de sus în jos).

Cea mai simplă relație de independență condițională codificată într-o rețea Bayesiană poate fi declarată astfel: un nod este independent de strămoșii săi, dat fiind părinții săi, unde relația strămoș / părinte este în raport cu o anumită ordonare topologică fixă a nodurilor. Prin regula lanțului de probabilitate, probabilitatea comună a tuturor nodurilor din graficul de mai sus este:

$$P = \left(SM, R, În, Ne\right) = P(SM) * P(R|SM) * P\left(În\middle|SM, R\right) * P(Ne|SM, R, În) \quad (3)$$

Folosind relații de independență condiționată, putem rescrie acest lucru astfel:

$$P = \left(SM, R, În, Ne\right) = P(SM) * P(R|SM) * P\left(În\middle|SM\right) * P(Ne|, R, În) \quad (4)$$

$$\Pr\ (\mathrm{R} = 1|Ne = 1) = \frac{\Pr(R=1, Ne=1)}{\Pr(Ne=1)} = \sum_{n,p} \frac{\Pr\left(SM=sm, R=r, \mathrm{În}=\mathrm{în}, Ne=1\right)}{\Pr(\mathrm{Ne}=1)} \quad (5)$$

$$\Pr(\mathrm{În} = 1|Ne = 1) = \frac{\Pr\left(\mathrm{În}=1, Ne=1\right)}{\Pr(Ne=1)} = \sum_{p,s} \frac{\Pr\left(Sm=sm, R=r, \mathrm{În}=1, Ne=1\right)}{\Pr(\mathrm{Ne}=1)} \quad (6)$$

unde:

$$\Pr(Ne = 1) \sum_{\mathrm{sm,în,r}} \Pr\left(Sm = sm, R = r, \mathrm{Î}n = 1, Ne = 1\right) \quad (7)$$

Concluzii

În articolul de față am încercat să scoatem în evidență posibilitatea integrării relevante a modelelor grafice şi reţele Bayesiene în viața noastră de zi cu zi, prin simplificarea proceselor de luare a unor decizii în raport cu implicarea sistemelor inteligente de natură artificială. Reţelele bayesiane sunt cunoscute si sub numele de retele probabilistice cauza-le, retele de probabilitate grafica, modele probabilistice cauza-efect si diagrame de influenta probabilistica, fiind de actualitate datorită folosirii lor ca o posibilă solutie pentru situatiile în care informatia avută la dispozitie este incompletă.

Referinţe

1. ION BALMUŞ, GHEORGHE CEBAN, ALEXEI LEAHU, INGA LISNIC, A. MOLOŞNIUC. *Teoria Probabilităţilor şi a Informaţiei în Sistemul de programe Mathematica (Teorie, indicatii metodice si probleme propuse),* UTM, Chişinău 2017
2. C. LYNNES, S. BERRICK, A. GOPALAN. *Application of Bayesian Classification to Content-based Data Management,* 2004
3. BEN-GAL I. *Bayesian Networks*, *Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons,* (2007)
4. ION BALMUŞ, GHEORGHE CEBAN, ALEXEI LEAHU, INGA LISNIC, A. MOLOŞNIUC. *Teoria Probabilităţilor şi a Informaţiei în Sistemul de programe Mathematica (Teorie, indicatii metodice si probleme propuse),* UTM, Chişinău 2017, pp. 31-42.
5. N. SEBE, M.S. LEW, I. COHEN, A. GARG, T.S. HUANG. Emotion Recognition Using a Cauchy Naive Bayes Classifier. In: *Proceedings of the 16 th InternationalConference on Pattern Recognition* (ICPR'02), (1) 2002, 10017
6. BEN-GAL I. Bayesian Networks. In: *Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons,* (2007)
7. JOANNA KAZMIERSKAA, JULIAN MALICKI. Application of the Naive Bayesian Classifier to optimize treatment decisions. In: *Radiotherapy and Oncology*, 86 (2008), 211–216
8. Understanding the Bias-Variance Tradeoff. [online], [accesat 19.02.2020]. Disponibil: http://scott.fortmann-roe.com/docs/BiasVariance.html