

Rețele bayesiene

Sinteză

Teorema lui Bayes

$$P(B|A) = P(A|B) \cdot P(B) / P(A)$$

$P(B|A)$ este probabilitatea a posteriori (engl. “posterior”)

$P(A|B)$ este verosimilitatea (engl. “likelihood”)

$P(B)$ este probabilitatea a priori (engl. “prior”)

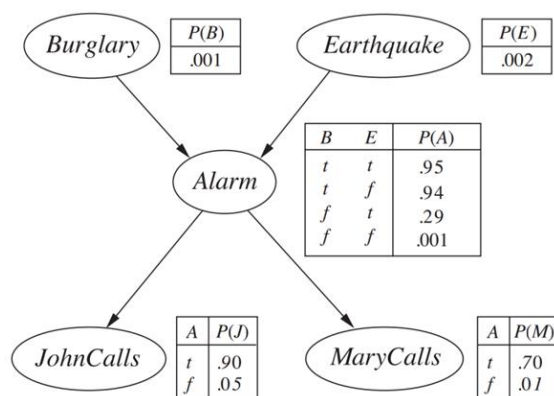
$P(A)$ este evidența (engl. “evidence”)

Relația este importantă deoarece putem calcula astfel probabilitățile cauzelor, date fiind efectele. Este mai simplu de cunoscut când o cauză determină un efect, dar invers, când cunoaștem un efect, probabilitățile cauzelor nu pot fi cunoscute imediat. Teorema ne ajută să diagnosticăm o anumită situație sau să testăm o ipoteză.

Structura unei rețele bayesiene

O *rețea bayesiană* este un model grafic probabilistic, adică un graf cu o mulțime de noduri, care reprezintă evenimente aleatorii, conectate de arce, care reprezintă dependențe condiționate între evenimente.

De exemplu:



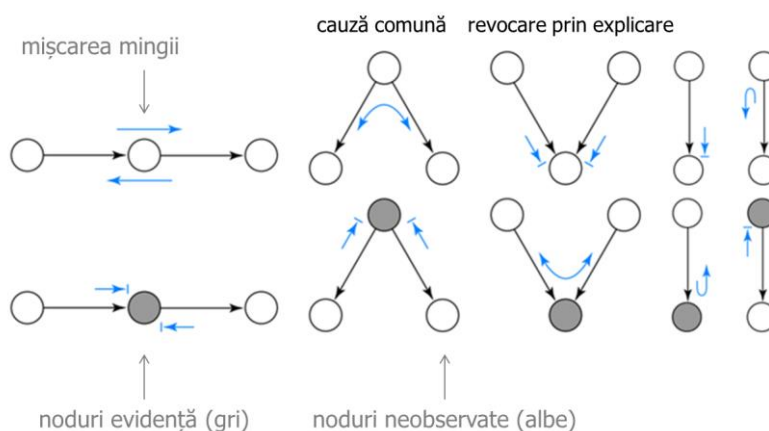
Ideea de bază este că un nod depinde doar de „părinții” săi, adică de o submulțime a nodurilor din rețea, nu de toate celelalte noduri, ceea ce ar implica folosirea distribuției comune de probabilitate. Altfel spus, fiecare nod este independent condiționat de predecesorii din șirul ordonat al nodurilor, dați fiind părinții nodului.

O rețea bayesiană este un graf orientat aciclic. Arcele pot forma bucle, dar nu pot forma cicluri.

Algoritmul Bayes-Ball

Reprezintă o modalitate de a determina relațiile de *independență* și *independență condiționată* într-o rețea bayesiană. Se presupune că o minge este trimisă dintr-un nod în rețea. Mingea trece în moduri diferite, în funcție de cine o trimite (fiu sau părinte) și starea nodului care o primește (observat/evidență sau neobservat). Nodurile la care mingea nu ajunge sunt independente (condiționat) de nodul de start.

Regulile de trimitere a mingii sunt cele din figura de mai jos:



Rețelele bayesiene codează relații de *corelație*. Dacă probabilitatea unui eveniment A variază împreună cu probabilitățile altor evenimente $\{B, C, \dots\}$, atunci A este corelat cu $\{B, C, \dots\}$. A poate fi cauzat sau nu de $\{B, C, \dots\}$. Toate evenimentele $\{A, B, C, \dots\}$ ar putea fi cauzate de alte evenimente necunoscute. O relație de corelație nu implică o relație de cauzalitate. O relație de cauzalitate implică o relație de corelație.

Inferențe exacte și aproximative

Sortarea topologică a unui graf este o ordonare liniară a nodurilor sale astfel încât, pentru fiecare arc $A \rightarrow B$, A apare înaintea lui B . Pentru o rețea bayesiană, sortarea topologică asigură faptul că nodurile părinte vor apărea înaintea nodurilor fiu, ceea ce este necesar pentru unii algoritmi sau pot să simplifice calculele.

Inferența probabilităților marginale calculează probabilitățile nodurilor în sine, în lipsa unor noduri evidență. Pentru un nod, se sumează probabilitățile sale condiționate de toate combinațiile de valori ale părinților, înmulțite cu probabilitățile părinților de a avea valorile respective.

Inferența prin enumerare este un algoritm prin care se poate răspunde la întrebări arbitrare despre evenimentele codate de o rețea bayesiană. Unele variabile (noduri) sunt observate, iar altele nu. Se calculează probabilitățile variabile de interogare separat pentru fiecare valoare. Pentru variabilele observate, se folosește valoarea observată. Pentru variabilele neobservate, se sumează

probabilitățile pentru toate valorile posibile. În final, probabilitățile determinate se normalizează, pentru a avea suma 1.

Inferența prin ponderarea verosimilității este un algoritm de inferență aproximativă stohastică. Se generează aleatoriu eșantionări (instanțieri) ale rețelei și se calculează probabilitățile dorite ca frecvențe relative de apariție. Nodurile fără părinți sunt instanțiate potrivit probabilităților lor marginale. Nodurile evidență iau mereu valorile observate. Valorile variabilelor neobservate au probabilități de apariție în conformitate cu probabilitățile nodurilor. Pentru fiecare eșantionare a rețelei, se calculează o pondere. Se repetă procesul pentru un număr prestabilit de eșantioane. În final, are loc o fază de normalizare, în care se calculează suma ponderilor cazurilor în care o variabilă de interogare a avut o anumită valoare, împărțită la suma ponderilor tuturor cazurilor.

Rețele bayesiane dinamice

O *rețea bayesiană dinamică* este o rețea bayesiană care reprezintă un model de probabilitate temporal. Este organizată în partiții (engl. “slices”) de timp. Fiecare partiție poate avea mai multe variabile de stare \mathbf{X}_t și variabile de evidență \mathbf{E}_t . Pentru a construi o RBD, trebuie specificate: distribuția a priori a variabilelor de stare $P(\mathbf{X}_0)$, modelul de tranziții $P(\mathbf{X}_{t+1} | \mathbf{X}_t)$ și modelul de observații sau de senzori $P(\mathbf{E}_t | \mathbf{X}_t)$.

O RBD se poate transforma într-o RB simplă prin desfășurare (engl. “unrolling”), după care s-ar putea aplica algoritmi de inferență de la RB simple. Însă complexitatea algoritmilor de inferență exactă specifici RB simple pentru RBD desfășurate este exponențială.

Filtrarea cu particule (engl. “particle filtering”)

Este un algoritm de inferență aproximativă pentru RBD.

Filtrare înseamnă determinarea lui $P(\mathbf{X}_t | \mathbf{E}_{0:t})$ la momentul curent t .

Mai întâi, se creează o populație de n particule cu starea eșantionată din distribuția a priori $P(\mathbf{X}_0)$. Apoi, se repetă următoarele faze pentru fiecare moment de timp:

1. Fiecare eșantion (stare a particulei) este propagat înainte prin eșantionarea următoarei stări pe baza modelului de tranziție $P(\mathbf{X}_{t+1} | \mathbf{X}_t)$;
2. Fiecare particulă este ponderată de probabilitatea pe care o atribuie noilor evidențe $P(\mathbf{E}_{t+1} | \mathbf{X}_{t+1})$;
3. Populația este reeșantionată pentru a genera o nouă populație de n particule. Fiecare nouă particulă este selectată din populația curentă. Probabilitatea de a fi selectată o particulă este proporțională cu ponderea sa. Noile particule nu au ponderi.