

Construirea unei rețele Bayesiene pentru modelarea unui anumit domeniu

Daniela Palcu
Gabriel Matasariu
Grupa 30238

April 1, 2018

Abstract

Vă vom prezenta construirea unei rețele Bayesiene pentru modelarea domeniului criminalității din Statele Unite ale Americii în anul 2016. Acest raport include 7 secțiuni care descriu problema propusă, construirea rețelei Bayesiene și pașii efectuați în modelarea problemei folosind această rețea.

1 Specificația problemei

Problema propusă în acest articol este aceea de a găsi anumite legături sau statistici în ceea ce privește arestările ce au avut loc în anul 2016 în Statele Unite ale Americii. Aceste legături se pot referi la:

- tipul de crimă comisă(jaf,omor,furt,etc.) și numărul lor
- rasa celui care a comis crima(white,blak,asian,indian,hawaiian,etc) și numărul lor

Datele folosite în încercarea rezolvării problemei sunt datele puse la dispoziție pe site-ul FBI cu privire la rata criminalității din anul 2016 în Statele Unite ale Americii precum și alte date referitoare la populația acestei țări preluate de pe Wikipedia.

2 Variabilele și domeniul lor

Pentru modelarea problemei propuse au fost introduse următoarele variabile care pot fi împărțite în 2 categorii:

- Rase
- Tipuri de crimă

Din categoria raselor, cele care fac parte din rețeaua noastră sunt următoarele:

- White
- Black or African
- Indian
- Asian
- Hawaiian

Din categoria tipurilor de crimă, cele care fac parte din rețeaua noastră sunt:

- Violent Crime
- Property Crime

Deoarece am dorit ca domeniul criminalității să fie mai extins, categoriile de crime au fost împartite la rândul lor în mai multe tipuri pentru a putea ilustra pe urma legăturile dintre acestea:

Violent Crime

- Rape
- Robbery
- Murder
- Aggravated Assault

Property Crime

- Burglary
- Larceny theft
- Motor Vehicle Theft

Aceste variabile au fost extrase din datele pe care le detinem, iar denumirile lor au fost pastrate în limba engleză pentru a nu pierde informația transmisă de acestea prin traducerea lor.

3 Rețeaua Bayesiană propusă

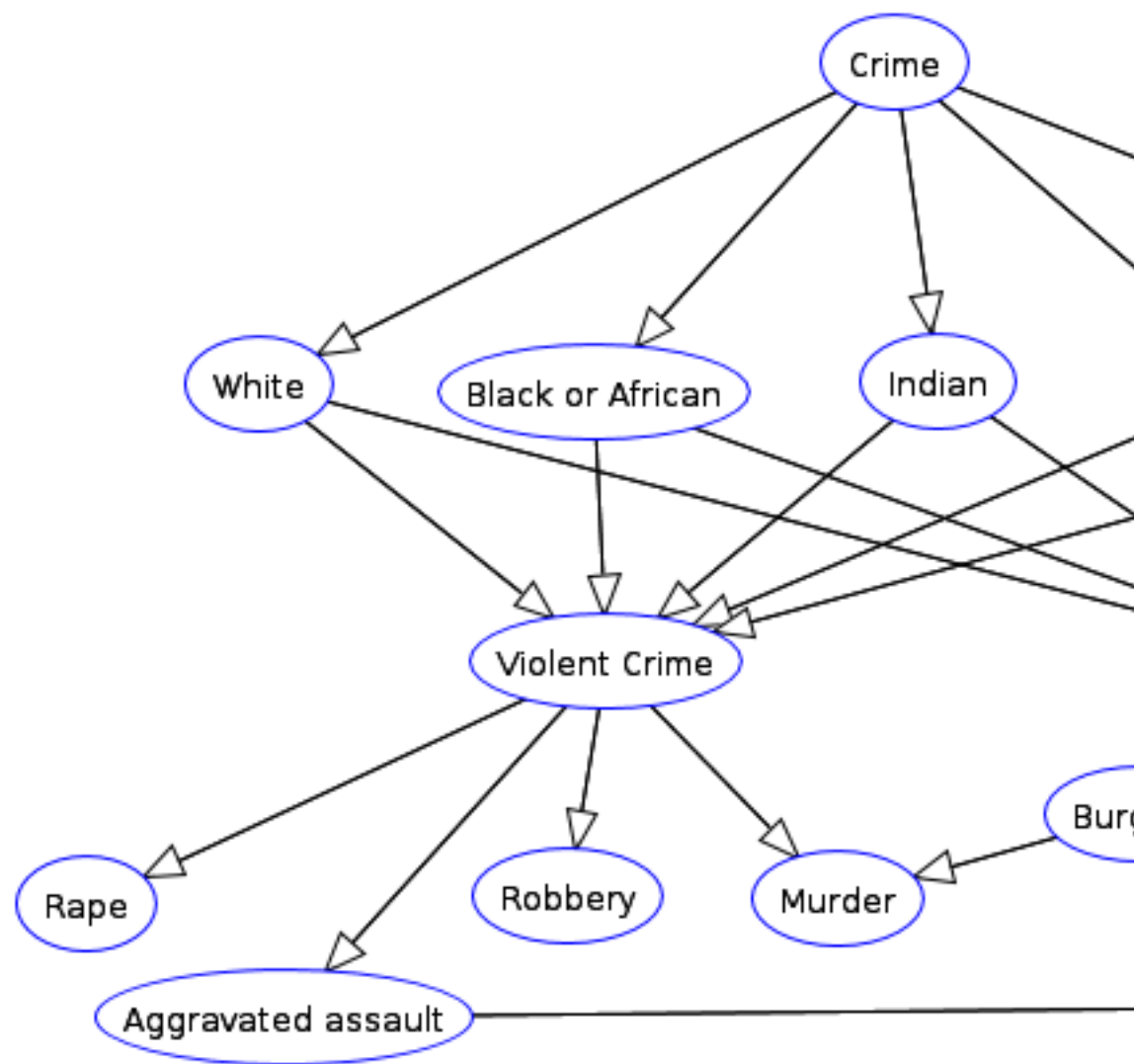
4 Descrierea rețelei Bayesiane(Justificare)

În această secțiune vor fi justificate prezenta sau absenta arcelor între nodurile rețelei Bayesiane.

- Crime – Rasa - prezenta arcelor poate fi justificată prin faptul că atunci când a fost comisă o crimă, atunci ea a fost comisă cu o anumită probabilitate de una din rasele ce se găsesc în nodurile de pe ”primul nivel” al rețelei
- Rasa – Rasa - absenta arcelor între diferite rase se justifică prin faptul că ilegalitățile comise de o anumită rasă sunt independente de ilegalitățile comise de orice altă rasă(desigur teoretic ar putea exista dependente în sensul că dacă un alb omorâ un negru, un alt negru va omorâ o persoană albă)
- Rasa – ViolentCrime/PropertyCrime - aceste arcuri se justifică prin faptul că din tabelele utilizate reiese că dacă o anumită persoană, de o anumită rasă, a comis o infracțiune, atunci aceasta poate fi încadrată fie la ViolentCrime fie la PropertyCrime
- ViolentCrime - PropertyCrime - desigur acest arc ar putea exista(pentru că anumite PropertyCrimes se soldează și cu un ViolentCrime), am ales ca aceste legături între diferite tipuri de crimă să le exprimăm în ultimul nivel al rețelei,cel în care vor fi reprezentate explicit tipurile de ilegalități
- ViolentCrime/PropertyCrime – CrimeType - nodurile se justifică prin faptul că în tabelele date ViolentCrime și PropertyCrime sunt împărțite în mai multe clase,ceea ce înseamnă că în rețeaua Bayesiană ca un tip explicit de ilegalitate reprezintă un anumit procent din toate ViolentCrimes/PropertyCrimes comise
- CrimeType – CrimeType - aceste arcuri se justifică prin faptul că, o anumită ilegalitate,poate conduce în anumite condiții la o altă ilegalitate (de exemplu, un jaf armat se poate solda și cu o crimă,etc)

5 Obținerea probabilităților

Majoritatea probabilităților care au fost obținute sunt bazate pe date reale, iar celelalte au fost obținute prin diferite metode de calcul care sunt prezen-



tate mai in jos. Pentru fiecare variabila in parte, va fi prezentat tabelul in care sunt incluse probabilitatiile acesteia si modul in care acestea au fost obtinute.

Din datele obtinute de la FBI si din datele de pe Wikipedia s-a constatat ca probabilitatea ca un om sa comita o ilegalitate in Statele Unite ale Americii in anul 2016 este de 70%.

Table 1: Crime

P(Crime = T)	P(Crime = F)
0.7	0.3

Probabilitatea ca un om de o anumita rasa sa comita o ilegalitate a fost calculata folosind formula:

$$\frac{\text{Numarul de persoane care au o anumita rasa si produc o ilegalitate}}{\text{Numarul total de persoane care produc o ilegalitate}} \quad (1)$$

Aceasta formula folosita se aplica tuturor celor 5 variabile care reprezinta rasa. Probabilitatea complementara de obtine folosind formula: $P(a) = 1 - P(a)$.

Probabilitatea ca un om de o anumita rasa sa nu comita o ilegalitate este calculata folosind formula:

$T1 = \text{Numarul total de persoane care au o anumita rasa}$

$T2 = \text{Numarul de persoane din rasa respectiva care produc o ilegalitate}$

$T = \text{Numarul total de oameni din SUA}$

$$\frac{T1 - T2}{T} \quad (2)$$

Probabilitatea complementara se obtine in mod similar cu cea prezentata anterior.

In tabelele II - VI sunt ilustrate aceste probabilitati care au fost obtinute pe baza formulelor de mai in sus.

Pentru calcularea probabilitatilor ca un anumit tip de ilegalitate sa aiba loc si sa fie produsa de o persoana de o anumita rasa, au fost obtinute din tabelele cu datele pe care le avem din resursele de la FBI. Dar se poate merge si pe o abordare intuitiva in sensul in care din totalul oamenilor de o anumita rasa care au comis o crima putem sa vedem cati produc o anumita categorie de crima. Aceste

Table 2: White

Crime	P(White = T)	P(White = F)
T	0.69	0.31
F	0.71	0.29

Table 3: Black or African

Crime	P(Black or African = T)	P(Black or African = F)
T	0.26	0.74
F	0.11	0.89

categorii sunt: Violent Crime si Property Crime. Dupa cum se poate observa din tabelele VII si VIII au fost pastrate doar acele variabile in care pe un rand doar una singura este adevarata, deoarece am pornit de premisa ca nu exista relatii de dependenta intre variabilele de pe nivelul 1.

Probabilitatiile pentru care un anumit tip de ilegalitate sa aiba loc au fost obtinute tot din datele din resursele pe care le avem. Dar, si aici putem sa avem o abordare intuitiva in sensul in care probabilitatea ca un anumit tip de ilegalitate sa aiba loc poate fi calculata impartind numarul total de ilegalitati de un anumit tip care au loc, la numarul total de ilegalitati din categoria din care face parte ilegalitatea respectiva.

De asemenea, se poate observa ca unele ilegalitati din categorii diferite se influenteaza reciproc. De exemplu, in cazul nostru Motor Vehicle Theft este infleuntat de catre Aggravated Assault si Murder de catre Burglary. Aceste probabilitati au fost approximate, intuitia in acest caz fiind aceea ca din numarul total de Burglary sau Aggravated Assault, cam in 10-20% din aceste cazuri au loc si celelalte tipuri de ilegalitati. De aici, reiesind probabilitatiile care sunt prezentate in tabelele XII si XII.

6 Interogări și răspunsuri ale rețelei

Interogarile la care reseaua dezvoltata ar trebui sa fie capabila sa raspunda sunt de tipul celor care urmeaza:

- O crima a fost comisa. Care este probabilitatea ca aceasta sa fi fost comisa de o persoana asiatica?

Table 4: Indian

Crime	P(Indian = T)	P(Indian = F)
T	0.02	0.98
F	0.007	0.993

Table 5: Asian

Crime	P(Asian = T)	P(Asian = F)
T	0.012	0.988
F	0.051	0.949

- O crima a fost comisa. Care este probabilitatea ca aceasta sa fi fost un jaf comis de o persoana asiatica?
- O crima a fost comisa. Care este probabilitatea ca aceasta sa fi fost un jaf comis de o persoana asiatica, si care s-a soldat si cu o crima?
- A fost comis un jaf.Care este probabilitatea ca acesta sa fi fost comis de o persoana neagra?
- A fost comisa o crima.Care este probabilitatea sa fi avut loc un jaf si sa fie comis de catre o persoana de rasa neagra sau africana?
Raspuns : $P = 0.03$
- A fost comisa o crima.Care este probabilitatea sa se fi produs si un jaf in timpul crimei?
Raspuns : $P = 0.07$
- A avut loc un furt de motocicletă. Care este probabilitatea sa se fi produs un asalt(Aggravated assault)?
Raspuns : $P = 0.08$

Pentru calcularea probabilitatilor s-a folosit formula lui Bayes:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

Table 6: Hawaiian

Crime	P(Hawaiian = T)	P(Hawaiian = F)
T	0.003	0.997
F	0.001	0.999

Table 7: Violent Crime

White	Black or African	Indian	Asian	Hawaiian	P(Violent Crime = T)	P(Violent Crim
T	F	F	F	F	0.59	0.41
F	T	F	F	F	0.375	0.625
F	F	T	F	F	0.018	0.982
F	F	F	T	F	0.014	0.986
F	F	F	F	T	0.013	0.987

Mai in jos vom ilustra cum este aplicata aceasta formula pentru ultima interogare, dar ideea este aceeaasi si pentru celelalte interogari. Avem urmatoarele variabile ale formulei:

AS = Agravated Assault

MV = Motor vehicle theft

$$P(AS | MV) = \frac{P(MV | AS) P(AS)}{P(MV)}$$

unde

$$P(AS | MV) = 0.0073 \text{ (din tabelul 15)}$$

$$P(AS) = 0.74 \text{ (din tabelul 11)}$$

$$P(MV) = 0.063 \text{ (din tabelul 15)}$$

Daca aplicam formula, vom obtine rezultatul 0.08. Acelasi rezultat pe care il obtinem si in urma interogarii retelei.

7 Concluzii

Una din limitările rețelei dezvoltate este aceea ca fiecare tip de crima nu este conectata in mod direct cu fiecare tip de rasa. Daca acest lucru s-ar fi intamplat, ca rezultatele obtinute ar fi fost mai exacte si in acelasi timp mai intuitive. O posibila imbunatatire evidenta a rețelei este reverificarea calculului probabilitatilor si verificarea corectitudinii datelor ce au deservit drept input pentru retea.

Table 8: Property Crime

White	Black or African	Indian	Asian	Hawaiian	P(Property Crime = T)	P(Property Crime = F)
T	F	F	F	F	0.68	0.32
F	T	F	F	F	0.28	0.72
F	F	T	F	F	0.017	0.983
F	F	F	T	F	0.012	0.988
F	F	F	F	T	0.003	0.997

Table 9: Rape

Violent Crime	P(Rape = T)	P(Rape = F)
T	0.045	0.955
F	0.0	1.0

Table 10: Robbery

Violent Crime	P(Robbery = T)	P(Robbery = F)
T	0.176	0.824
F	0.0	1.0

Table 11: Aggravated Assault

Violent Crime	P(Aggravated Assault = T)	P(Aggravated Assault = F)
T	0.74	0.26
F	0.0	1.0

Table 12: Murder

Violent Crime	Burglary	P(Murder = T)	P(Murder = F)
T	T	0.015	0.985
T	F	0.022	0.978
F	T	0.0	1.0
F	F	0.0	1.0

Table 13: Burglary

Property Crime	P(Burglary = T)	P(Burglary = F)
T	0.153	0.847
F	0.0	1.0

Table 14: Lacerny Theft

Property Crime	P(Lacerny Theft = T)	P(Lacerny Theft = F)
T	0.77	0.23
F	0.0	1.0

Table 15: Motor Vehicle Theft

Property Crime	Aggravated Assault	P(Motor Vehicle Theft = T)	P(Motor Vehicle Theft = F)
T	T	0.0073	0.9927
T	F	0.063	0.937
F	T	0.0	1.0
F	F	0.0	1.0