

Raport

Alesia Filinkova

336180

1 Treść ćwiczenia

Celem ćwiczenia było stworzenie i analiza sieci Bayesowskiej modelującej zależności pomiędzy wybranymi cechami związanymi z przestępstwami w USA na podstawie zbioru danych z lat 1980-2014. Następnie należało stworzyć generator danych, który symuluje nowe obserwacje zgodnie z nauczona siecią.

2 Doprecyzowanie

Dane wejściowe zostały ograniczone do cech:

- Victim Sex, Victim Age, Victim Race
- Perpetrator Sex, Perpetrator Age, Perpetrator Race
- Relationship, Weapon

Zbiór danych został przefiltrowany poprzez usunięcie wierszy zawierających wartości Unknown w celu poprawnego estymowania parametrów sieci.

3 Cel i opis eksperymentów

3.1 Cel

Celem eksperymentu było:

1. Wydobycie struktury zależności pomiędzy zmiennymi przy użyciu uczenia struktury sieci Bayesowskiej.
2. Estymacja prawdopodobieństw warunkowych dla uzyskanej sieci.
3. Analiza wpływu poszczególnych cech na inne zmienne.
4. Implementacja generatora danych losowych zgodnie z rozkładem reprezentowanym przez sieć.

3.2 Opis eksperymentów:

- **Uczenie struktury sieci:** Przeprowadzono proces uczenia struktury sieci Bayesowskiej przy użyciu metody structure learning.
- **Uczenie parametrów:** Po uzyskaniu struktury sieci dokonano estymacji prawdopodobieństw warunkowych.
- **Analiza wyników:** Zwizualizowano sieć oraz dokonano analizy uzyskanych rozkładów prawdopodobieństwa.
- **Generowanie nowych danych:** Opracowano funkcje, która na podstawie niepełnych obserwacji losowo generuje brakujące wartości.

4 Przygotowanie środowiska i danych

Skrypt można uruchomić przez terminal za pomocą polecenia:

1. `git clone https://gitlab-stud.elka.pw.edu.pl/afilinko/wsi.git`
2. `python3 -m venv venv`
3. `source venv/bin/activate`
4. `cd /lab7`
5. `pip install -r requirements.txt`
6. `python3 main.py`

5 Wyniki

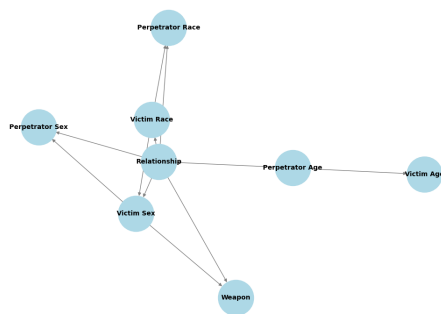


Figure 1: graf struktury sieci Bayesowskiej (skierowany acykliczny graf (DAG))

- Wzły (wierzchołki) reprezentują zmienne w danych (np. Victim Age, Weapon, Perpetrator Race).

- Krawędzie (strzałki) oznaczają skierowane zależności między zmiennymi, wykryte podczas uczenia struktury (structure learning.fit).

Częściowa obserwacja: "Victim Sex": "?", "Victim Age": "20", "Victim Race": "?", "Perpetrator Sex": "Male", "Perpetrator Age": "?", "Perpetrator Race": "Asian/Pacific Islander", "Relationship": "Friend", "Weapon": "Strangulation"

Wygenerowane wyniki:

- 'Victim Sex': 'Male', 'Victim Age': '20', 'Victim Race': 'Black', 'Perpetrator Sex': 'Male', 'Perpetrator Age': '17', 'Perpetrator Race': 'Asian/Pacific Islander', 'Relationship': 'Friend', 'Weapon': 'Strangulation'
- 'Victim Sex': 'Male', 'Victim Age': '20', 'Victim Race': 'Asian/Pacific Islander', 'Perpetrator Sex': 'Male', 'Perpetrator Age': '39', 'Perpetrator Race': 'Asian/Pacific Islander', 'Relationship': 'Friend', 'Weapon': 'Strangulation'
- 'Victim Sex': 'Male', 'Victim Age': '20', 'Victim Race': 'White', 'Perpetrator Sex': 'Male', 'Perpetrator Age': '33', 'Perpetrator Race': 'Asian/Pacific Islander', 'Relationship': 'Friend', 'Weapon': 'Strangulation'

6 Wnioski

- Zależności między zmiennymi: Wyniki potwierdzają znane korelacje, np. wpływ wieku na wybór broni czy związek rasy ofiary i sprawcy.
- Skuteczność modelowania: Sieć Bayesowska skutecznie odwzorowuje rzeczywiste rozkłady i pozwala na generowanie realistycznych nowych obserwacji.
- Ograniczenia: Sieć nie uwzględnia czynników poza zestawem danych, takich jak kontekst społeczny czy motywacje sprawców.

Podsumowanie: Ćwiczenie pozwoliło na praktyczne zastosowanie metod uczenia struktury i parametrów sieci Bayesowskiej. Uzyskana sieć poprawnie odwzorowała zależności pomiędzy analizowanymi zmiennymi, a opracowany generator danych umożliwił symulacje nowych przypadków zabójstw zgodnie z rozkładami wynikającymi z modelu.