Heart Attack Analysis using Bayesian Network

Kankana Ghosh
Master's Degree in Artificial Intelligence, University of Bologna
kankana.ghosh@studio.unibo.it

February 19, 2023

Abstract

This mini-project aims to perform a vertical case study on Bayesian Network. I used the python library pgmpy to implement and explore some of the major functionalities of the Bayesian Network. I found that Bayesian Network works quite efficiently in finding inferences and how the data are related.

Introduction

Domain

Heart Attack is one of the most serious disease in the segment of cardiovascular disease that occurs due to the interruption of blood circulation to muscle of the heart which damages the heart muscle. Diagnosing heart diseses is a crucial task as it depends on different factors including cholesterol, genetic heart disease, high blood pressure, low physical activity to mention some.

An enormous amount of data is being generated on a daily basis and acquiring knowledge from these data is essential. This report implements a probabilistic graphical model Bayesian Network to represent the knowledge and find inferences and dependencies of the features and predict which features are most helpful in predicting Heart Attack.

This report presents an original Bayesian Network generated based on research on the connection between various features present in the dataset used.

Aim

The purpose of this project is to implement partof the Bayesian network and experiment with some of its concepts. It also includes a brief Exploration with Exact and Approximate Inferencing under different conditions.

Method

The project mainly focuses on using the pgmpy library to build the Bayesian Network and use several methods[1] to implementits various functionalities and run queries. This project also presents a brief comparison of the different inferencing approach; Exact Inference: testing using four different heuristics for variable elimination functions provided by pgmpy library; Approximate Inference: using likelihood weighted sampling.

Results

The outcomes show that the different heuristics need almost the same time to reach to the conclusion and the order of variable elimination are similar except for the MinFill heuristic which differs by an insignificant amount. Thus, all four heuristics work well for the simple case study. Similarly for the Exact and Approximate Inferencing do not show much variation in results and work equally well.

Model

The Bayesian Network is built based on the Heart Attack Analysis and Prediction dataset that contains 14 common features. Below Fig.2 is an interpretation of the created model:

- The nodes: age, sex, ChestPainType, RestingBP Cholestrol, FastingBS, RestingECG, MaxHR ExerciseAngina, oldpeak, ST_Slope, no.ofVes Thalassemia and HeartAttack.
- The connections between each node is created based on analysis of the domain and thorough research through various scientific medical publications and online resources. Below Fig.1 is a snapshot of the generated CPT for HeartAttack:

CPT of HeartAttack:

ChestPainType	ChestPainType(0)	...	ChestPainType(3)	ChestPainType(3)
Cholestrol	Cholestrol(120-200)	...	Cholestrol(320+)	Cholestrol(320+)
RestingBP	RestingBP(110-130)	...	RestingBP(130+)	RestingBP(90-110)
HeartAttack(0)	0.6636363636363637	...	0.5	0.5
HeartAttack(1)	0.3363636363636364	...	0.5	0.5

Fig. 1 CPT of HeartAttack

Analysis

Experimental setup

The network is analyzed by finding the independencies and descendants of each node. The Markov Blanket and the active trails of the nodes are found out by using the pgmpy functions. The model is experimented with various probability queries both for Exact and Approximate Inferencing approaches. The queries are helpful in establishing situations of an individual's risk of Heart Attack. Some major risk factors include High Cholesterol level, high Blood Pressure, Maximum Heart Rate, a type of Chest Pain. Elderly male has more risk of Heart Attack. These are the certain assumptions that are expected to give positive results while querying the model.

Result

The model generated a 0.8154 probability, i.e an 82% possibility of Heart Attack given a Resting Blood Pressure of 110-130, Maximum Heart Rate of 130-160 and Chest Pain of type Atypical Angina. People with Cholesterol of 200-240 has higher risk of Heart Attack. As per the assumptions elderly male people have a little higher possibility but they alone cannot predict risk of Heart Attack of a person.

Conclusion

The Bayesian Network is quite efficient and works very well for small datasets. It is flexible and provides interpretable representations. Querying through various targets and evidences helped in good knowledge discovery and helped in learning a lot about the domain. Although it might have some limitations in working with larger datasets with huge number of features as creating a network and establishing connection between them can be time consuming and also requires expertise of the domain.

Links to external resources

a link to the dataset heart.csv

References

1. Machine Learning Based Heart Attack Prediction Machine learning-based heart attack prediction: A... | F1000Research https://www.nia.nih.gov/health/heart-health-and-aging#:~:text=The%20most%20common%20aging%20change,more%20common%20as%20we%20age.

2. https://github.com/mwritescode/EndoBayes

3.https://www.healthline.com/health/high-cholesterol/levels-by-age#:~:text=As%20we%20age%2C%20cholesterol%20levels,more%20frequent%20testing%20is%20recommended.

4. https://www.healthline.com/health/high-cholesterol/levels-by-age#:~:text=As%20we%20age%2C%20cholesterol%20levels,more%20frequent%20testing%20is%20recommended.

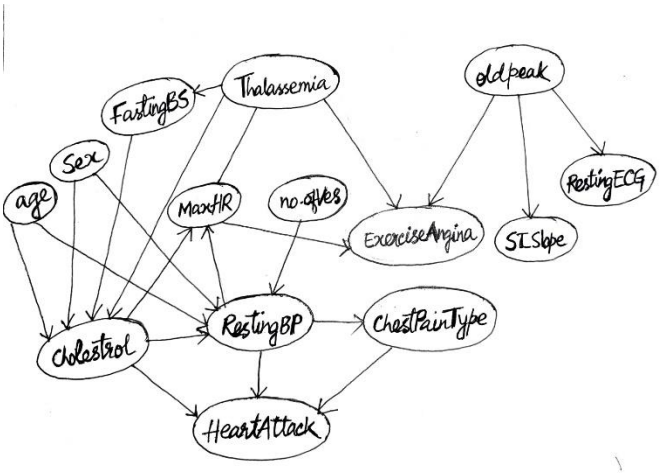

Fig. 2 The Bayesian Network