ARTIFICIAL INTELLIGENCE (E016330)
ALEKSANDRA PIZURICA
GHENT UNIVERSITY
2020/2021 AY

Assistants: Srđan Lazendić, Shaoguang Huang,
Nina Žižakić, Nicolas Vercheval
Professor: Aleksandra Pizurica

Practical Session 2: Bayesian Networks

1 Introduction

The purpose of this practicum is that the student gains practical experience in several modeling and application aspects of Bayesian networks. A simple medical diagnosis problem is considered. In order to model the Bayesian network of this diagnosis problem, we will use the Python library *pomegranate*, whose documentation is available on its official website [2]. The provided jupyter notebook can be used as a basis for the implementation.

2 Assignment

2.1 Medical diagnosis

The concepts learned about Bayesian networks will be put in practice with a medical diagnosis problem, which is based on a fictitious chest clinic model [1].

Consider the following qualitative knowledge about the medical diagnosis of several respiratory diseases:

A symptom like *shortness* of breath can be caused by the following diseases: *pneumonia*, *lung cancer* and/or *bronchitis*. Moreover, patients with *pneumonia* and/or *bronchitis* often have a very annoying *loose cough*. *Pneumonia* and/or *lung cancer* are usually associated with severe *chest pain*. *Pneumonia* often causes severe *fever*, but this can also be caused by *common cold*. However, a *common cold* can be recognized by a *runny nose*. Sometimes symptoms like *loose cough*, *chest pain* and/or *shortness of breath* occur without explanation, or are caused by other diseases not considered here. Sometimes diseases occur together. A *weakened immune system* increases the risk of *pneumonia*. Also, *lung cancer* increases this risk. *Smoking* is a serious risk factor for *bronchitis* and *lung cancer*.

The foreseen expert system is meant to infer the preliminary diagnosis and probabilities of the considered diseases only based on the observed symptoms and the accumulated expert knowledge (statistical data), without having access to any further tests. E.g., the system can operate online posing the relevant questions to the user and making inference in real time based on the answers. Based on the results of the Bayesian inference, further tests (including some standard tests the considered diseases) should be recommended.

Now consider the following quantitative information about a population of patients:

- Statistics (finding percentage):

  - Weakened immune: 5,3%
  - Smoking: 29%
  - Pneumonia: 1,8%
  - Lung cancer: 3,7%
  - Bronchitis: 9,0%
  - Common cold: 37%
  - Fever: 8,5%
  - Loose cough: 19%
  - Chest pain: 14%
  - Shortness of breath: 18%

- Correlations:
  - 80% of the population diagnosed with bronchitis suffer from shortness of breath.
  - 50% of the population without bronchitis, but with pneumonia and /or lung cancer suffer from shortness of breath, 10% of the population for which the mentioned diseases cannot be determined suffer from shortness of breath.
  - 90% of the population with pneumonia and/or lung cancer suffer from chest pain, 10% of the population for which the mentioned diseases cannot be determined suffer from chest pain.
  - 90% of the population with pneumonia and/or bronchitis suffer from loose cough, 10% of the population for which the mentioned diseases cannot be determined suffer from loose cough.
  - From the population with common cold, about 20% has high fever. In 95% of the cases, common cold is accompanied by a runny nose. Only 1% of the patients without common cold has a runny nose. 0.1% of the population for which the mentioned diseases cannot be determined suffer from high fever.
  - 94% of the population with pneumonia has a high fever

- Risk Factors:
  - 10% of the smokers have lung cancer, and only 1% of the non-smokers.
  - 30% of the smokers have bronchitis, and only 1% of the non-smokers.
  - 30% of the people with weakened immune system, and about 5% of the people with lung cancer are diagnosed with pneumonia. Only 0.1% of the population without these risk factors is diagnosed with pneumonia.

2.2 Questions

Build a medical expert system using a Bayesian network with the information provided in previous sections.

1. Determine the variables of the network and the network's structure (topology) that reflects the causal relationships (include a figure). These relationships can be deduced from the qualitative description presented in section 2 and using common sense.(10%)

2. Determine the number of parameters that are needed for representing the network based on the compact representation (e.g., $P(X = \mathit{True} \mid \mathit{parents}(X))$). (5%)

3. Write a function that groups the variables in subsequent stages in order to schedule the message updates, and describe concisely the idea in the report. (5%)

4. (a) Write down the conditional probabilities making use of the quantitative information (e.g., $P(X = \mathit{True} \,|Y, Z) = 0.1$) (10%). (b) Calculate using the quantitative information the complete conditional probability table for each variable; if the quantitative information provided is insufficient to fill in the CPT of the nodes, propose a further assumption. Any assumption must have common sense and be consistent up to a small percentage error with the statistics. Explain your approach (15%).

Run your system and verify the diagnostic guidelines described below extracted from a medical textbook. Explain the results and support your explanation with the posterior probabilities provided by the expert system.

5. In the case of high fever without a runny nose, pneumonia must be considered. Why? (10%)

6. Lung cancer is often found in patients with chest pain, shortness of breath, no fever, and usually no loose cough. Support this statement with the findings of your network. (10%)

7. Bronchitis and lung cancer often coexist. For instance, patients with bronchitis often develop lung cancer and vice versa. However, these diseases have no causal relationship. That is, bronchitis is not caused by lung cancer, and lung cancer is not caused by bronchitis. Can you explain why these diseases often occur together? (10%)

Take advantage of the expert system to evaluate the following cases:

8. A patient suffers high fever and a severe loose cough (although he is not a smoker). He speaks nasally. What is your diagnosis (a weakened immune system may be excluded)? In order to confirm your diagnosis (i.e., discard any other disease), what other evidence should be noticed? (10%)

9. A patient suffers of shortness of breath and has a weakened immune system. This patient is a heavy smoker. What are the possible diseases? What else is needed to provide a confident diagnosis? (5%)

10. Is Pneumonia directly causing Shortness of Breath, or the correlation is to be attributed to common factors? To justify your statement and give a quantitative estimate of its direct influence, write a new Model as before but replace Pneumonia as a dependency of Shortness of Breath with an independent clone variable. (10%)

3 Submission

Include the jupyer notebook you used to build the network (this can be downloaded from Colab) to the report. Report your results from the exercises above and discuss the results in a summary report. For each q (diagnosis, explanation, advice, ...) derived from the evidence e, report the conditional probability $P(q|e)$ obtained with the model.

Important remarks:

- Write concise answers to the questions. Indicate clearly the number of each task/question when answering it.
- The overall score may be influenced by the report presentation too (pay attention to the clarity of answers and avoid sloppiness).
- The submitted file should be named after both authors and specify the project subject (e.g. Nicolas Vercheval & Shaoguang Huang: *BayesianNetworks_nverchev-shuang.zip*).
- It is the student's responsibility to check that the submitted notebook is working and leads to the same results of the report. Avoid any situation that prevents the correct evaluation of your report.
- Deadline: **November 18, 2020, 23:59 CET**

References

[1] S. Lauritsen and D. Mirror Halter. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Royal Statistical Society,*, 1988.

[2] pomegranate. Python library for bayes nets and other probabilistic models. https://pomegranate.readthedocs.io/en/latest.