Alessandro Allegranzi

BU MET CS 767

Assignment 6: Bayesian Network

6/19/2024

06/19/2024

MET CS 767 Assignment 6: Bayesian Networks

*Alessandro Allegranzi*

Create your own application of a Bayesian network, like the “sprinkler” problem in the notes. You can do this by hand, via a Python program, or using an online Bayesian network tool.

Please leave the gray text and the headings unchanged etc.

# Requirements for your application

Supply the requirements, including the nature of its inputs and that of its outputs.

The following Bayesian Network can help in understanding how different factors influence voting behavior in an election, aiding in the analysis of voter decisions and the development of political strategies.

**Events (Nodes):**

1. **Political Affiliation (PA)**: The political leaning of a voter (e.g., 0: Conservative, 1: Liberal, 2: Independent).
2. **Economic Outlook (EO)**: The voter's perception of the economic future (0: Positive, 1: Negative).
3. **Vote (V)**: The actual voting decision (0: Conservative, 1: Liberal).

**Dependencies:**

* **Political Affiliation** influences **Actual Vote**.
* **Economic Outlook** influences **Actual Vote**.
* **Economic Outlook** influences **Political Affiliation. (**Lower economic outlook affects political affiliation towards conservative.**)**

**Requirements:**

* **Inputs**:
  + **Political Affiliation**: An integer representing the political leaning of a voter.
  + **Economic Outlook**: An integer representing the voter's perception of the economic future.
* **Outputs**:
  + The probability distribution of the **Vote**, given the political affiliation and economic outlook.

I have implemented the above network in python using the pgmpy library and will attach the .py file with my submission [1]. [GH link](https://github.com/1-8192/bu_cs_767/blob/main/moduel_6/bayesian_network.py)

# Diagram

Provide a figure like the one in the notes for the Bayesian network.

I created the diagram on the [BayANet website](https://profgavinbrown.github.io/projects/bayes_nets/). The probabilities are arbitrary values, since real data is not available. I will include a [json file](https://github.com/1-8192/bu_cs_767/blob/main/moduel_6/bayes.json) of the diagram with the submission.

A diagram of a political process

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

# Example 1

Give an example of an input and the resulting outputs, with an explanation of the computation.

**Input and output of python app:**

“What is the probability of a liberal political affiliation given a liberal vote and positive economic outlook?”

The input is a positive economic outlook (1) and a liberal vote (1), which should lead to a likelier liberal affiliation (1).

result = inference.query(*variables*=['Political Affiliation'], *evidence*={'Economic Outlook': 1, 'Vote': 1})

+--------------------------+------------------------------+

| Political Affiliation | phi(Political Affiliation) |

+==========================+==============================+

| Political Affiliation(0) | 0.2319 |

+--------------------------+------------------------------+

| Political Affiliation(1) | 0.5072 |

+--------------------------+------------------------------+

| Political Affiliation(2) | 0.2609 |

+--------------------------+------------------------------+

The answer to the initial question is .5072 using the python model.

**Computation:**

1. **P(V=Liberal∣P=Liberal,E=Positive)**:
   * From the CPD of Vote = **0.8**
2. **Compute P(P=Liberal∣E=Positive)**:
   * From the CPD of Political Affiliation given Economic Outlook = **0.5**
3. **Compute P(V=Liberal∣E=Positive):**

P(V=Liberal∣E=Positive)=(0.5⋅0.25)+(0.6⋅0.5)+(0.4⋅0.25) = **0.65**

1. **Compute P(P=Liberal∣V=Liberal,E=Positive)**:
   * Substitute the values into Bayes' theorem:

P(P=Liberal∣V=Liberal,E=Positive) = 0.8⋅0.50 / .65 = **0.6154**

The calculated probability is 0.6154. It does not match the code output exactly, but that is expected given the Bayesian Network model does not simply run Bayes’ but applies other operations as well [1].

# Example 2

Give an example of an input and the resulting outputs, with an explanation of the computation.

**Input and output:**

“What is the probability of a conservative political affiliation given a liberal vote and a negative economic outlook?”

The input is a negative economic outlook (0) and a liberal vote (1), which should lead to a less likely liberal affiliation (1) than in example 1.

result\_two = inference.query(*variables*=['Political Affiliation'], *evidence*={'Economic Outlook': 0, 'Vote': 1})

+--------------------------+------------------------------+

| Political Affiliation | phi(Political Affiliation) |

+==========================+==============================+

| Political Affiliation(0) | 0.6545 |

+--------------------------+------------------------------+

| Political Affiliation(1) | 0.2727 |

+--------------------------+------------------------------+

| Political Affiliation(2) | 0.0727 |

The answer to the initial question is .6545 using the python model.

**Computation:**

**Compute P(V=Liberal∣P=Conservative,E=Negative)**

* From the CPD of Vote = **0.4**

 **Compute P(Pol=Conservative∣E=Negative)**:

* From the CPD of Political Affiliation given Economic Outlook = **0.6**

 **Compute P(V=Liberal∣E=Negative)**:

P(V=Liberal∣E=Negative)=(0.6⋅0.6)+(0.2⋅0.2)+(0.2⋅0.2) = **0.44**

 **Compute P(P=Conservative∣V=Liberal,E=Negative)P(P=Conservative∣V=Liberal,E=Negative)**:

* Substitute the values into Bayes' theorem:P(P=Conservative∣V=Liberal,E=Negative) = 0.4 ⋅ 0.60 / .44 = **0.5455**

The calculated probability is 0.5455. It does not match the code output exactly, but that is expected given the Bayesian Network model does not simply run Bayes’ but applies other operations as well [1].

# Scaling

Imagine a real-world Bayesian network built to assess the economic impact of connected events, and implemented as in the your example. What would the main obstacles be to its practical development and use? Avoid generalities about Bayesian networks; concentrate on your application and its extrensions.

A major obstacle in developing and using a Bayesian network like the one described for assessing voter intent lies in the accurate and comprehensive data collection required to define the conditional probability distributions (CPDs). For real-world applications, especially those involving complex economic systems, the relationships between variables can be intricate and influenced by many factors. Gathering sufficient, reliable data to accurately model these relationships can be challenging due to the dynamic nature of economic systems, where the influence of one variable on another can change over time due to evolving market conditions, policy changes, and unforeseen global events. Data on voting intent and behavior can be especially hard to gather given the private nature of voting in the US. Additionally, the model's simplifications and assumptions, necessary for computational tractability, might not capture the full complexity of real-world interactions, leading to potential inaccuracies in predictions.

Another significant challenge is the computational complexity associated with performing inference in large-scale Bayesian networks. As the number of variables and their states increases, the computational resources required to perform exact inference (e.g., using the Variable Elimination algorithm) can grow exponentially, making it impractical for real-time or large-scale applications. Approximate inference methods can mitigate this to some extent but may introduce their own inaccuracies. Furthermore, the model's utility and accuracy are contingent upon its continuous update and validation against new data and emerging economic theories. This necessitates an ongoing commitment to data collection, model refinement, and validation to ensure the Bayesian network remains a reliable tool for economic impact assessment, posing logistical and resource-related challenges.

# References

[1] “Bayesian Network”. *Pgmpy Docs*. https://pgmpy.org/models/bayesiannetwork.html

# Evaluation

