

# ▼ The Understanding about Bayesian Networks

Written by Zhian Chen

## Personal information

|              |   |
|--------------|---|
| Student Name | Zhian Chen  |
| Student ID   | 13083112  |
| GitHub Link  | <a href="https://github.com/baojian123/Uts_ML_ID13083112">https://github.com/baojian123/Uts_ML_ID13083112</a>   |
| Colab Link   | <a href="https://colab.research.google.com/github/baojian123/UTS_ML_ID13083112/blob/master/Copy_of_A1_13083112(2).ipynb">https://colab.research.google.com/github/baojian123/UTS_ML_ID13083112/blob/master/Copy_of_A1_13083112(2).ipynb</a> |

## 1.Content

### 1.1 Introduction

This report will provide an introduction the paper "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks" by Greory F. Cooper which proves the calculation of probabilistic inference using bayesian belief networks is an NP-hardness problem.

### 1.2 Bayesian Belief Network

Before we talk about that paper, we first need to introduce the Bayesian Belief Network, because that paper describes the Bayesian Belief Network shortly. Subsequently, we denote the Bayesian Belief Network as BN.

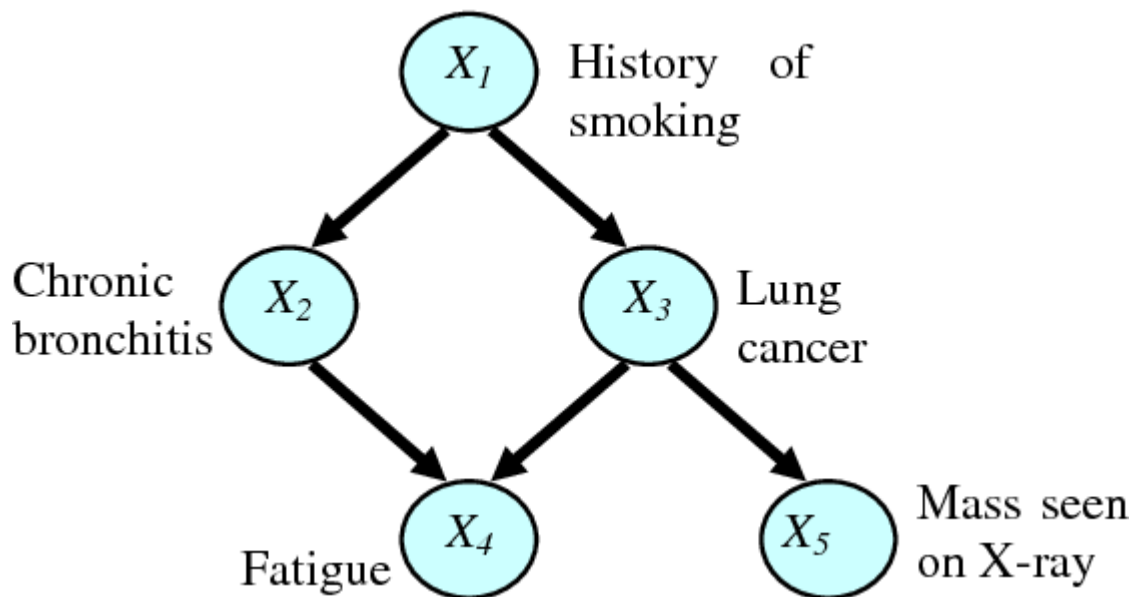
The BN has a little bit difference with other machine learning method. Its theory bases largely on the statistic and probabilistic knowledge. This algorithm will construct a graphic structure for probabilistic relationship between several events. By BN, we just need to consider the known dependencies rather than to build a fully-connected dependency matrix. Moreover, the representation of graphic strucure can be more intuitive, and we will discuss that later. The structure of the BN is a direct acyclic graph (DAG), and this will be related to the conditional independence

One of the reasons why we need BN is that there are some problems which cannot be calculated by Bayesian theory. The Bayesian theory can be presented as follow:

$$P(B|A) = \frac{P(A|B)*P(B)}{P(A)} \quad (1)$$

For instance, sometimes doctor need to diagnose a patient by other related factors. There is an example given by Neapolitan(2004). Suppose we know whether the patient smoke hold a direct impact to whether they have bronchits or whether they have lung cancer, and these two diseases have a direct

influence to whether they experience fatigue. Furthermore, whether the patient have lung cancer has a direct effect to whether their chest X-ray is positive. Obviously, only Bayesian theorem, as shown in (1), cannot predict the result with this complex relationship structure. Furthermore, it is too abstract to understand the casual relationship between these variables based on the above description. However, if we use the BN to present the probabilistic relationship as shown as Figure 1.1, it will be clearer than the description.



**Figure 1.1 The Bayesian network of whether patients have bronchitis or a lung cancer**

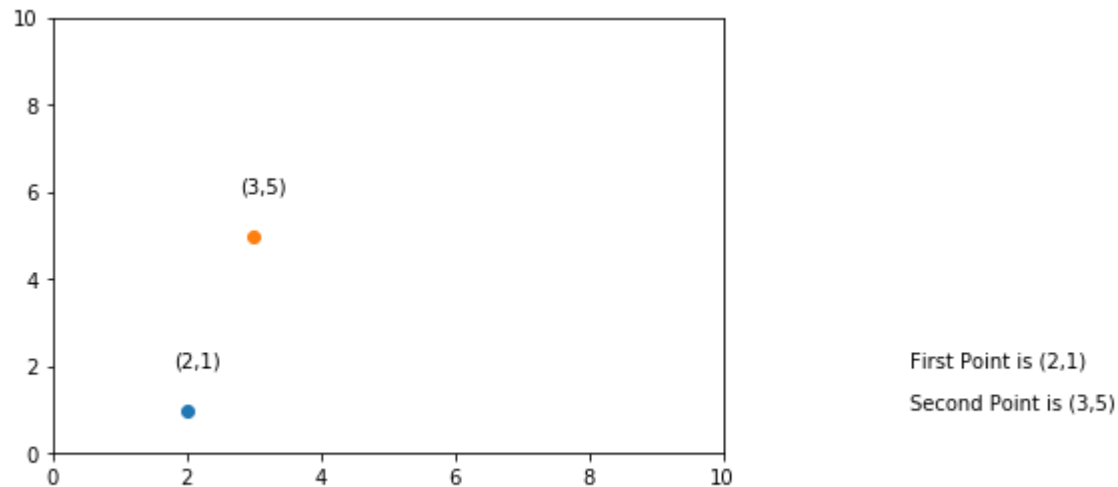
There is another example for representation. In Figure 1.2, we have two point on a 2-dimensional space. While the coordinate of first point is (2,1), the coordinate of second one is (3,5). When we want to identify their position relationship on a space, a graph structure is clearer than the coordinate.

```

from mpl_toolkits import mplot3d
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
fig = plt.figure()
plt.plot(2,1,'o')
plt.plot(3,5,'o')
plt.xlim(0,10);
plt.ylim(0,10);
plt.text(1.8,2,"(2,1)");
plt.text(2.8,6,"(3,5)");
plt.text(12.8,2,"First Point is (2,1)");
plt.text(12.8,1,"Second Point is (3,5)");
plt.text(4,-2,"Figure 1.2: the comparison of different representation");

```





### 1.3 Conditional Probability

For computing the probability of the result of a BN, we need to compute the conditional probability in terms of their dependencies. Therefore, in this section we will introduce the probability. In order to calculate the probabilistic function, there are some approaches provided by Neapolitan (2004) and we will present as follow.

At the beginning, probabilistic theory relies on the repeated experiment, such as tossing a thumbtack, dropping a dice hundreds of times and drawing a card from a desk of 52 cards. For the last two instance, all events are **indifferent**. Indifference means that we can treat the all events as the same, because they have the equal probability to happen.

Nevertheless, the events of tossing a thumbtack have different probabilities due to the irregular shape of thumbtack, but we still can obtain their probabilities though repeated experiment. Suppose that we define the event "head" when the thumbtack we toss lands on its head, and for the event that the thumbtack lands with the edge of the head, we call it "tail". If we toss the thumbtack 10000 times and the event "head" happen 3373 times, we would estimate the probability of the event "head" to be 33.73% approximately. This method for calculating the probability by repeat experiments are called **relative frequencies**.

In addition to the relative frequencies, Neapolitan also provide an instance about **the subjective probability** or called **the degree of belief**. This terms demonstrates that when an event cannot be predicted and cannot be repeat several times at the same condition, to what extent you can determine that this event will occur?

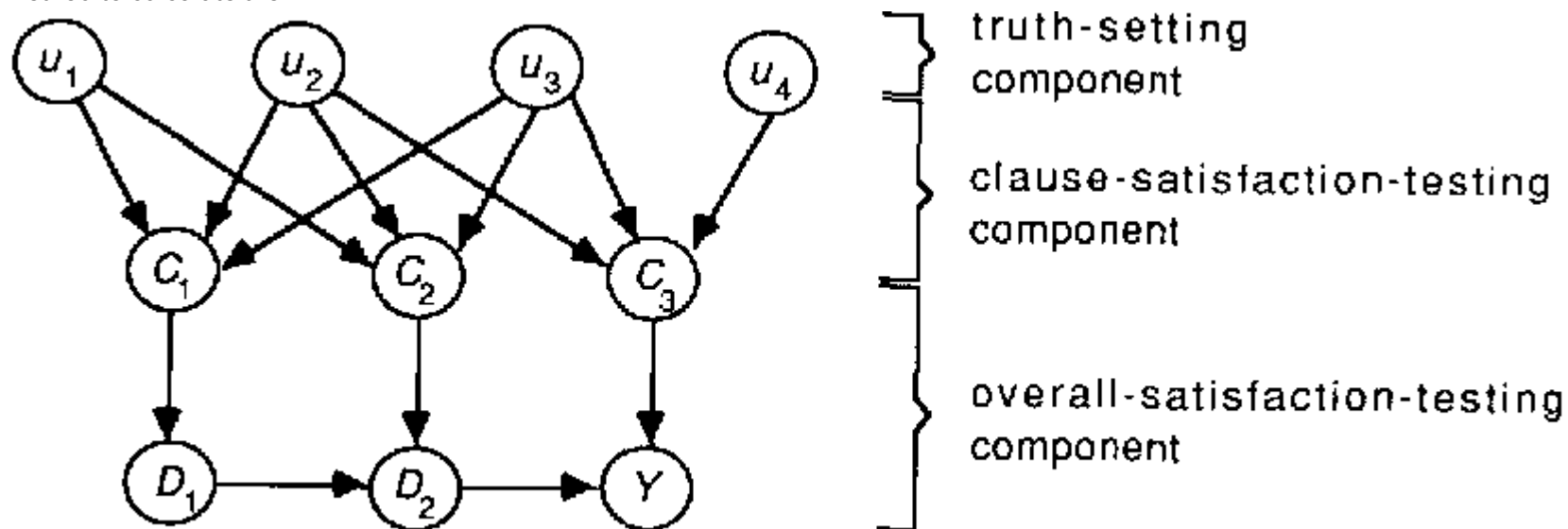
The probabilistic inference using Bayesian network (PIBNET) relies on conditional probability. Conditional probability can be defined as the probability of event A when the event B has occurred, denoted by  $P(A|B)$ . It also can be obtained by calculating the area of event A in B sample space.

## 1.4 Content of Paper

As a classifier, Bayesian network need to predict the result. so we need a probabilistic inference algorithm which can infer the probability of an events by other known events.

This paper proves the computational complexity of probabilistic inference using Bayesian belief network is a NP-hard problem by proving that the decision problem version of probabilistic inference using Bayesian belief network (PIBNETD) also is an NP-hard problem. Because PIBNETD problem calculates  $P(Y = T) > 0? 1 : 0$  while PIBNET problem calculates  $P(Y = T)$ . If we can prove PIBNETD problem is NP-hard, we also can prove PIBNET problem is NP-hard.

Bayesian belief network can be separated by multiply connected belief network and singly connected belief network. Building the multiply connected belief is a NP-hard problem, it means it is imposible to build a general and efficient probabilistic inference algorithm. However, there is some special method to calculate the PIBNET.



A belief network structure. The probabilities in this belief network are defined in Section 2.1.

Figure 1.3 the transformation from 3SAT problem to Bayesian Network

## 2. Innovation

To prove a problem is NP-hard, we need to reduce a known NP-hard problem to current problem and show this reduction can be done efficiently. If a problem is NP-hard and NP simultaneously, this problem is a NP-complete problem. The NP-hard problem means the computational complexity of the problem is pretty high. One of the famous NP-hard problem is Travel Salesman Problem. One salesman want to travel all  $N$  cities. There is an airline between any two cities, but each airline has different cost. Their business can only reimburse  $C$  dollar. We need to find a path and the sum of cost is lower than  $C$ . The time complexity of the solution for this problem is  $O(N!)$  for finding a Hamiltonian cycle in  $N$  cities. In the worst case, there is no one path can be lower than the outlay. Then we need to traverse all of the solution space tree to build a Hamiltonian cycle.

Rather than other researches, this paper want to prove that the probabilistic inference using Bayesian network is an NP-hardness problem by a new method.

s-t network reliability problem is a problem that there is a graph consist of many edges, and for each edge, there is a probability to fail. What we need to calculate is the probability about finding a path of unfailed edges from node  $s$  to node  $t$ . Provan and Ball (1983, cited in Cooper 1990) has shown that the s-t network reliability problem is a NP-hard problem for DAG. Hence, we can reduce s-t network to DAG and apply it into the PIBNET problem.

This paper model the 3SAT problem by Bayesian network, and then prove the PIBNETD problem is a NP-hard problem.

### 3. Technical Quality

Bayesian network is a state-of-art machine learning method. It can be applied into different domains. Moreover, This paper involves lots of mathematical concept and problem, like 3SAT problem and s-t network reliability problem. they provide a new way to prove that the PIBNET problem is NP-hard. One of the advantage is that we can calculate the joint probability of  $n$  events easily rather than list all the samples in the sample space and calculate their relative frequency. And there is some good idea in this paper. For example, the overall-satisfaction testing components in Figure 1.3 use a set of intermediate dummy node  $D_k$  to divide a large AND gate into several small AND gate and restrict the indegree of each node in belief network less than 4. because when the number of clauses increases, the indegree of node  $Y$  would increase if we link all the clause  $C$  to  $Y$ .

### 4. Application and X-Factor

According to Cooper (1990), one of the advantages using 3SAT problem to prove that the PIBNET is NP-hard is that we can solve some medicine problems by modeling them to 3SAT problem. This paper uses the 3SAT problem to map the disease diagnosis problem. As shown as Figure 1.3, we can treat the set  $U$  as different diseases and the set  $C$  as different symptoms, to transform the diagnosing process to 3SAT problem. As proven as above, the 3SAT problem can be reduced to PIBNET problem. Therefore, the BN can be applied in medicine domain.

We suggest that the further research can identify the relationship between CNN and BN. Because Nerual Network is also a DAG. While the CNN model decomposes the image into different parts by several convolutional kernels, the BN also decomposes the sample space into different subspace by calculating the conditional probability.

## 5.Presentation

This paper presents the proof with a clear logic. The section structure of the whole paper is conducted step by step. We can follow the section to understand the concept from the low level to the high level. The figure in this paper can be used several times. such as giving an example of the structure of Bayesian network and 3SAT problem. Some significant concepts, like NP-hard proving method, have been presented in a clear way.

## Reference

Cooper, G.F. 1990, 'The computational complexity of probabilistic inference using bayesian belief networks', *Artificial Intelligence*, vol. 42, no. 2, pp. 393-405.

Neapolitan, R.E. 2004, *Learning Bayesian Networks*, Pearson Prentice Hall.

