**Name: Peter Sunny Shanthveer Markappa**
**Subject: Knowledge Representation**
**Student No.: R00208303**

# 1.1 Logical Reasoning

There are several ways to represent the facts and things in computer science.

| Language | Ontological Commitment | Epistemological Commitment |
|---|---|---|
| Propositional Logical | Facts | True / False / Unknown |
| First Order Logic | Facts, Objects, Relations | True / False / Unknown |
| Temporal Logic | Facts, Objects, Relations | True / False / Unknown |
| Probability Theory | Facts | Degree Belief ε [0,1] |
| Fuzzy Logic | Facts with degree of truth | Known interval values |

**Propositional Logic:** They represent only facts; the result will be True or False or Unknown

**First Order Logic:** *FOL is representing the language that humans know in a way that computer can understand. There are many ways that computer can understand the logic and this is one of the ways. This FOL is more expressive than Propositional Logic*

**Operators that are used in First Order Logic are:**

∧ --- Conjunction ∨ --- Dis-Junction ⇒ --- Implication ¬ --- Negation
⇔ --- Double Implication ∃ --- Existential Quantifier ∀ --- Universal Quantifier

## *Axiom*

**GreatGrandParent(a,d) ⇔ ∃ b, c Child(d,c) ∧ Child(c,b) ∧ Child(b,a)** → A is great-grandparent of d, d is child of c, c is child of b, b is child of a

**Grandchild(c,a) ⇔ ∃ b Child(c,b) ∧ Child(b,a)** → C is grandchild of a, c is child of b. b is child of a.

**Brother(x,y) ⇔ ∃ Male(x) ∧ Sibling(x,y)** → x and y are brothers, They both are siblings where x is male

**Sister(x,y) ⇔ ∃ Female(x) ∧ Sibling(x,y)** → x and y are sisters, They both are siblings where x is female

**Daughter(d,p) ⇔ Female(d) ∧ Child(d,p)** → D is daughter of p, d is child of p where d is female

**Son(s,p) ⇔ ∃ Male(s) ∧ Child(s,p)** → S is son of p. s is child of p where s is male

**Ancestor(a,x) ⇔ Child(x,a) ∨ ∃b Child(b,a) ∧ Ancestor(b,x)** → A is ancestor of x . x is child of a. b is child of a. b is ancestor of x.

**FirstCousine(c,d) ⇔ ∃ p1,p2 Child(c,p1) ∧ Child(d,p2) ∧ Sibling(p1,p2)** → C and d are first cousins. C is child of p1. d is child of p2. P1 and p2 are siblings.

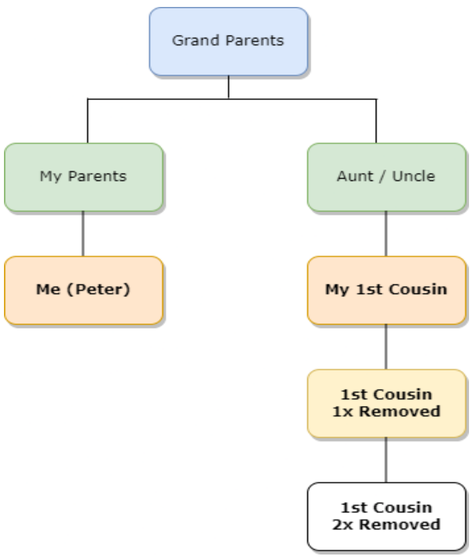
**Brother-In-Law(b,x) ⇔ ∃m Spouse(x,m) ∧ Brother(b,m)** → B is brother In law of x. x is spouse of m. b is brother of m.

**Sister-In-Law(s,x) ⇔ ∃m Spouse(x,m) ∧ Sister(s,m)** → S is sister in law of x. x is spouse of m. s is sister of m.

**Aunt(a,c) ⇔ ∃p Child(c,p) ∧ [Sister(a,p) ∨ Sister-In-Law(a,p)]** → A is aunt of c. c is child of p. a is sister or sister In law of p.

**Uncle(u,c) ⇔ ∃p Child(c,p) ∧ [Brother(u,p) ∨ Brother-In-Law(u,p)]** → U is uncle of c . c is child of p. u is brother or brother-in-law of p.

### Definition of $m^{th}$ cousin n times removed




Let me consider some of the definition which make easier to understand the above question

**First Cousin:** First Cousins are the people in the family who have same Grand Parent.

If I take my example my aunts or uncle's children will be my first cousin. If consider the family chart of Red Hugh O'Neill,

**Second Cousin:** who has same great grand parent as me, but grand parents will not be the same.

If I take my example my second cousin is my mom's cousin's children

**Third Cousin:** Who has same great great Grand Parents. So on…

### Now definition of mth cousin n times removed means:

Here the term **removed** refers to the number of generations separating the cousins.
So, if I say **"my 1st cousin once removed will be the child of my first cousin"**
if I say **"my 1st cousin Two times removed will be the grand child of my first cousin"**
If I say **"my second cousin once removed will be child of my second cousin"**

## Basic Facts

Matthew and Siobhan are parents of Red Hugh O'Neill → **(Parent(Matthew, Hugh) ∧ Parent(Siobhan, Hugh))**

Hugh is spouse of Siobhan_OD → **Spouse(Siobhan_OD,Hugh) ∧ Female(Siobhan_OD) ∧ Male(Hugh)**

Hugh is brother of Art and Brian → **Brother(Hugh, Art) ∧ Brother( Hugh, Brian)**

Hugh is Parent of Hugh og, Sarah, Alice, Henry → **(Childof(Hugh og , Hugh) ∧ Childof(Sarah, Hugh) ∧ Childof(Alice, Hugh) ∧ Childof(Henry, Hugh))**

Sarah is spouse of Arthur → **Spouse(Sarah, Arthur)**

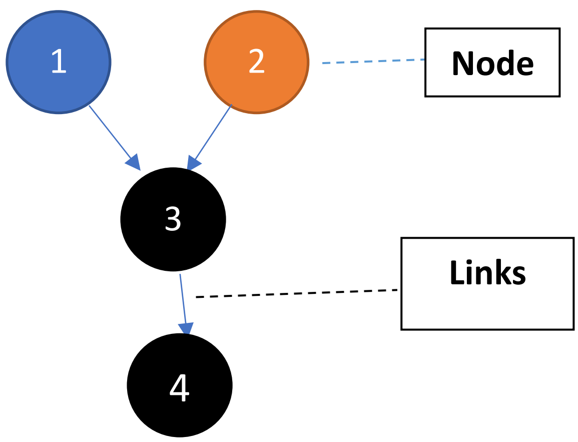
Alice is spouse of Randal → **Spouse(Alice, Randal)**

Hugh Magennis, Evelyn and Catherine are grandchildren of Hugh → **Grandchild(Hugh Magennis, Hugh) ∧ Grandchild(Evelyn, Hugh) ∧ Grandchild(Catherine, Hugh)**

# 1.2 Bayesian Networks

***Bayesian Network falls under the category of Probabilistic Graphical Modelling (PDM) technique that is used to compute uncertainties by using the concept of probability.***

Bayesian Network is also called *belief network and casual Probabilistic network* they are used to module uncertainties by using Directed Acyclic Graphs.




***A directed acyclic graph will basically tell the uncertainty of event occurring based on the conditional probability distribution of each random variable, a conditional probability table is used to represent this distribution of each variable in the network.***

**Joint Probability:** It is a statistical method where two or more events happening a same time. In measure of two or more events occurring at the same time i.e., P(A and B), the probability of this intersection of A and B can be written as P(A∩B)

**Conditional Probability:** Conditional Probability of an event 'B' is the probability that the event will occur given that an event 'A' has already occurred.

P(B|A): Probability of event B occurring, given that event A occurs.

If A and B are dependent events: P(B|A) = P(A and B) / P(A)

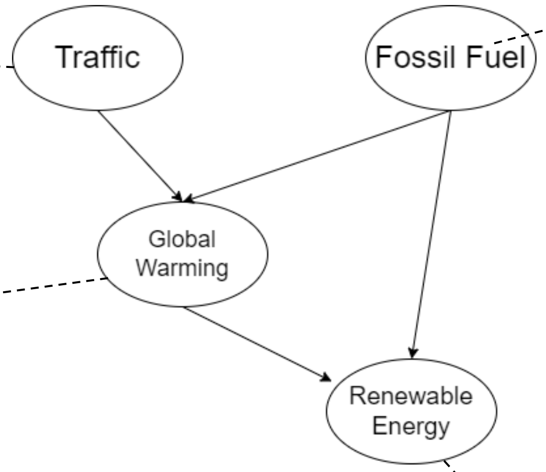
If A and B are independent events then P(B|A) = P(B)

Here I have built two different models with relation with Traffic, Fossil Fuel, Global Warming and Renewable Energy, second model I took AI and Employed.

**Description of the model Created:** The Global Warming is the child variable of the Traffic and Fossil Fuel so we can say that Global Warming is dependent on two variables Traffic and Fossil Fuel, i.e., *due to traffic and use of fossil-fuel the pollution will increase which will lead to global warming.* Renewable Energy is child variable of Global Warming and Fossil Fuel means Renewable energy is also dependent on Fossil Fuels and Global Warming, here we can say that increasing in the global-warming and using of more fossil-fuels leads us to move towards the renewable energy i.e., using renewable energy will reduce the global warming and use of fossil-fuels which are harmful to nature. Where as in second model the Employed variable is Dependent on AI, let's say if a student study AI properly then he will get Employed in company.

| P(Traffic) |
|---|
| 0.01 |

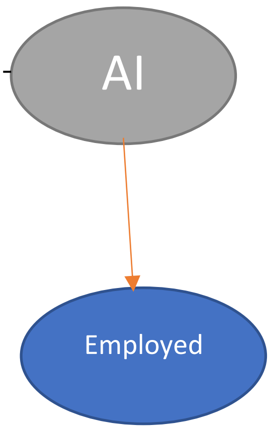
| P(Fossil Fuels) |
|---|
| 0.002 |




| Traffic | Fossil Fuels | Global Warming |
|---|---|---|
| **True** | True | 0.95 |
| **True** | False | 0.94 |
| **False** | True | 0.30 |
| **False** | False | 0.001 |

| Global Warming | Fossil Fuels | Renewable Energy |
|---|---|---|
| **True** | True | 0.90 |
| **True** | False | 0.85 |
| **False** | True | 0.70 |
| **False** | False | 0.001 |

| P(AI) |
|---|
| 0.5 |




| AI | Employee |
|---|---|
| **True** | 0.05 |
| **False** | 0.90 |

In the above distribution table of Global warming, we can see that if Traffic is True and Fossil Fuel is true then surely there is more chance of increase in the global warming, and if both are false then it is obvious that there will be no global warming as there will be less pollution, if we see the renewable energy table then if both global warming and fossil fuels are true i.e., if people are using more fossil fuels and if there is global warming in nature then probability of moving towards the renewable energy is more then when people use no fossil fuel and if there is no global warming.

**Advantages:**

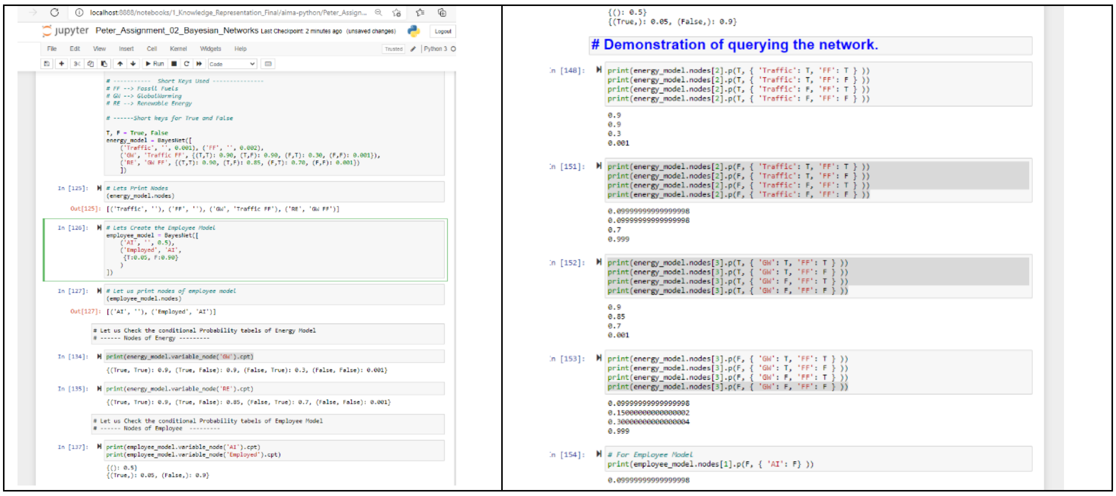
1. It has several features that can be used in real-life data analysis and management questions
2. It provides natural way of handling the missing data
3. It allows us to combine data with domain knowledge
4. It tells the casual relationship between variables
5. It has method which will avoid the overfitting of data and also show very good prediction accuracy even with small amount of data set even if the data has missing values.

**Dis advantages:**

1. Its ability to deal with continues data is limited.
2. It does not support feedback loops because it is acyclic graph. Spatial dynamics and temporal dynamics can be modelled in Bayesian network using separate network for each time slice which is very tedious task.

I have referred a paper for above section [1]

**Demonstration of Conditional Probability Tables and Querying of the network**



**Querying Network**

P(GLOBAL WARMING = TRUE | Traffic= TRUE , FOSSIL FUEL = TRUE)

P(GLOBAL WARMING = TRUE | Traffic= TRUE , FOSSIL FUEL = FALSE)

P(GLOBAL WARMING = FALSE | Traffic= TRUE , FOSSIL FUEL = TRUE)

P(RENEWABLE ENERGY = TRUE | Global Warming= TRUE , FOSSIL FUEL = False)

P(RENEWABLE ENERGY = FALSE | Global Warming= FALSE , FOSSIL FUEL = TRUE)

```
energy_model = BayesNet([
    ('Traffic', '', 0.001), ('FF', '', 0.002),
    ('GW', 'Traffic FF', {(T, T): 0.95, (T, F): 0.94, (F, T): 0.30, (F, F): 0.001}),
    ('RE', 'GW FF', {(T, T): 0.90, (T, F): 0.85, (F, T): 0.70, (F, F): 0.001})])

# Lets Create the Employee Model
employee_model = BayesNet([
    ('AI', '', 0.5),
    ('Employed', 'AI',
     {T: 0.05, F: 0.90}
     )])
```

## 1.3.1 Learning Developing and Evaluation Model

https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/

*This is link for the code downloading-→ winequality-white.csv is the file to be downloaded and kept it in aima-data as it is, it has all the data in single column so I have written the code that will split the data and read it using DataSet used in aima-python.*

I have taken the data set of Wine Quality, it is a small data set with 4898 instances and 12 attributes, the data is discrete, the last column tells the quality of the wine, the quality is checked from 0-10 range. *Other features are fixed-acidity volatile-acidity, citric-acid, residual-sugar, chlorides, free-sulphur-dioxide, total-sulphur-dioxide, density, pH, sulphates, alcohol, quality.*

This dataset best suits for classification. When a new wine feature is given, it will classify to which quality the wine belongs. Models can be easily build using Naïve Bayes classifier when the dataset is small or medium size. Many a times this will outperform even highly sophisticated classification methods. This classifier is used in different streams of analysis like sentiment analysis, diagnosis of disease, spam email classifier for this reason.

**Bayes Theorem:** It states that, for 2 events X and Y, if we know the conditional probability of Y given X and the probability of Y, then it is possible to calculate the probability of Y given X

**Explanation of Bayes Theorem for the above definition:**

$$P(\frac{Model}{Data}) = \frac{\mathrm{P}\left(\frac{\mathrm{Data}}{\mathrm{Model}}\right)\ P(Model)}{\mathrm{P(Data)}}$$

**P(Model/Data) → is Posterior Probability** **P(Data/Model) → is likelihood**

**P(Model) → is Prior Probability** **P(Data) → Model Evidence**

Let me consider the dataset example with some changes in their values for demonstration; Let us consider there are 125 instances with wine-quality_5 has 100 and there are 25 instances with wine-quality_6. Here we can see that class of wine_quality_5 is 4 times more then the class of wine quality 6. So, with new feature during classification is four times likely to have membership of wine quality 5 then 6. In Bayesian Analysis this belief is known as **The Prior Probability**. So, in the below calculation I have taken prior_probability with wine_quality_5 and wine_quality_6, I have guessed it has 0.8 and 0.2 which is my best guess, but the probability could also be different also.

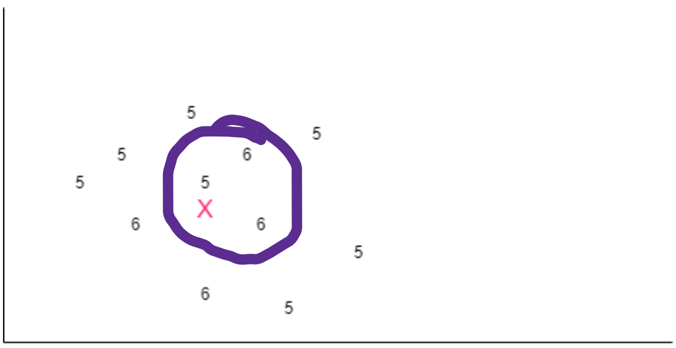
$$Prior\ Probability\ of\ \mathrm{wine\_Quality\_5} = \frac{\mathrm{class\ of\ wine_{quality_5}}}{\mathrm{wine_{Quality_5}} + \mathrm{wine_{Quality_6}}} = \frac{100}{125} = \frac{4}{5} = 0.8$$

$$Prior\ Probability\ of\ \mathrm{wine\_Quality\_6} = \frac{\mathrm{class\ of\ wine\_quality\_6}}{\mathrm{wine\_Quality\_5} + \mathrm{wine\_Quality\_6}} = \frac{25}{125} = \frac{1}{5} = 0.2$$

**Here is the prior probability of actual data set**

```
------------------- Printing Prior Probability Table-------------
Printing Prior Probability Table = {0: 0.0, 1: 0.0, 2: 0.0, 3: 0.004083299305839118, 4: 0.03327888934258881, 5: 0.2974683544303797, 6: 0.44875459371171905, 7: 0.17966516945692118,
 8: 0.03572886892609228, 9: 0.0010208248264597796, 10: 0.0}
```

**Probability of Likelihood:** This will tell what probability will be observing with data when a particular model is given as input. This is computed by picking a particular value from the model parameter and then we will compute how likely it would be for the observed data as output of experiment. Let us explain using diagram




Even though the prior probability indicate that 'X' may belong to Wine_Quality_5, the likelihood indicates otherwise; that is, it will indicate the new Class membership of 'X' is Wine_Quality_6 (such that wine_quality_6 must be more inside the circle)

So, in my code basically I'm calculating of each quality of wine with the total number of data instance.

```
probability_prior_table[0] = len(wineQualityDataArray[wineQualityDataArray[:, -1] == 0])
/ wineQualityDataArray.shape[0]
```

**Model Evidence:** this will tell what would the probability of observing data irrespective of the input model parameter. This p(data) is important because the value that we get will be in normalising constant. This normalising constant will make sure the resulting distribution is true probability distribution and the sum of all outcomes if the event is equal to 1.

**In Bayesian Analysis, the final classification is taken by combining all this information to get posterior probability.**

**Posterior Probability:** **P(Model/Data)** this will tell what we think the probability of the model being correct after doing the experiment, given the data that resulted.

**Probabilistic Inference:** Using the probabilistic model which describes the statistical problems which are in terms of probability theory and probability distribution.

## 1.3.2 Naïve Bayes Learner

**Write a clear and concise description of the Naive Bayes Classifier considering the headings outlined: probability, conditional independence, and Bayes theorem**

Initially the bayes theorem, considered all the variables as dependent on one another. This conditional dependency leads to complications in real world. Hence, researches took place to resolve the issue and the resolution was to consider the variables independent of one another. Therefore, the concept of conditional independence came into usage. This refined model of bayes theorem with conditional independence is called naive bayes classifier.

There are different variants in Naïve Bayes Classifier 1) Gaussian Naive Bayes: which works best with continuous types of data 2) Bernoulli Naïve Bayes: Here features are vectorised in binary 3) Multinomial Naïve Bayes: This is non-Binary version of Bernoulli Naïve Bayes 4) Complement Naïve Bayes: this will implement Complement Naïve Bayes algorithm, this is suited for imbalanced data set, this will use statistical complement of each class to compute the model weight.

I have used Multinomial Naïve Bayes Variant because the features of the data are not dependent on one another i.e., we have discrete data, like the citric acid that is present in the data (or wine) is not at all depend on the chlorides i.e., increasing the value of any column will not affect the value of another column.

Accuracy of the Model:
Accuracy = 72.02939975500205
Confusion Matrix which I have got for my dataset

Actual Values

| | | Positive Quality 3 | Positive Quality 4 | Positive Quality 5 | Positive Quality 6 | Positive Quality 7 | Positive Quality 8 | Positive Quality 9 |
|---|---|---|---|---|---|---|---|---|
| Predicted Values | Positive Quality 3 | 9 | 1 | 4 | 5 | 1 | 0 | 0 |
| | Positive Quality 4 | 0 | 54 | 49 | 53 | 7 | 0 | 0 |
| | Positive Quality 5 | 0 | 8 | 1069 | 327 | 51 | 2 | 0 |
| | Positive Quality 6 | 1 | 6 | 293 | 1691 | 202 | 5 | 0 |
| | Positive Quality 7 | 0 | 1 | 38 | 214 | 626 | 1 | 0 |
| | Positive Quality 8 | 0 | 0 | 2 | 57 | 37 | 79 | 0 |
| | Positive Quality 9 | 0 | 0 | 0 | 3 | 2 | 0 | 0 |

Here I have put a simple confusion table to show the prediction with the actual value, The Yellow colour background will tell that the values that I have predicted and the actual value matching number, ie., lets take quality_5, total of 1069 values are matching where the blue colour indicate the incorrect prediction of the model, i.e., let us take quality_3 in this 5 models are predicted as Quality_6, 1 as Quality_4, 4 and Quality_5 and so on, so finally the accuracy of the model is 72% that is can prediction 72 of new given features accurately and rest 28 will be classified in wrong prediction.

## References:


1. https://www.sciencedirect.com/science/article/pii/S0304380006006089?casa_token=O38qyMObR3YAAAAA:2MH0jsvjBq49JzAKUr_FJcNaejiptHTKeFHhli1_zilvV2Vcz-CtAnHPWazH4KFo2bArArA#!