ASSIGNMENT II

Artificial Intelligence

Submitted by:

Nitish Arya

2017CSC1061

On May 11, 2020

To:

P.D Sir

Probabilistic Model for COVID-19

- **Sample Space:** The set of all possible worlds, mutually exclusive and exhaustive.

  Our Sample Space is as such:

  - A single human with its characteristics (Age, Height, etc…) lives in a normal realistic world with this new COVID-19 with other people.
  - Our space is localized to a certain region where the person lives but is affected by global decisions.
  - Primary focus is on medical side of COVID-19.
  - A possible world is diagnosing a human with COVID-19 and related scenarios.

    - Ω: Sample Space, ω: Elements of the space (a particular possible world).
    - $0 \leq P(\omega) \leq 1$ for every ω and $\Sigma(\omega \in \Omega) \, P(\omega) = 1$.

- **Notations:**

  - **Textual**
    1. <u>First letter uppercase :- Random Variables. They have specific domain.</u>
       a. All Random Variables in our model: **Breath, Touch, Symptom, Test, Solution, Incubation/Treatment.**
       b. Domains:

       Breath ∈ {Inhaled near a corona +ive person, Inhaled from a common AC}.

       Touch ∈ {Touched a corona +ive person, Touched an item with virus}

       (Touched means touched your face, nose, eyes or ears).

       Symptom ∈ {Fever, Dry Cough, Runny Nose, Nausea}.

       Test ∈ {(all different kinds of Test for COVID-19)count = 10(say)}.

       Solution ∈ {Lockdown, Hydrochloroquine, Sanitization}.

       Incubation/Treatment ∈ {Incubation, Allopathy, Ayurvedic, Homeopathy,

       Yunani, Yoga, No Treatment}.

2. All lowercase :- Boolean Variables. Either True or False.

      a. All Boolean Variables in our model: *asymptomatic, corona positive, recovery, reinfection, death.*

o **Figurative**

  1. **Boolean Variables:**

  2. **Random Variables:**

  3. **Clusters(Similar Variables):**

  4. **Dependencies:**

- **How to Inference from our model:**

  ➢ Each Node(Random/Boolean Variable) has a **Conditional Probability Table(CPT)** <u>containing dependent probabilities of its parents</u>.

    For ex:

    For the Node **Incubation/Treatment** we have the following **CPT**:

x	P(**Incubation/Treatment = x** |*corona positive)*	P(**Incubation/Treatment = x** |*¬corona positive)*
Incubation		
Allopathy		
Ayurvedic		
Homeopathy		
Yunani		
Yoga		
No Treatment		

    <u>There will be values in the blank spaces.(Data wasn't available, hence only one **CPT** is shown.)</u>

➢ An example to inference a required probability from our model:

Q. What is the probability that a patient having Fever is corona positive?

1. We have to find P(***corona positive | Symptom =*** Fever**).**
2. This value is directly fetched from the **CPT** of Node ***corona positive*** having all other variables(parents of Node ***corona positive*** except Fever) as False.

Q. What is the probability that a patient dies if he is corona positive?

1. We have to find P(***death | corona positive***).
2. This probability can't be directly fetched from the **CPT** of Node ***death*** as Node ***corona positive*** is not a parent of Node ***death.***
   Hence we first formulate,

   P(***death | corona positive***) = α P(***death, corona positive***)

   = α $\sum$(hidden variables) P(***death, corona positive*** and all other hidden variables).

   i.e,

   α $\sum$(hidden variables) P(***death, corona positive,*** *all Random Variables with each value from its domain, all Boolean Variables with each True and False*).

Consider a small case where hidden variables are only ***asymptomatic*** and **Test.**
Now the probability would be,

P(***death | corona positive***) = α $\sum$(**Test**) $\sum$(***asymptomatic***) P(***death, corona positive, asymptomatic,*** **Test)**

 = α $\sum$(**Test**) $\sum$(***asymptomatic***) P(***death)*** P(***corona positive| asymptomatic,*** **Test)**

 = α P(***death)*** $\sum$(**Test**) $\sum$(***asymptomatic***) P(***corona positive| asymptomatic,*** **Test)**

 = α P(***death)*** $\sum$(**Test**) P(***corona positive| asymptomatic*** =True***,*** **Test)** P(***corona positive| asymptomatic*** =False***,*** **Test)**

 = α P(***death)*** $\sum$(**Test = {x,y,…})** P(***corona positive| asymptomatic*** =True***,*** **Test)** P(***corona positive| asymptomatic*** =False***,*** **Test)**

Now we do the above summation for 10 different values of Test, find the respective probabilities from the respective **CPT's** and multiply all of them to get our final P(***death | corona positive***).

- **Key points:**

a) We used Bayesian Network and its properties rather than Joint

   Probability Distribution table because we have a total of 11 Nodes with

   28(Random Variable Domain count)+5*2(Boolean Variable Domain

   count). If we build a Joint Probability Distribution table for 11 'Boolean

   variables', we would have 2^11 = **2048 rows**. Whereas in Bayesian

   Network we have 2^7 + 4*2 = 128 + 8 = **136 rows** of **CPT's**.

b) The Bayesian Network is acyclic. Any dependent probability can be

   calculated by the formulating algorithm and using the **CPT's**. Hence it

   can also give **posterior probabilities.**

c) If the Nodes are ordered numerically from 1 to 11, we can calculate

for any variable X_i (i = 1,2,…11) in the network

$P(X_i \mid X_{i-1,\ldots,}X_1) = P(X_i \mid Parents(X_i))$

Given $Parents(X_i)$ is a proper subset of $\{X_{i-1},\ldots,X_1\}$.

- **Limitations:**

  ➤ Since **CPT** of a Node requires dependent probabilities of its Parent Nodes, <u>our model needs to collect various data to fill in the the **CPT's** of each Nodes.</u>

  ➤ The formulation for a query requires summing over all values of hidden variables which might grow large when Random Variables have multiple domain values. Hence, we should use an algorithm and probable a computer for calculating such probabilities.

  ➤ This model is very limited in scope as the variables and their domains both will be much large in number. Also its complexity is very low and dependencies might be incorrect.
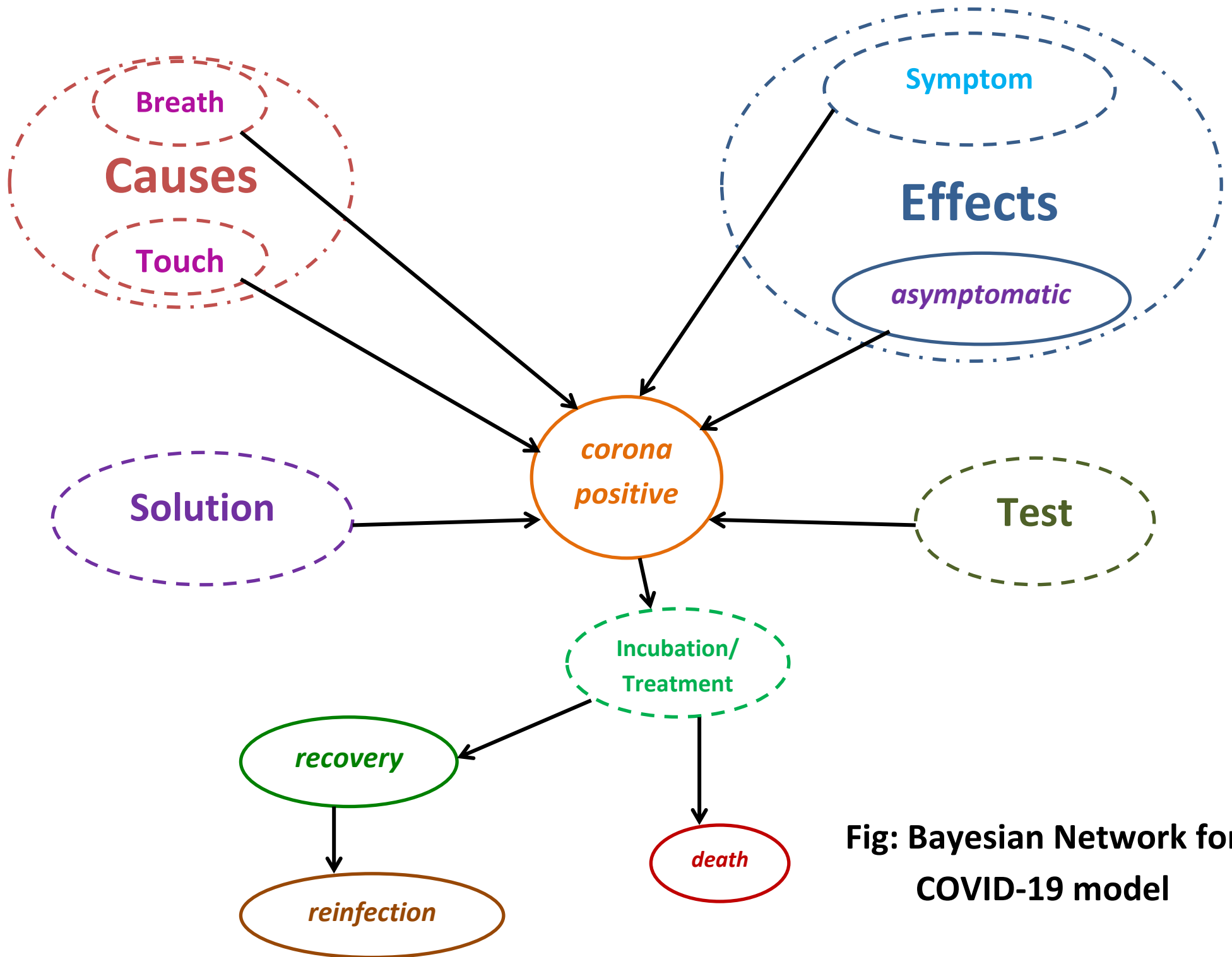
Fig: Bayesian Network for COVID-19 model

This is the end.

Pardon for any spelling mistakes.