Structured Statistical Models of Inductive Reasoning: An Application

Akash Pahariya & Vivek Agarwal

November 13, 2010

Abstract

Common concepts such as animals, cities et cetera confer to highly structured arrangements in their respective spaces. We as human beings continuously construct and improve these structures based on our daily activities and experiences. The construction of these structures can be explained by the probabilistic inference about the relations and similarities between different entities of a domain.

In this project we have tried to form an efficient organizational structure for some of the diseases based the extent to which they show similar symptoms. We have attempted to understand how a doctor decides on what disease an individual is suffering form based on the already existing mental structures. In this attempt we also propose a simple model which could be further used to create **Doctor-Bots** with artificial Intelligence.

We have used the model of structure learning specifically 'taxonomic model', described in the paper ("Structured Statistical Models of Inductive Reasoning', Kemp & Tenenbaum).

The code for structure learning is available on:
http://www.psy.cmu.edu/ ckemp/code/structstatmodels.html
The code utilized for this project in specific with all relevant changes and corrections is available on:
http://home.iitk.ac.in/ agvivek/se367/project/code.zip

1 Introduction

"Going beyond given information", there are two fundamental ways in which this can be achieved viz. **Deductively** or **Inductively**. In Deductive inferences, we draw conclusions that were not explicitly stated but can be derived from the given information. Whilst, in Inductive inferences we conclude in a systematic manner which is one of the most likely outcomes in the given context.

In Inductive Inference, the authors[1] view can be developed as a two step process; a combination of Statistical Inference & Structured Learning. Both of these models have previously been implemented individually, but their combination is an intuitive extension, since they tend to complement each other. Structured Learning is insensitive to noise, however brings out the underlying core efficiently. On the other hand Statistical Inference is sensitive to noise however fails to bring out the underlying essence.

The approach can thus be summarized as:

(i) Generation of a prior structure which captures the relevant relationships between the categories in the domain;
(ii) Identification of a stochastic process that captures the knowledge about how these properties tend to distribute themselves.

For Example: A natural representation for animals in the biological context(evolutionary) is a hierarchal tree structure where they are located at the leaves of the tree. Also the stochastic property is that of "Diffusion" as it ensures the symmetry of two way relationship between different animals is maintained. It also acknowledges the fact that the farther the animals lesser the similarities between them.

2 Kemp & Tenenbaum, 2009 (Appendix C): A Summary

A tree can be represented as a pair (ST, SL) where ST refers to the topology of the tree and SL is a vector specifying the lengths of its branches.

Our task is to find ST & SL in such a way so that we can maximize

$p(ST\ ,\ SL \mid D,\ T,\ F) = p(D \mid ST\ ,\ SL,\ T) * p(ST,\ SL \mid F)$

where,
T is the diffusion process,
F is the unknown structure,
S is the structures(tree) &
D is the data set

To achieve this we first identify the best topology ST and then search for the best set of branch lengths.

(i) We employ a *greedy search* to identify the topology which includes first starting with the most basic topology in which every entity is connected through an edge to single node. Every time after witnessing some training data we try to

split one of the current nodes. And after each split we rearrange some of the nodes in order to improve the posterior *p(ST | D, F, T)*. This process continues till we get a best structure.

(ii) After the above step, a *gradient based search* is used to find the SL, which maximize *p(SL | ST, D, T)*

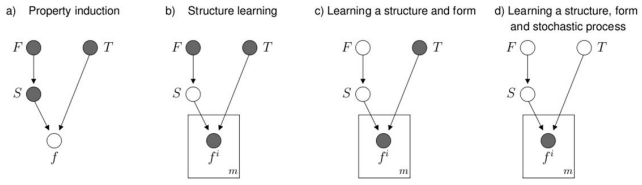


Figure One: Courtesy, Kemp & Tenenmaun, 2009

3 Project Description & Challenges

The motivation behind choosing this project has been to gain a deeper insight into the "Structured Statistical Model", which offers a comprehensive blend of the Statistical Approach along with the Structured Approach.

As a part of the project, we have attempted to understand & recreate the mental structure of how a doctor thinks, and based on all inputs(symptoms) conclude what the problem(disease) is. This has also opened a way to creating **Doctor-Bot** with **artificial intelligence** to diagnose & recommend corrective action.

To achieve this we have employed the taxonomic model; since this best captures the underlying relationship between diseases.

4 iDoctor

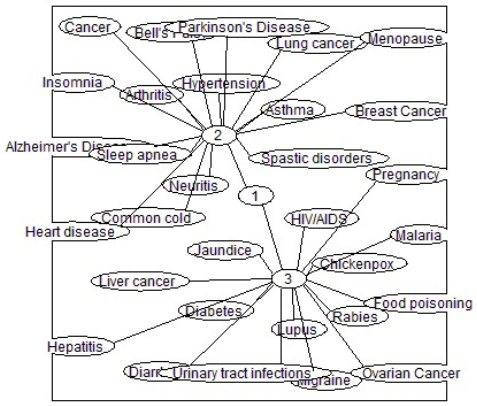
iDoctor works out of a basic set of 31 disease-81 symptom database. In its current format the user inputs his/her symptoms on a scale of 0 to 1. After Execution of the model a tree-leaf structure is obtained. By locating the disease in a particular branch; one can conclude the kind of disease one might be infected with.

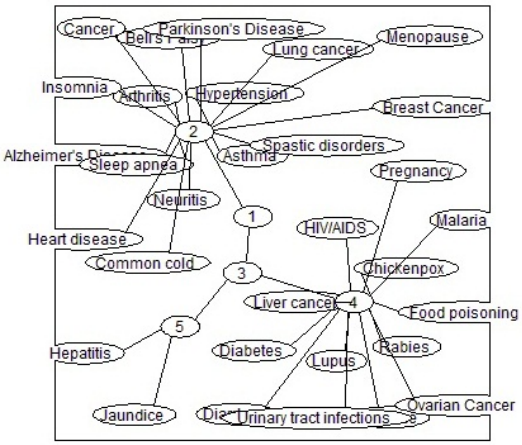
We have obtained the "Master Data Set" based on the inputs of three doctors. Each of whose data set is appended in **Appendix A**. All three of the data sets where then combined to create a *"graded"* feature set. The model then uses this as an input to try & understand how the diseases may be related & plots a graph.

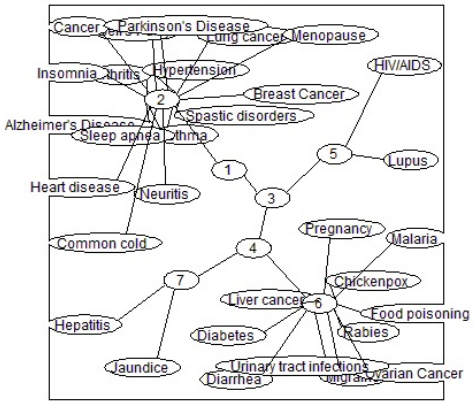
To diagnose a particular disease we request an individual to input symptoms on a scale of 0 to 1 into the dataset. The model is run again, the location of the Disease to be found is located on a branch. It is then concluded that the disease is most likely to be like the members of its branch. A close examination with the original tree helps concluding the disease.

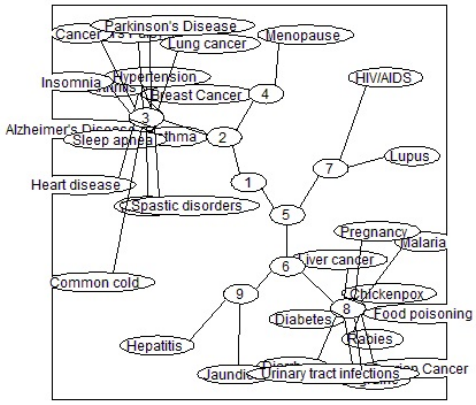
5 Results

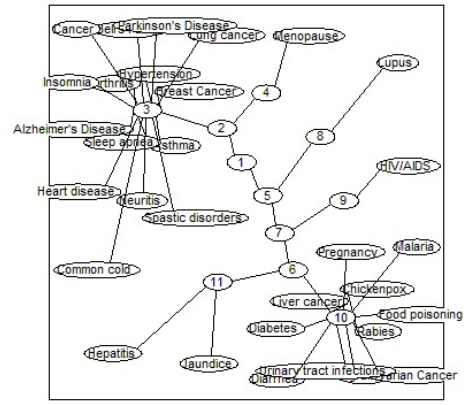
Initial Dataset

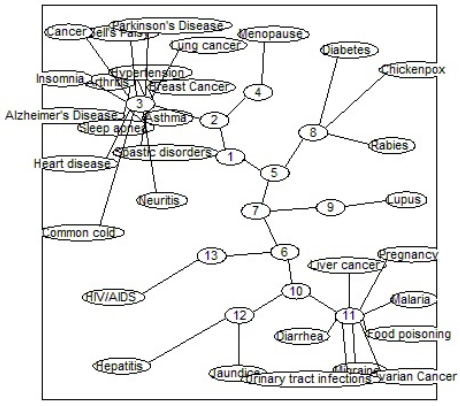


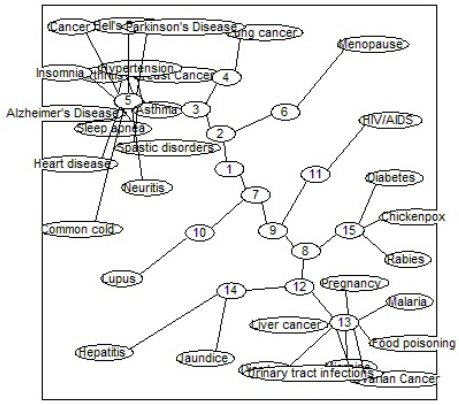


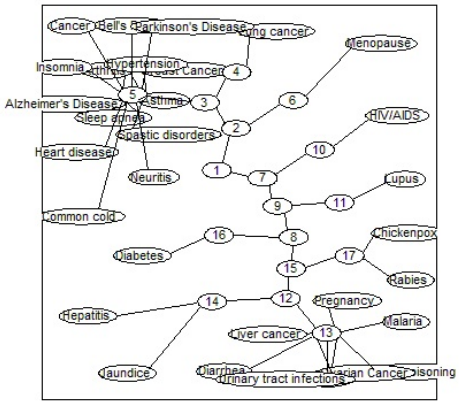

Cancer
Parkinson's Disease
Lung cancer
Menopause
Insomnia
Hypertension
HIV/AIDS
Breast Cancer
Spastic disorders
Alzheimer's Disease
Sleep apnea
Asthma
Lupus
Heart disease
Neuritis
Pregnancy
Malaria
Common cold
Chickenpox
Liver cancer
Food poisoning
Hepatitis
Diabetes
Rabies
Jaundice
Urinary tract infections
Diarrhea
Ovarian Cancer

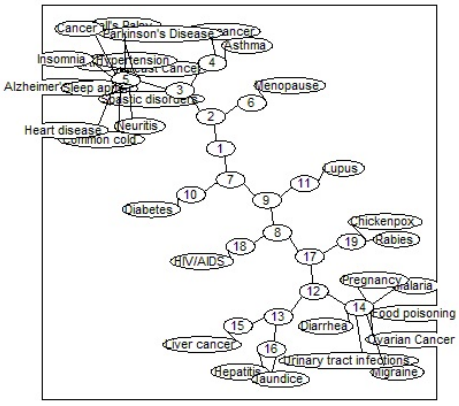

Parkinson's Disease
Cancer
Lung cancer
Menopause
Hypertension
Insomnia
Breast Cancer
HIV/AIDS
Alzheimer's Disease
Sleep apnea
Asthma
Lupus
Heart disease
Spastic disorders
Pregnancy
Malaria
Liver cancer
Common cold
Chickenpox
Food poisoning
Diabetes
Rabies
Hepatitis
Jaundice
Urinary tract infections
Cancer

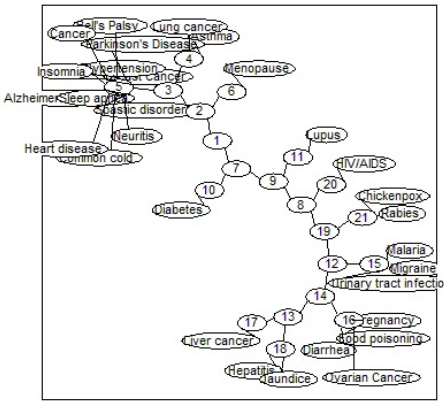

Parkinson's Disease
Cancer
Lung cancer
Menopause
Lupus
Hypertension
Insomnia
Breast Cancer
HIV/AIDS
Alzheimer's Disease
Sleep apnea
Asthma
Heart disease
Neuritis
Spastic disorders
Pregnancy
Malaria
Common cold
Chickenpox
Liver cancer
Food poisoning
Diabetes
Rabies
Hepatitis
Jaundice
Urinary tract infections
Ovarian Cancer

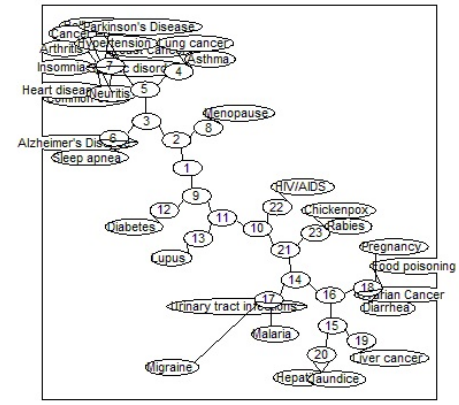

Cancer
Parkinson's Disease
Menopause
Lung cancer
Diabetes
Chickenpox
Insomnia
Hypertension
Breast Cancer
Alzheimer's Disease
Asthma
Sleep apnea
Rabies
Spastic disorders
Heart disease
Neuritis
Lupus
Common cold
Pregnancy
Liver cancer
HIV/AIDS
Malaria
Diarrhea
Food poisoning
Hepatitis
Jaundice
Urinary tract infections
Ovarian Cancer

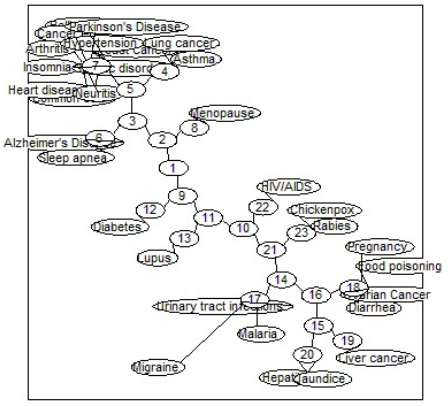

Cancer
Parkinson's Disease
Lung cancer
Menopause
Insomnia
Hypertension
Breast Cancer
Alzheimer's Disease
Asthma
Sleep apnea
Spastic disorders
Heart disease
Neuritis
HIV/AIDS
Diabetes
Common cold
Chickenpox
Rabies
Lupus
Pregnancy
Malaria
Liver cancer
Food poisoning
Hepatitis
Jaundice
Urinary tract infections
Ovarian Cancer

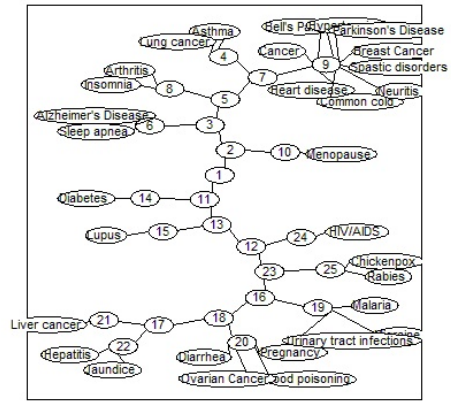

Cancer
Parkinson's Disease
Lung cancer
Menopause
Insomnia
Hypertension
Breast Cancer
Alzheimer's Disease
Asthma
Sleep apnea
Spastic disorders
Heart disease
Neuritis
HIV/AIDS
Lupus
Common cold
Diabetes
Chickenpox
Rabies
Pregnancy
Hepatitis
Liver cancer
Malaria
Jaundice
Diarrhea
Urinary tract infections
Ovarian Cancer

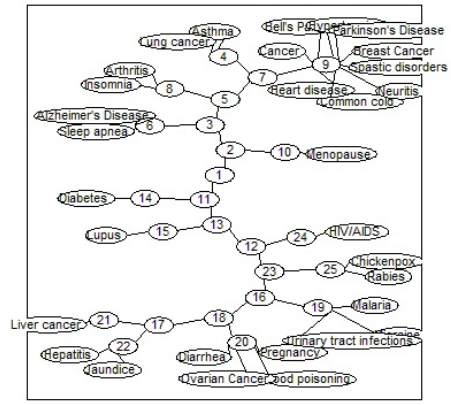

Asthma
Insomnia
Alzheimer
Menopause
Heart disease
Neuritis
Lupus
Diabetes
Chickenpox
Rabies
HIV/AIDS
Pregnancy
Malaria
Food poisoning
Diarrhea
Ovarian Cancer
Liver cancer
Urinary tract infections
Migraine
Hepatitis
Jaundice

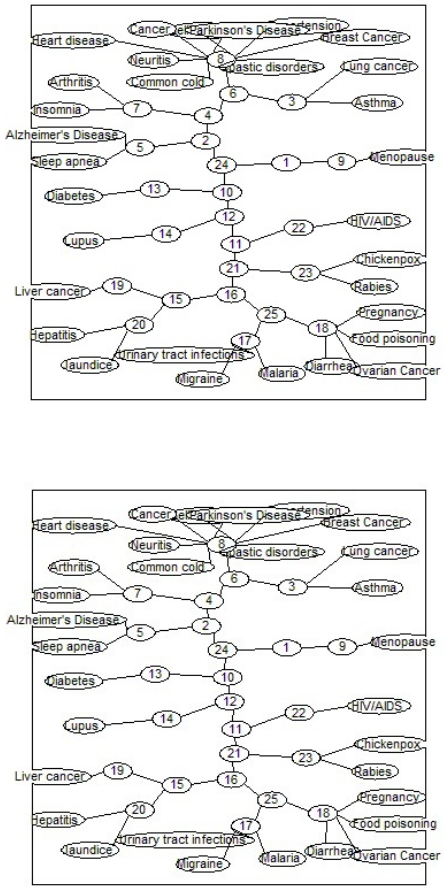

Lung cancer
Asthma
Menopause
Alzheimer
Neuritis
Heart disease
Lupus
HIV/AIDS
Diabetes
Chickenpox
Rabies
Malaria
Migraine
Urinary tract infections
Pregnancy
Liver cancer
Food poisoning
Diarrhea
Hepatitis
Jaundice
Ovarian Cancer


Parkinson's Disease
Lung cancer
Asthma
Heart disease
Menopause
Alzheimer's
Sleep apnea
HIV/AIDS
Diabetes
Chickenpox
Rabies
Lupus
Pregnancy
Food poisoning
Ovarian Cancer
Diarrhea
Malaria
Liver cancer
Migraine


Parkinson's Disease
Asthma
Insomnia
Heart disease
Neuritis
Menopause
Alzheimer's Dis
Sleep apnea
HIV/AIDS
Chickenpox
Rabies
Diabetes
Pregnancy
Lupus
Food poisoning
Ovarian Cancer
Diarrhea
Malaria
Liver cancer
Migraine
Hepat
Jaundice


Asthma
Parkinson's Disease
Cancer
Breast Cancer
Spastic disorders
Arthritis
Insomnia
Heart disease
Neuritis
Common cold
Alzheimer's Disease
Sleep apnea
Menopause
Diabetes
Lupus
HIV/AIDS
Chickenpox
Rabies
Malaria
Liver cancer
Urinary tract infections
Pregnancy
Hepatitis
Jaundice
Diarrhea
Ovarian Cancer
Food poisoning

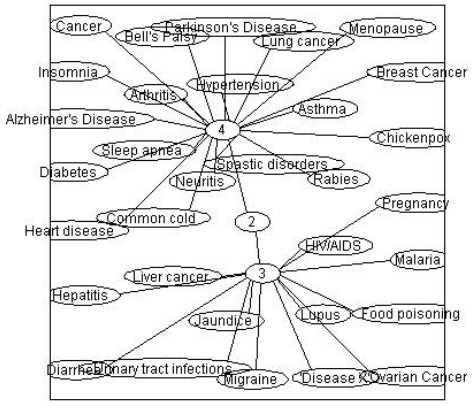

Asthma
Parkinson's Disease
Cancer
Breast Cancer
Spastic disorders
Arthritis
Insomnia
Heart disease
Neuritis
Common cold
Alzheimer's Disease
Sleep apnea
Menopause
Diabetes
Lupus
HIV/AIDS
Chickenpox
Rabies
Malaria
Liver cancer
Urinary tract infections
Pregnancy
Hepatitis
Jaundice
Diarrhea
Ovarian Cancer
Food poisoning

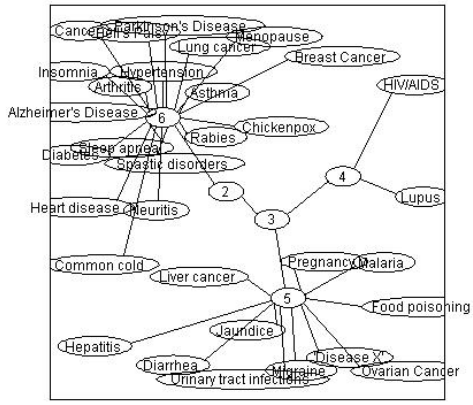


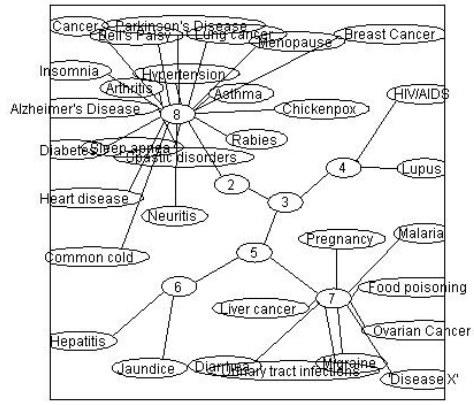
For Detection

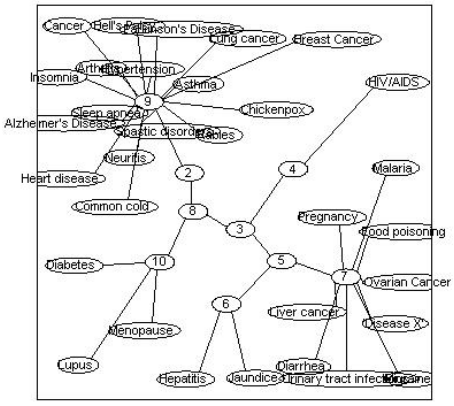
Disease X: had the following Symptoms:

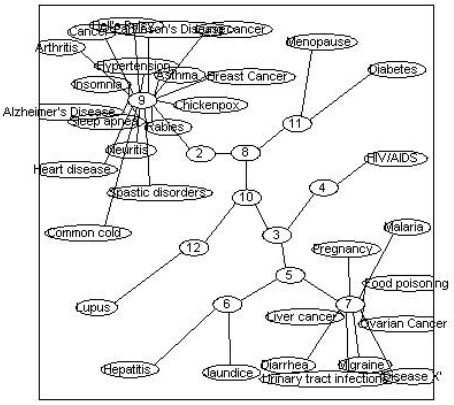
Abdominal Symptoms (Bloating, Pain, Cramps, Discomfort), Anemia, Dark Colored, Urine, Diarrhea, Dizziness, Facial Symptoms (Expression Changes, Flushing, Muscle Spasm, Pain, Swelling, Tingling), Knee Symptoms, Loose stool, Muscle Syptoms (Stiffness, Weakness, Spasms), Palpitations, Sexual problems, Shortness of breath, Skin Symptoms, Swelling, Vomiting, Wheezing, Yellow Skin

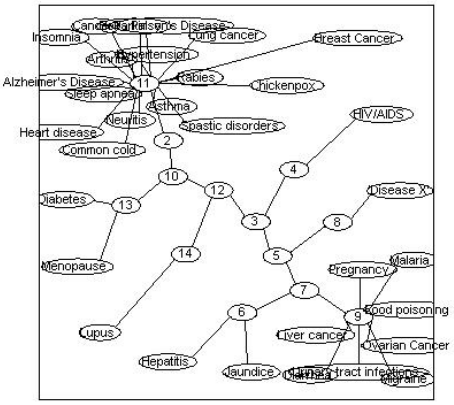

Cancer
Parkinson's Disease
Menopause
Bell's Palsy
Lung cancer
Insomnia
Breast Cancer
Hypertension
Arthritis
Asthma
Alzheimer's Disease
4
Chickenpox
Sleep apnea
Spastic disorders
Diabetes
Rabies
Neuritis
Pregnancy
Common cold
2
Heart disease
HIV/AIDS
Malaria
3
Liver cancer
Hepatitis
Lupus
Food poisoning
Jaundice
Diarrhea
Urinary tract infections
Migraine
Disease X'
Ovarian Cancer

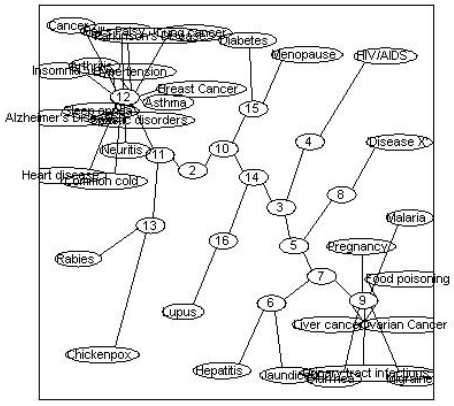

Cancer
Parkinson's Disease
Bell's Palsy
Menopause
Lung cancer
Breast Cancer
Insomnia
Hypertension
HIV/AIDS
Arthritis
Asthma
Alzheimer's Disease
6
Chickenpox
Rabies
Diabetes
Sleep apnea
Spastic disorders
4
Lupus
2
Heart disease
Neuritis
3
Common cold
Pregnancy
Malaria
Liver cancer
5
Food poisoning
Jaundice
Hepatitis
Disease X'
Diarrhea
Migraine
Ovarian Cancer
Urinary tract infections

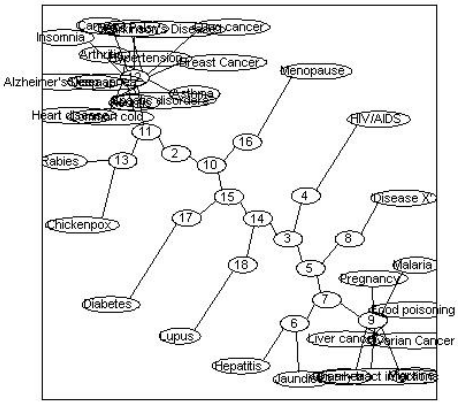

Cancer
Parkinson's Disease
Bell's Palsy
Lung cancer
Menopause
Breast Cancer
Insomnia
Hypertension
Arthritis
Asthma
HIV/AIDS
Alzheimer's Disease
8
Chickenpox
Rabies
Diabetes
Sleep apnea
Spastic disorders
4
Lupus
2
Heart disease
3
Neuritis
Pregnancy
Malaria
Common cold
5
6
7
Food poisoning
Liver cancer
Ovarian Cancer
Hepatitis
Jaundice
Diarrhea
Urinary tract infections
Migraine
Disease X'

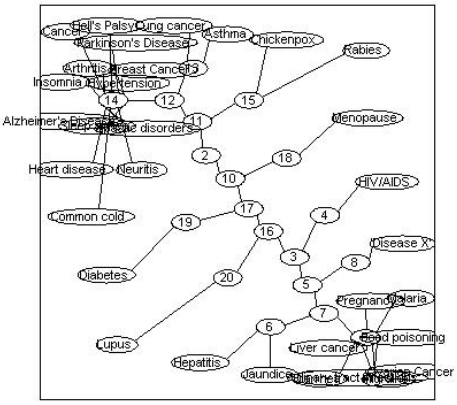

Cancer
Cancer
Parkinson's Disease
Lung cancer
Breast Cancer
Arthritis
Hypertension
Insomnia
Asthma
HIV/AIDS
9
Chickenpox
Sleep apnea
Alzheimer's Disease
Spastic disorders
Rabies
Neuritis
Heart disease
2
4
Malaria
Common cold
8
Pregnancy
3
Food poisoning
Diabetes
10
5
7
Ovarian Cancer
6
Liver cancer
Disease X
Menopause
Lupus
Hepatitis
Jaundice
Diarrhea
Urinary tract infection

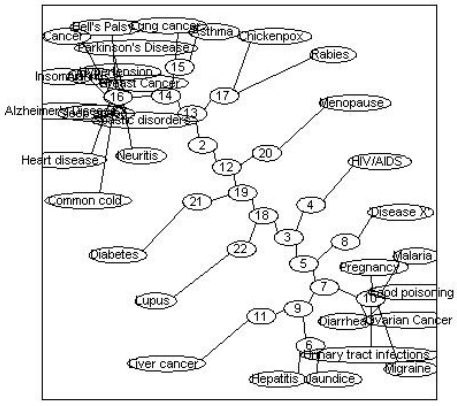

Arthritis
Menopause
Hypertension
Diabetes
Asthma
Breast Cancer
Insomnia
9
Chickenpox
Alzheimer's Disease
11
Sleep apnea
Rabies
Neuritis
2
8
HIV/AIDS
Heart disease
Spastic disorders
10
4
Common cold
3
Malaria
12
Pregnancy
5
Food poisoning
Lupus
6
7
Liver cancer
Ovarian Cancer
Hepatitis
Jaundice
Diarrhea
Migraine
Urinary tract infection
Disease X

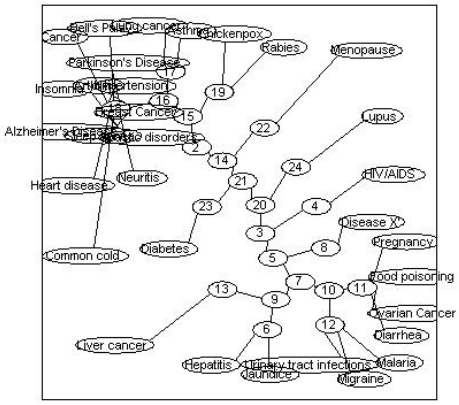

Insomnia
Lung cancer
Breast Cancer
Arthritis
Hypertension
Alzheimer's Disease
11
Rabies
Chickenpox
Sleep apnea
Asthma
HIV/AIDS
Neuritis
Spastic disorders
Heart disease
2
Common cold
4
10
12
Disease X
Diabetes
13
3
8
14
5
Menopause
Pregnancy
Malaria
7
6
9
Food poisoning
Lupus
Liver cancer
Ovarian Cancer
Hepatitis
Jaundice
Urinary tract infection
Migraine

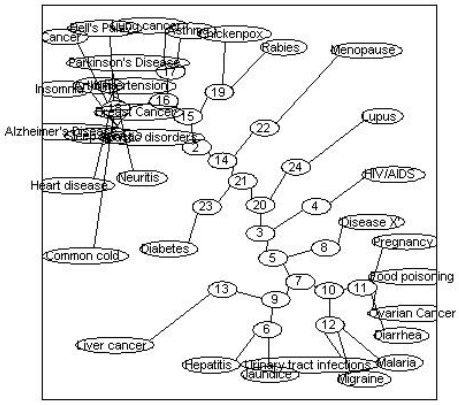

Cancer
Diabetes
Menopause
HIV/AIDS
Insomnia
Breast Cancer
Asthma
Disease X
Neuritis
Heart disease
Common cold
Malaria
Rabies
Pregnancy
Food poisoning
Lupus
Liver cancer
Ovarian Cancer
Chickenpox
Hepatitis

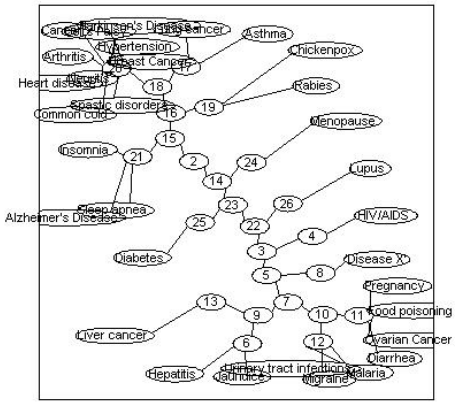

Insomnia
Breast Cancer
Menopause
HIV/AIDS
Disease X
Malaria
Pregnancy
Food poisoning
Liver cancer
Ovarian Cancer
Rabies
Chickenpox
Diabetes
Lupus
Hepatitis

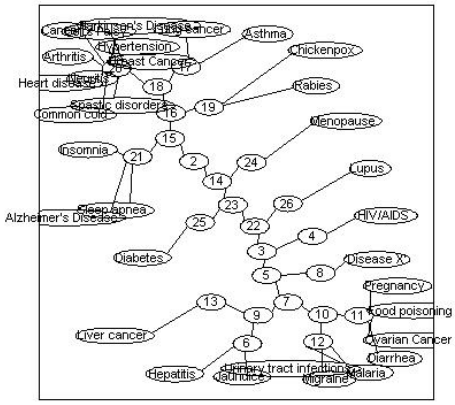

Cancer
Bell's Palsy
Lung cancer
Asthma
Chickenpox
Rabies
Parkinson's Disease
Arthritis
Breast Cancer
Insomnia
Hypertension
Menopause
Heart disease
Neuritis
HIV/AIDS
Common cold
Disease X
Diabetes
Malaria
Pregnancy
Food poisoning
Lupus
Liver cancer
Hepatitis

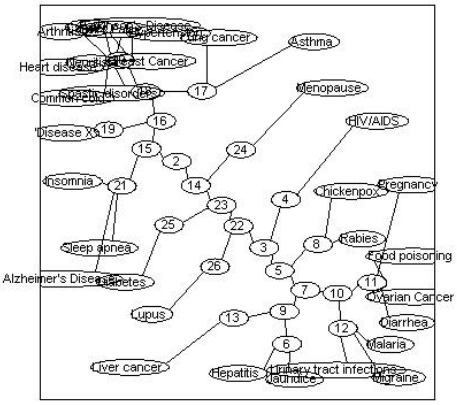

Bell's Palsy
Lung cancer
Asthma
Chickenpox
Cancer
Parkinson's Disease
Rabies
Menopause
Heart disease
Neuritis
HIV/AIDS
Common cold
Disease X
Diabetes
Malaria
Pregnancy
Lupus
Food poisoning
Diarrhea
Ovarian Cancer
Liver cancer
Urinary tract infections
Migraine
Hepatitis
Jaundice

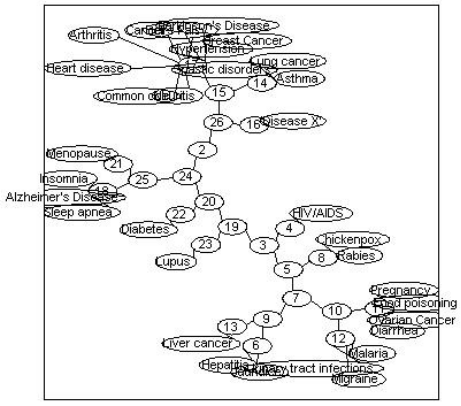

Cancer
Chickenpox
Rabies
Menopause
Parkinson's Disease
Lupus
Heart disease
Neuritis
HIV/AIDS
Disease X
Pregnancy
Common cold
Diabetes
Food poisoning
Ovarian Cancer
Diarrhea
Liver cancer
Hepatitis
Malaria
Jaundice
Migraine


Cancer
Chickenpox
Rabies
Menopause
Parkinson's Disease
Lupus
Heart disease
Neuritis
HIV/AIDS
Disease X
Pregnancy
Common cold
Diabetes
Food poisoning
Ovarian Cancer
Diarrhea
Liver cancer
Hepatitis
Malaria
Jaundice
Migraine


Asthma
Chickenpox
Arthritis
Heart disease
Rabies
Spastic disorders
Common Cold
Menopause
Insomnia
Lupus
Sleep apnea
Alzheimer's Disease
HIV/AIDS
Diabetes
Disease X
Pregnancy
Food poisoning
Liver cancer
Ovarian Cancer
Diarrhea
Hepatitis
Malaria
Jaundice
Migraine


Asthma
Chickenpox
Arthritis
Heart disease
Rabies
Spastic disorders
Common Cold
Menopause
Insomnia
Lupus
Sleep apnea
Alzheimer's Disease
HIV/AIDS
Diabetes
Disease X
Pregnancy
Food poisoning
Liver cancer
Ovarian Cancer
Diarrhea
Hepatitis
Malaria
Jaundice
Migraine


Asthma
Heart disease
Menopause
HIV/AIDS
Disease X
Insomnia
Chickenpox
Pregnancy
Rabies
Sleep apnea
Food poisoning
Alzheimer's Disease
Diabetes
Ovarian Cancer
Lupus
Diarrhea
Malaria
Liver cancer
Hepatitis
Jaundice
Migraine



6 Inference

Based on the results the following was conclusions can been drawn:

(i) Most of the Diseases tend to pair up well. Some of them pair up based on their origin like Hepatitis, Jaundice. There are also pairs which are disconnected Pregnancy, Food Poisoning, Ovarian Cancer, Diarrhoea. Of these Pregnancy & Ovarian Cancer and Food Poisoning & Diarrhoea are connected in pairs but not as a whole. This can be attributed to the limitations of the dataset & can be bettered with more noise free dataset.

(ii) The method allows to detect a disease, and potentially can also be used to identify new diseases. In the above case, the disease is an outlier an can be only traced to closest neighbours for some hints. Which suggests that this is a new disease.

7 Future Work

In our modelling we have only been able to utilize the structured approach to the fullest, the statistical approach that should have been used to arrive at the working database has been used in a very limited fashion. With more **"organized"** inputs one can use the *"Bayesian Inference Engine"* to build the "master database". This will help us in exploring the potential of the model to the fullest. In the current version of *iDoctor* a very limited & rudimentary data set has been employed. Detail & correctness of this data set determines the accuracy of the model. One of the many & essential improvements would be to reorganize the symptoms in such a fashion that first a major branch of disease can be arrived at and then further detailed analysis can be carried out.

This will not only save computation time, but also help in looking at greater details. Very certainly a lot can also be done with the interface.

8 Conclusion

The "Structured Statistical Approach" has myriad applications & immense potential to unlock many underlying phenomena. In our humble attempt we have been able to model this approach in part by implementation of the structured approach in complete & the statistical approach in part we have gained a better understanding of the "Structured Statistical Approach".

9 Acknowledgement

(i) Dr. Amitabh Mukherjee
Department of Computer Science & Engineering,
Indian Institute of Technology, Kanpur

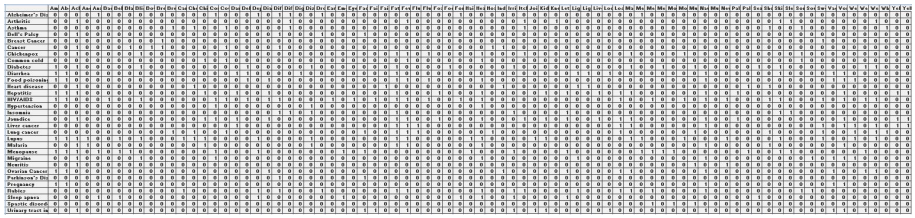
(ii) Dr. Divya Balasundaram

(iii) References:
(1) "Structured Statistical Models of Inductive Reasoning", Kemp & Tenenbaum (2009)

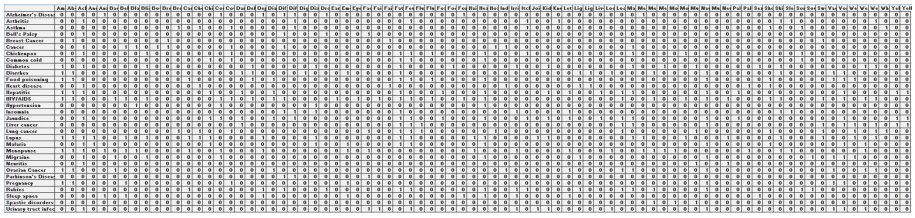
10 Appendix

10.1 Appendix A

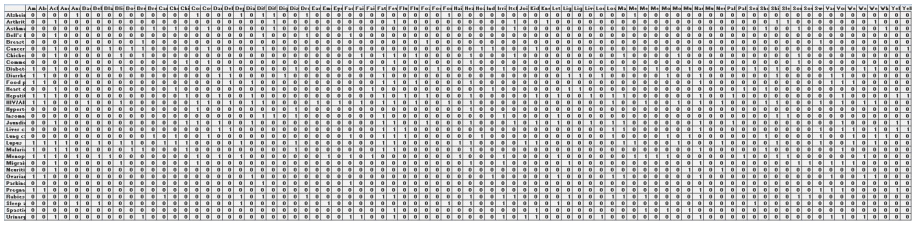
Doctor One



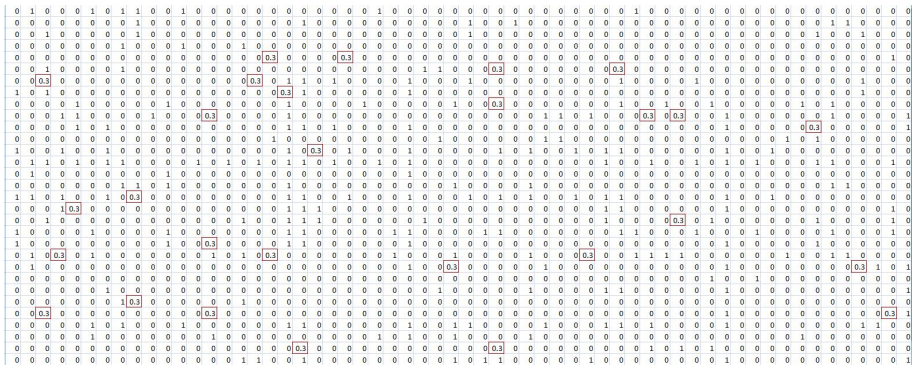
Doctor Two



Doctor Three



Final Data Set



10.2 Appendix B

List of Diseases:
Alzheimer's Disease
Arthritis
Asthma
Bell's Palsy
Breast Cancer
Cancer
Chickenpox
Common cold
Diabetes
Diarrhea
Food poisoning
Heart disease
Hepatitis
HIV/AIDS
Hypertension
Insomnia
Jaundice
Liver cancer
Lung cancer
Lupus
Malaria
Menopause
Migraine
Neuritis
Ovarian Cancer
Parkinson's Disease
Pregnancy
Rabies
Sleep apnea
Spastic disorders
Urinary tract infections

10.3 Appendix C

List of Symptoms: Amenorrhea (Absence of periods)
Abdominal Symptoms (Bloating, Pain, Cramps, Discomfort)
Aches
Anemia
Anxiety
Bacterial Infections
Behavioral Changes
Bladder Problems
Blisters

Bowel Symptoms
Breast Symptoms (Lump, Shape Change, Thickning, Skin Ridges)
Breathing difficulty
Cardiac Symptoms
Chest Ache
Chills
Confusion
Cough
Dark Colored Urine
Dehydration
Depression
Diarrhea
Difficult Swallowing
Difficulty Speaking
Digestive Symptoms
Dizziness
Drooling
Ear Symptoms
Emotional Symptoms
Eye Symptoms (Conjunctivitis)
Facial Symptoms (Expression Changes, Flushing, Muscle Spasm, Pain, Swelling, Tingling)
Failure to thrive
Fainting
Fatigue
Fever
Fluid Retention
Flu-like symptoms
Foamy Urine
Foot Symptoms (Numbness, Tingling)
Forgetfulness
Hair Loss
Headache
Hoarseness
Indigestion
Irritability
Itching
Joint Symptoms (Pain, redness, Stiffness, Swelling, Tenderness)
Kidney Symptoms (Failure, Infection, Inflammation)
Knee Symptoms
Lethargy
Light Sensitivity
Light-headedness
Liver failure
Loose stool
Loss of appetite

Malaise
Memory Problems
Mental Symptoms
Menstrual Symptoms (Changes, Absence, Bleeding)
Mood Symptoms (Change, Swing)
Mouth Symptoms (Blisters, Infections, Sores, White Spots)
Muscle Symptoms (Stiffness, Weakness, Spasms)
Nausea
Muscle Symptoms (Stiffness, Weakness, Spasms)
Neurologic Syptoms (Problems, Disorders)
Pale feces
Palpitations
Sexual problems
Shortness of breath
Skin Symptoms
Sleep Problems
Sore throat
Sound sensitivity
Swelling
Vague symptoms
Vomiting
Weakness
Weight Gain
Weight Loss
Wheezing
Yellow eyes
Yellow Skin