Bayesian Medical Expert Systems

Mert Cem Taşdemir

Boğaziçi University

CmpE 492: Graduation Project

Advisor: Assoc. Prof. Ali Taylan Cemgil

mert.tasdemir@boun.edu.tr

June 5, 2017

Submitted to the Department of Computer Engineering in partial fulfillment of the requirements for the degree of Bachelor of Science
Boğaziçi University, Spring 2017

Overview

Objectives

The project consists of two parts:

- Finding the set of possible diagnoses given a set of symptom states i.e. whether they are found or not
- Finding the best question to be asked to the patient

State-of-Art

- Knowledge-based systems are outdated
- Neural Networks are of the leading approaches
- Bayesian Models are the other one leading
    - Exact inference
    - Variational inference
    - Sampling

Data Description

Data consists of m x n binary matrix showing the relations between diseases and symptoms.

```
1 0 0 0 0 0 0 0 0 1
0 0 1 1 0 0 1 0 0 0
0 0 0 1 1 0 0 0 0 0
0 0 0 0 1 0 0 0 1 1
0 0 0 0 0 0 0 0 0 0
1 1 0 0 0 0 0 0 0 1
1 1 1 0 1 0 0 1 0 0
0 0 0 0 1 1 0 0 0 1
0 0 0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0 0 1
0 0 1 0 0 0 0 1 0 0
0 1 0 0 1 1 1 0 0 0
0 1 0 0 0 1 0 0 0 1
1 0 0 0 1 0 1 0 0 1
0 0 0 0 1 0 0 0 0 0
0 0 1 0 0 1 0 0 0 0
0 1 1 0 0 0 0 1 0 1
1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 1 0 0 0 1 0
```

Figure: Sample 20 symptoms 10 diseases network data

$ds_{ij} \sim \mathcal{BE}(\left[0,1\right]; \pi_0, 1-\pi_0)$ where $\pi_0 \gg 1-\pi_0$ and ds_{ij} indicates if the symptom i is present for d_j, disease j. ds_{ij} i 1 when the symptom s_i is observed for d_j, and 0 when not.

Network Model

Noisy-OR model is adopted on a two layer Bayesian Network. Probability of observing a sympton given a disease is calculated as 1−production of not observing the symptom given each disease.Probability of not observing s_i given a disease set d is:

$$p(s_i = 1|d) = \theta_0 \prod_j \theta^{ds_{ij} d_j}$$

Probability of observing s_i given a disease d is:

$$p(s_i = 1|d) = 1 - \theta_0 \prod_j \theta^{ds_{ij} d_j}$$

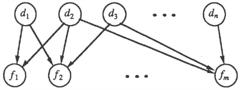

Figure: Two layer Bayesian Network

Problem Definition

- The problem is called maximum-a-posteriori(MAP) estimation, that is, the aim is to maximize the posterior probability of a disease set to be present given a set of symptom state set -whether observed or not, using the evidence found in the previous slide :
$$d^* = \arg\max_d p(s|d)p(d)$$
where d^* is a set of diseases that form a possible diagnosis.

Description

Finding the best diagnosis is one crucial part of this project and to be able to do that, one needs to know the states of the symptoms. Yet it is neither possible nor feasible to know about all the symptoms. Therefore asking the right question is also an essential problem.

Question-Generation Strategies

- Strategy based on Relative-Entropy
- Strategy based on Symptoms
- Strategy based on Diseases
- Strategy based on both Symptoms and Relative-Entropy

Strategy based on Relative-Entropy

Finding the question that reduces the Shannon Entropy more than all the others. To do so, for all symptoms:

1. Calculate the probability of observing a diagnosis before the examined symptom is known
2. Calculate the probability of observing the diagnosis when the examined symptom is known
3. Compare the relative-entropy of the first two result
4. Choose the unknown symptom that maximizes the 3^{rd} part

Strategy based on Symptoms

1. Rank the symptoms according to the number of diseases with which they are related
2. Chose the unknown symptom that has the highest rank

Strategy based on Diseases

1. Rank the diseases according to the number of symptoms with which they are related
2. Choose the highest ranked disease
3. Ask the symtoms related to the chosen disease
4. When there is no symptom related to the chosen disease, choose the disease that has the highest rank amongst the diseases that are not examined yet

Strategy based on both Symptoms and Relative-Entropy

1. Rank the symptoms according to the number of disease with which they are related
2. Choose choose highest X percentage of the symptoms
3. Calculate relative-entropy for chosen diseases
4. Choose the question that has the maximum relative-entropy i.e. causes maximum Shannon Entropy reduction
5. Note: X needs to be optimized, 25% seems like a good initialization

The strategies are compared according to the average number of symptoms needed to be known before the inference engine reaches to the state where all the symptom states are known.

Small Network Results

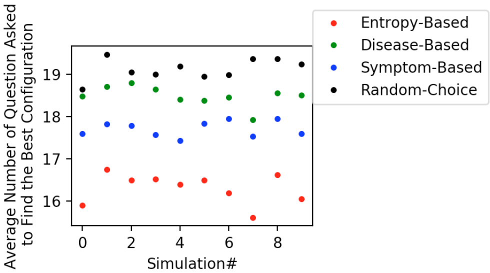

Figure: Small network with 10 diseases 20 symptoms

Mid-size Network Results

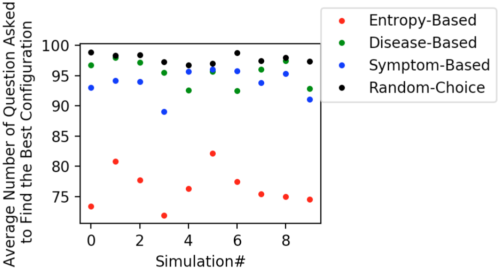

Figure: Mid-size network with 30 diseases 100 symptoms

Conclusion

- QGS based on diseases slightly outperforms random choice in both networks
- QGS based on symptoms slightly outperforms QGS based on diseases
- QGS based on relative-entropy outperforms both of the others, yet it's computationally infeasible in large networks
- The hybrid strategy outperforms the first two, yet it fails to beat the last one mentioned, yet it's computationally more feasible than it

Future Work

- Variational inference methods will be used
- Sampling methods like Gibbs sampling will be used
- New question generation techniques based on symptom disease relations will be tested
- Parallelized version of the QGS based on relative entropy and the exact inference algorithm will be implemented

References I

Hıdır Yüzügüzel, A. Taylan Cemgil, and Emin Anarım.
Query ranking strategies in probabilistic expert systems.
In *22nd Signal Processing and Communications Applications Conference*, 2014.

Hıdır Yüzügüzel and A. Taylan Cemgil.
Uzmandoktar: An expert system for diagnosis of agricultural plant diseases.
Boğaziçi University, 2013.

David Barber.
Bayesian reasoning and machine learning, feb 2017.

David Barber.
Machine learning - a probabilistic approach, 2006.

References II

Tommi S. Jaakola and Michael I. Jordan.
Variational probabilistic inference and the qmr-dt network.
Journal of Arti cial Intelligence Research, 10:291–322, may 1999.

William J. Long.
Medical informatics: reasoning methods.
Artificial Intelligence in Medicine, 23(1):71–87, aug 2001.

Ethem Alpaydin.
Introduction to Machine Learning (Adaptive Computation and Machine Learning series).
The MIT Press, 2014.

Bastian Wemmenhove, Joris M. Mooij, Wim Wiegerinck, Martijn Leisink, Hilbert J. Kappen, and Jan P. Neijt.
Inference in the promedas medical expert system.

References III

Adam Zagorecki and Marek J. Druzdzel.
Knowledge engineering for bayesian networks: How common are noisy-max distributions in practice?
IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, 43(1), jan 2013.

Jonathan D. Nelson.
Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain.
Psychological Review, 112(4):979–999, 2005.

Wim Wiegerinck, Bert Kappen, and Willem Burgers.
Bayesian networks for expert systems, theory and practical applications.

References IV

Igor Kononenko.
Machine learning for medical diagnosis: History, state of the art and perspective.
Artificial Intelligence in Medicine, 23(1):89–109, aug 2001.

Promedas - probabilistic medical diagnostic advisery system.

S R Soni, A Khaunteta, and M Gupta.
A review on intelligent methods used in medicine and life science.
In *International Conference and Workshop on Emerging Trends in Technology, Mumbai, India, February 25–26, 2011*, ICWET '11, pages 703–706, New York, NY, USA, 2011. ACM.

Thank you for watching