An Expert System for Medical Differential Diagnosis

Mert Cem Taşdemir

Advisor: Assoc. Prof. Ali Taylan Cemgil

Boğaziçi University Computer Engineering Department

Objectives

The project consists of two essential parts. One of them is to find the best set of diagnosis given a set of symptoms. The other is to find the best question to be asked to the patient to find the diagnosis.

Network Model & Approach

For the first part we tried to estimate maximum the probability of observing a disease combination $d*$, given set of symptoms $s_{1:k}$ where $k < j$, and $d*$ consists of $d_{1:m<t}$'s where t is a constant that reflects the maximum number of concurrent diseases *and we took it 3 for the sake of computational complexity*. In other words, we tried to make a *Maximum a-Posteriori(MAP)* estimation.

MAP:

$$d* = \arg\max_{d} p(s|d)p(d)$$

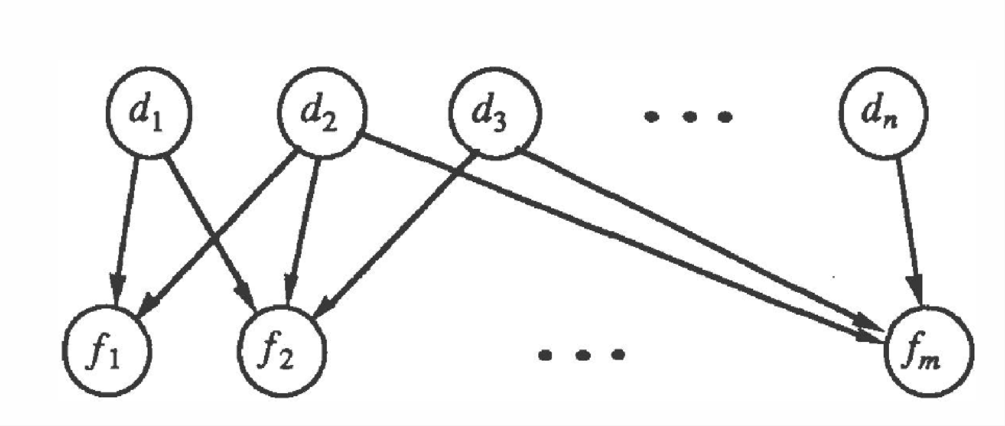

Figure: Two layer Bayesian Network for QMR-DT

State-of-Art

- Old school knowledge-based systems are outdated
- Neural Networks are on one cutting edge
- Bayesian methods are also quite popular.
- QMR-DT is the most popular probabilistic model on which this project is built.

Question Asking Strategies

- **Relative-Entropy Based Strategy :** Minimize the Shannon Entropy
- **Strategy Based on Symptoms :** Rank the symptoms according to their number of occurance in distinct diseases
- **Strategy Based on Diseases :** Rank the diseases according to the number of symtoms that they are related
- **Strategy Based on Symptoms and RE :** Hybrid of first two

Important Result

Large networks require so much power that makes the model nonutilizable after some point. So the model can be used for reduced problems.

Mathematical Section

Data description:

$$ds_{ij} \sim \mathscr{B}\mathscr{E}([0,1]; \pi_0, 1-\pi_0),\ \pi_0 >> 1-\pi_0 \quad (1)$$

Probability of not observing s_j is:

$$p(s_i = 0|d) = \theta_0 \prod_j \theta^{ds_{ij}d_j}$$

Probability of observing s_j is:

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

θ_0 is the probability of not observing a symptom when no disease related to it is present.

Minus log posterior:

$$-\mathscr{L} = -logp(d|s) \propto -logp(s|d) + log(1/p(d))$$

Results

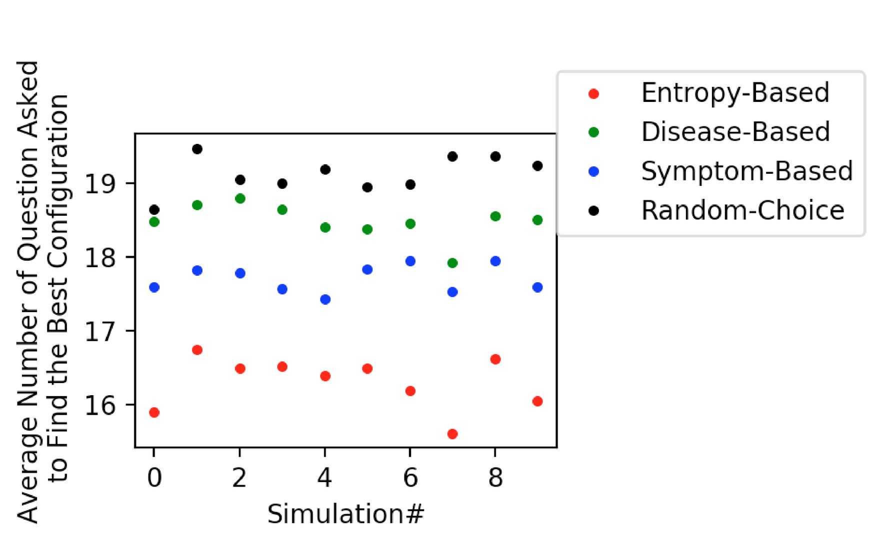

Figure: QGS Comparison - Small Network(20 Symptoms)

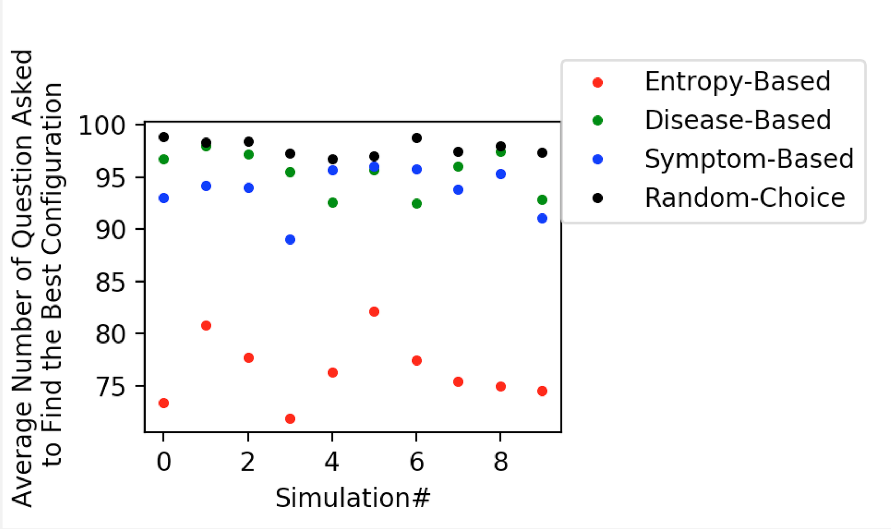

Figure: QGS Comparison - Large Network(100 Symptoms)

Conclusion

As mentioned, it's not working efficiently or even not tolerable for large networks. So, nonetheless it's very efficient and high scoring model for small networks, a scalable model seems necessary.

Future Work

Delving into variational methods and their application to this domain since the inference is infeasible in large models. Besides them, Monte Carlo methods should be examined.

As Confucius said "The man who moves a mountain begins by carrying away small stones." and this project was one of the smallest ones.

Special Thanks to

- My advisor Ali Taylan Cemgil for allowing me to participate this presentation after too much confusion of which I am responsible for, and helping me to make the simple and extremely necessary move.
- Hıdır Yüzügüzel for his great project -UzmanDoktar, project report and implementation.