6.S897/HST.956 Machine Learning for Healthcare

Lecture 11: Machine learning for Differential Diagnosis

Instructors: David Sontag, Peter Szolovits

1 Introduction

Differential diagnosis is the distinguishing of a particular disease or condition from others that present similar clinical features. Differential diagnostic procedures are used by physicians to diagnose the specific disease in a patient, or, at least, to eliminate any imminently life-threatening conditions. Example: In the evaluation of Cough, there might be two possible diagnosis: Acute Bronchitis and Common Cold. Acute Bronchitis will be considered as differential diagnosis even if the final diagnosis is common cold.

1.1 Guyton's Model of Cardiovascular Dynamics

Arthur Guyton (1919-2003) was an American physiologist who modelled the basic functioning of the cardiovascular system. It can be used for understanding the dynamics of heart based on the condition of the heart. Based on the observed signs, it could be used for detecting diseases. However it was not successful because it required a lot of tuning which is not possible in practical scenarios.

2 Models for Diagnostic Reasoning

This lecture will cover different models for diagnostic reasoning, most of which are based on associations between diseases and signs or symptoms. To clarify the difference between signs and symptoms, a sign refers to any objective evidence of disease, which is usually perceived by the patient. On the other hand, a symptom refers to subjective evidence of disease recognized by the patient, physician, nurse, or someone else. The various methods of Diagnosis and its reasoning will be covered one by one

2.1 Flow Chart

It is one of the earliest method that involves asking standard set of questions and follow the path on the flow chart based on the answers given by the person and then come to conclusion about the diagnosis.
It is not a good method because it is fragile, very specific and doesn't take unusual cases into consideration.

2.2 Card Selection

In this process, card for each patient is taken and based on symptoms they are having, an extra hole around punch hole (corresponding to the symptom) is cut on the card. After collecting the cards from the people. The cards can be segregated based on symptoms by lifting them using the punch holes corresponding to the symptom. The patients that have a those symptom, the card will fall down. Using this way we can get the patients that have same symptoms and diseases. But if there are large number of symptoms then this method will fail as very less/ no cards will fall down.

2.3 Naive Bayes

The Bayes rule is given in the equation below. It is powerful tool to determine what should be the probability of person having disease after a given observation, symptom or a test result.

$$P_{i+1}(D_j) = \frac{P_i(D_j)P(S|D_j)}{\sum_{k=1}^{n} P_i(D_k)P(S|D_k)}$$

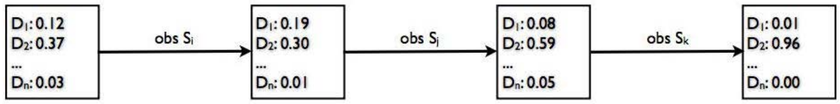

Figure 1: After observing each symptom, revising the distribution according to Bayes rule.

It assumes the conditional independence of the symptoms Si.

$$P(S_1, S_2, ..., S_n|D_i) = P(S_1|D_i)P(S_2|D_i)...P(S_n|D_i)$$

Probability of a person having disease can be updated (Figure 1) given the symptoms by using the update equation given below:

$$P^j(D_i|S_1, ..., S_j) = \frac{P^{j-1}(D_i)P(S_j|D_i)}{P^{j-1}(S_j)} = \frac{P^{j-1}(D_i)P(S_j|D_i)}{\sum_{i=1}^{n} P^{j-1}(D_i)P(S_j|D_i)}$$

2.3.1 Odds Likelihood

Since Bayes rules involves a lot of multiplication, it can be converted into the Odds using the equation below:

$$O = P/(1 - P) = P/\neg P$$

Odds-likelihood form of Bayes rule is below:

$$L(S|D) = P(S|D)/P(S|\neg D)$$

$$O(D|S_1, .., S_n) = O(D)L(S_1|D)...L(S_n|D)$$

Performing the log transformation of the Odds Likelihood form of Bayes rule, a different form is obtained that involves additions.

$$\begin{aligned}\log O(D|S_1, .., S_n) &= \log[O(D)L(S_1|D)...L(S_n|D)\\ &= \log[O(D)] + \log[L(S_1|D)] + ... + \log[L(S_n|D)]\\ &= W(D) + W(S_1|D) + ... + W(S_n|D)\end{aligned}$$

2.4 Rationality

Principle of rationality says that actions should be aimed towards maximizing the expected utility. In the scenarios in which the decision is to be taken that involves life vs death then action should be taken such that utility is maximised.

2.4.1 Decision Tree for Gangrene Case

A case from New England Medical Center Hospital, from 1970's was discussed, that involved a man suffering from Gangrene and decision is to be taken to continue medicines or amputate the foot. If the medicines fail then the person might die or there might be need of amputation of whole leg. In such cases decision trees can be made as shown in the figure 2. The values of utility for each decision can be obtained from the patients.

Demo of Acute Renal Failure Program [GKES73] was shown in the class that involved finding potential causes, when the person "stopped peeing" or amount of urine discharged is very less. Such models involve use of some kind of probabilistic graphical model to reach the conclusion.

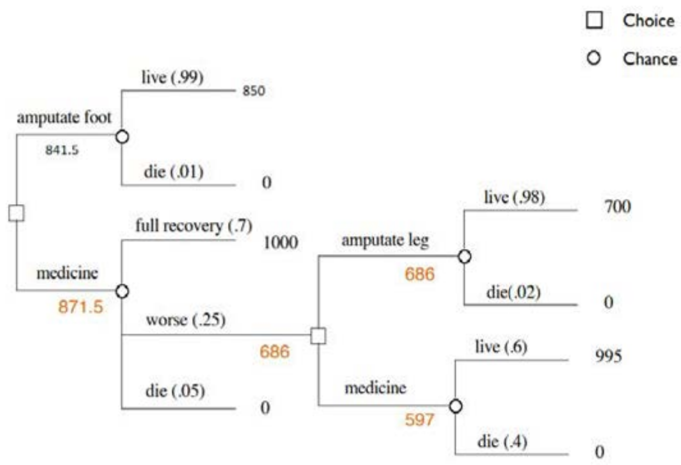

Figure 2: Folding Back of decision Trees

2.5 Bipartite Graph Model

This model is used for the diagnosis of the multiple diseases. The Diseases (Di) are considered as independent variables. The manifestations depends only on which diseases are present. Manifestations (Mi) are conditionally independent. The problem with the Bipartite graph model is they are computationally intractable as the number of undirected cycles increases. A typical graphical model is shown in figure **??**

2.5.1 Dialog/Internist/QMR [MPJM82]

It was a model that was developed in 1982 based on the Bipartite Graph models. For Each Disease Dx, the associated manifests are listed with the invoking strength and frequency. For each manifest Mx, the importance of the manifest is listed.
It uses abductive logic for inference. The steps for inference are:

- List Mx of a case
- Evoke Dxs with high evoking strengths from Mxs
- Score Dx, positive or negative based on evoking strength of observed manifestation.
- Form differential around the highest scoring Dx.

To accommodate the multi-hypothesis diagnosis, they used the heuristic.

- Set aside the complimentary hypothesis and manifestations predicted by them.
- Solve diagnostic problem among competitors.
- Eliminate confirmed hypotheses and manifestations explained by them.
- Repeat as long as there are coherent problems among the remaining data.

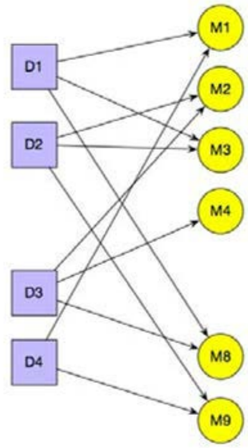

Figure 3: Bipartite Graph Model.

2.6 1990s Evaluation of Diagnostic Systems

[BWS+94] evaluated the methods QMR, DXplain, Iliad, Meditel on 105 cases that were labelled by the experts. The results evaluated the methods based on the Coverage, Correctness, Rank, Relevance, Comprehensiveness and Value addition. From the Evaluation some of the results were: None of the method performed consistently better or worse on all the measures. Although sensitivity and specificity were not impressive, the programs had the additional functions that were not evaluated like interactive displays of signs and symptoms.

2.7 QMR DT model

It interprets QMR data as the bayesian network with the following assumptions:

- Marginal independence of Dx, conditional independence of Mx.
- Binary Dx and Mx.
- Causal independenceleaky noisy-OR.
- No distinction between Mx that predispose to a Dx and those that are a consequence of the Dx.
- Priors on Dx estimated from health statistics.
- Estimate leak for each Mx from Importance values.

It used the method similar to the QMR method for inference. It used likelihood weighting to estimate posteriors.

There exists variety of symptoms checkers [SLGM15] that can be used for self diagnosis. Symptoms checker uses the some of the above methods or new methods that can give you the possible diagnosis based on the questions asked to the user.

2.8 Rationality under resource constraint

There are some scenarios in which decision is to be made under a small amount of time. Because of lack of time, the utilities decrease[Hor90]. In such scenarios there should be fast inference algorithms that don't

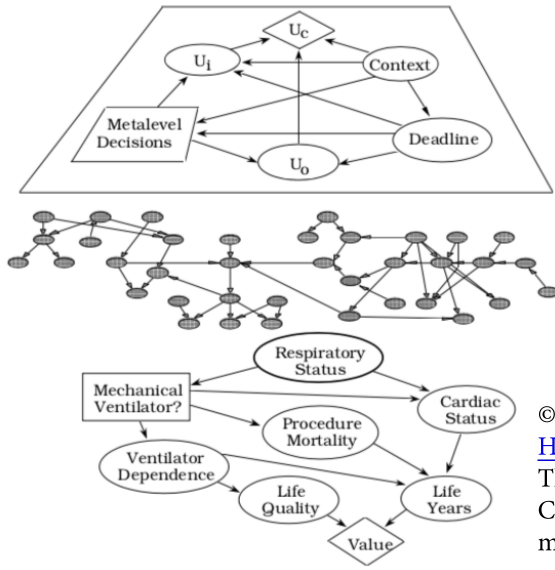

Figure 4: A representation of a time-pressured decision problem. From top to bottom, the three sections of the figure portray (a) the decision-theoretic metareasoning problem, (b) a belief network representing propositions and dependencies in intensive-care physiology, and (c) a closeup on the respiratory status node, and its relationship to the current decision problem.

take much time. So the value of computation also needs to be taken into consideration. An example of the time pressured decision problems modelled as belief network as shown in figure 4 [HCH89].

2.9 Reinforcement Learning

Some of the new machine learning methods have developed that can be used for speeding up the diagnosis. One of the method is to formulate the diagnosis problem as the Reinforcement learning problem. It is a method of training networks based on the rewards. If network outputs correct diagnosis then it gets positive reward. It gets negative reward it outputs incorrect diagnosis. The model is standard q-learning framework, using double-deep NN strategy. The state space is set of +ve and -ve findings. The action space is ask about the finding or it should conclude about the diagnosis. One of the architecure [PTLC18] is to learn a dual Neural Network architecture as shown in the figure.

References

[BWS+94] Eta S Berner, George D Webster, Alwyn A Shugerman, James R Jackson, James Algina, Alfred L Baker, Eugene V Ball, C Glenn Cobbs, Vincent W Dennis, Eugene P Frenkel, et al. Performance of four computer-based diagnostic systems. *New England Journal of Medicine*, 330(25):1792–1796, 1994.

[GKES73] G Anthony Gorry, Jerome P Kassirer, Alvin Essig, and William B Schwartz. Decision analysis as the basis for computer-aided management of acute renal failure. *The American journal of medicine*, 55(4):473–484, 1973.

[HCH89] Eric J Horvitz, Gregory F Cooper, and David E Heckerman. Reflection and action under scarce resources: Theoretical principles and empirical study. In *IJCAI*, pages 1121–1127, 1989.

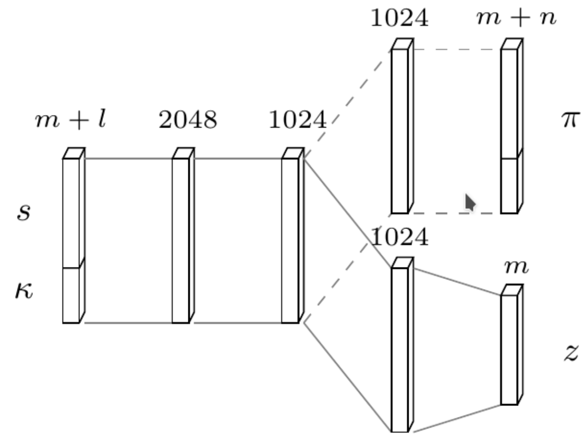

Figure 5: Dual neural network architecture, which comes at the moment of the final action of The upper branch is the policy π of an agent. classification, is rather weak to the agent. Thus, the The lower branch is the feature rebuilding part concept of not wasting queries can be slow to learn of sparse features.

[Hor90] Eric J Horvitz. *Rational metareasoning and compilation for optimizing decisions under bounded resources*. Knowledge Systems Laboratory, Medical Computer Science, Stanford University, 1990.

[MPJM82] Randolph A Miller, Harry E Pople Jr, and Jack D Myers. Internist-i, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 307(8):468–476, 1982.

[PTLC18] Yu-Shao Peng, Kai-Fu Tang, Hsuan-Tien Lin, and Edward Chang. Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. In *Advances in Neural Information Processing Systems*, pages 7333–7342, 2018.

[SLGM15] Hannah L Semigran, Jeffrey A Linder, Courtney Gidengil, and Ateev Mehrotra. Evaluation of symptom checkers for self diagnosis and triage: audit study. *bmj*, 351:h3480, 2015.