Inferring causal strength and structure in Bayesian networks: approaches and applications in medicine

Julie Vaughn
6.804/9.66 Final Project

Author's Note

This project began with an interest in applying the concepts of this class to understanding how doctors think. I am very interested in applications of AI to understanding medical data, and wondered if we can model the way that doctors learn from patient data – the best doctors are arguably the best because they have a lot of experience identifying diseases and treating a large number of patients. In this vein, I became curious about how humans infer causal structure from data, and especially medical data. The problem of causal structure learning of Bayesian networks (i.e., what is the underlying cause and effect structure in a network of concepts) turned out to be a very complex and interesting problem, perhaps more so than inference over a pre-determined network. I explored a large number of papers in this space, and in the end decided to base a large part of this project on the paper *Structure and strength in causal induction.* (Griffiths & Tenenbaum, 2005). I hope that the work described here is reasonable and appropriate, given that I am entirely new to this field. Many assumptions made about this project are based on casual conversation with the TAs in this class or a friend, and I try to denote that the sources of these assumptions where possible. Please see the accompanying Python notebooks for my original implementations of the concepts described here.

Abstract

Inference of causal relationships is an important and often misleading element of human cognition. In this paper, we seek to model how a human might learn the causal structure of simple, 2-layer Bayesian networks that follow a noisy-OR structure by observing data sampled from this structure. We accomplish this through the use of traditional rational models of causality and correlation in psychology – namely, ΔP, causal power, and mutual information to determine the presence or absence of a link between two neighbors. The structures assumed using thresholds on these measures to determine causality is then compared to the original structure from which the data is sampled. It was found that the structure tended to vary significantly in most cases, suggesting that humans may cognitively have a difficult time inferring complex causal structures from data alone, in the absence of domain knowledge.

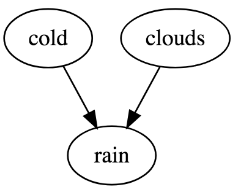
Introduction

Causality

The induction of causal relationships is a critical tool for understanding the world. Across many fields of modern science, numerous statistical tools have been developed in the last century to infer causal relationships from experimental results. However, these fields were developed long before these statistical tools were formalized (Griffiths & Tenenbaum, 2005). It is therefore interesting to investigate the methods by which humans cognitively infer causal relationships without directly employing quantitative calculations, as these methods are also timelessly paramount to the process of learning and survival in general. For example, after consuming a particular substance and then experiencing an allergic reaction to it, it is important that a human consider their actions and conclude that the consumption of that particular substance is responsible for their reaction. There are also interesting results that arise as a result of humans being more or less susceptible to inferring causal relationships – some superstitions may arise from the confusion of coincidental correlation with causation (i.e., "when I wear my favorite socks, it helps my favorite team wins the superbowl"). The obfuscation of a causal relationship is caused in part by the existence of several causal factors for a particular effect. Effectively, it makes sense that many leading theories about causality assert that we cannot accurately infer a causal relationship without the ability to directly perturb the system and observe the effects of it subsequently in time. In this project, the focus is on probabilistic (i.e. non-deterministic) cause and effect relationships where variables are first categorized into causes and effects (modeling a particular latent directed structure perhaps based on logical beliefs about the world). This paper will first explore two historical rational models of causality, ΔP (Jenkins & Ward, 1965) and causal power (Cheng, 1997) across a number of causal scenarios. We will then investigate a specific case of inferring the structure of medical symptom/disease networks and the implications of doing so.

Bayesian Networks

Bayesian networks (BNs) are a useful way of modeling the interplay between multiple cause and effect relationships. They are particularly efficient at compactly representing joint probability distributions (Pearl, 1988). An example Bayesian network and corresponding conditional probability table (CPT) is shown in Figure 1 below.



P(Cold)
.5

P(Clouds)
.6

Clouds	Cold	P(Rain)
n	n	.1
n	y	.3
y	n	.6
y	y	.75

Figure 1: Bayesian network with probabilities of each Bernoulli variable shown.

If a causal representation of a domain can be represented with a graphical BN, we may assume that learning cause and effect relationships in the domain may be accomplished through the use of structure learning (Pearl, 2000). Note that structure refers primarily to the shape of the graph (i.e. what nodes and edges exist, and in what formation are they connected). A number of algorithms for the inference of structure have proven successful at modeling several domains (Tong & Koller, 2001; Drton & Maathuis, 2016). Specifically, several Bayesian and frequentist models have been developed for active structure learning of directed acyclic graphs (DAGs), as this is a less complex problem than inferring structure for cyclic structures (Drton & Maathuis, 2016). It should be noted that, in this project, inferring structure from statistical measures is meant more to model how a human would infer structure from observing the data, rather than how to actually most accurately determine a structure from data.

Methodology

Overview

Overall, the purpose of this project is to investigate the accuracy of models of human perceptions of causality. We will first consider established cognitive representations of causality in the elemental causal induction (i.e. one cause and one effect) case, and see how effectively they perform in different scenarios.

We will then consider a variety of two-layer Bayesian networks composed of Bernoulli random variables. These networks will comprise of a causal layer and an effect later, in which there is a number of disease nodes (all mutually independent) which then may or may not cause a number of symptom nodes (also mutually independent from the other symptoms).

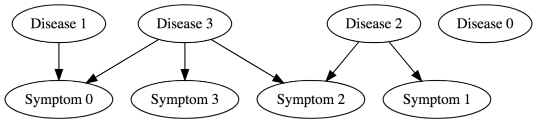


Figure 2: An example symptom/disease network generated in this project

The networks will assume a noisy-OR structure (Pearl, 1988) (see methodology section *Functional Causal Relationships).* These Bayesian networks can then be used to stochastically generate an arbitrary number of patient data points. The task, then, is to use this generated data to infer the structure again. We will additionally use the observed mutual information here as a measure of causality here, because it may be used as a measure of variable independence and information flow. Furthermore, it mimics the human tendency to mistake correlation for causation and may provide an interesting model for the way humans will infer causal structure from correlational data.

Associated Programs

The experiments described in this paper are carried out in several Jupyter notebooks in the attached zip file. The *Medical Bayes Net* notebook is the final form of the code in the other two notebooks. They are written in Python, with the assistance of the Pyro, Numpy, CausalGraphicalModel, and Math libraries. Class models in the *Medical Bayes Net* notebook for "Disease" objects and "MedBayesNet" objects may be of particular use for future experiments involving this structure of networks.

ΔP as a Measure of Causal Strength

A long-standing proposal for modeling the human perception of the strength of causality is the ΔP model. It was first suggested by Jenkins & Ward (1965), and later explored by Allan (1980, 1993). It is a measure that concerns the extent to which two variables co-vary, i.e. are both present or absent. The formula for calculating ΔP for some cause c and some effect e is shown below:

$$\Delta P = \frac{N(e^+,c^+)}{N(e^+,c^+)+N(e^-,c^+)} - \frac{N(e^+,c^-)}{N(e^+,c^-)+N(e^-,c^-)}$$

$$= P(e^+|c^+) - P(e^+|c^-)$$

A common interpretation of ΔP values is that if the value is large and positive, the cause is generative, if the value is large and negative, it is preventative, and if the value is close to zero it is not causally linked to the effect. This measure has been widely used and discussed in both philosophy and psychology. In this paper, we will evaluate the accuracy of it in predicting a causal relationship from data generated from a Bayesian network.

Causal Power

Cheng (1997) proposed causal power as an alternative to ΔP, arguing that human judgments of causality are not as purely covariational as ΔP would suggest. The equation Cheng proposes for generative causes is:

$$p_c = \frac{\Delta P_c}{1 - P\left(e \mid \bar{c}\right)}.$$

And for non-generative causes:

$$p_c = \frac{-\Delta P_c}{P\left(e \mid \bar{c}\right)}$$

In a sense, Cheng adds a term to the denominator that allows us to better account for the performance of the "control group" in the study – how we think about cause differently when the effect is frequently observed without the cause. For example, in the case where both control and treatment groups do not exhibit an effect, the equation becomes an undefined measure rather than a negative measure as with ΔP, better reflecting the perceptions of the researcher (that no causal relationship may be inferred). In this paper, we will consider the accuracy of this measure in predicting causal structure in Bayesian networks as well.

Mutual Information

Mutual information is a further estimate of correlation we will investigate in this paper. In information theory, it is equivalent to the KL divergence of a joint distribution and its constituent component variables. In a sense, it is the distance between these two distributions. If we were to perfectly capture the probabilities of two independent variables A, and B, it follows that

$$P(X, Y) – P(Y)*P(X) = 0$$

By definition of conditional independence. If this is the case, then the mutual information of the variables should be 0:

$$\operatorname{I}(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

Thus, when evaluating structure of a Bayes net, we may try inferring a casual relationship according to when the MI exceeds a particular epsilon close to 0.

Functional Causal Relationships

There are several causal relationships discussed by Griffiths & Tenenbaum (2005). The example graph used in the paper is shown below:

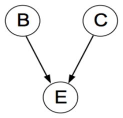


Figure 3: The graph to the left represents both a known cause and background cause contributing to an effect

These relationships involve the the possible relationships between a background and direct cause and their subsequent effect. These are the noisy-OR, the noisy-AND-NOT, and the linear hypothesis. Griffiths and Tenenbaum gave the following explanation and calculations for those relationships:

- Noisy-OR – applies when we can attribute the effect to any of it's causes, and the effect does not occur when none of it's causes are present. It assumes that all causal relationships described are generative. It can be roughly generalized for the example graph as "If B occurs or C occurs, then E will occur, otherwise it will not occur". The terms w_0 and w_1 are probability weights, while b and c are either 1 or 0 (if they are present in a patient or not, respectively).

$$P_1(e^+|b,c;w_0,w_1) = 1 - (1 - w_0)^b(1 - w_1)^c$$

- Noisy-AND-NOT – applies when we can attribute one cause to be generative and others to be preventative.

$$P_1(e^+|b,c;w_0,w_1) = w_0^b(1 - w_1)^c,$$

- Linear – Presence of a cause will increase/decrease the probability of an event by a linear amount.

$$P_1(e^+|b,c;w_0,w_1) = w_0 \cdot b + w_1 \cdot c.$$

Medicine and Causality

For the purpose of this paper, we will assume that medical diseases are purely generative. We will assume that diseases will never inhibit symptoms, and thus the probability of a disease generated as a result of multiple possible causes will only increase as more of these causes are observed to be present. This can be mathematically implemented by assuming that diseases and symptoms exhibit the noisy-OR relationship described in the previous section. Each present disease that is a cause of a symptom contributes a $(1\text{-}w_n)$ term to the overall expression for the probability of a symptom:

$$p(symptom) = 1 - (1 - w_0)(1 - w_1)(1 - w_2) \ldots (1 - w\_n)$$

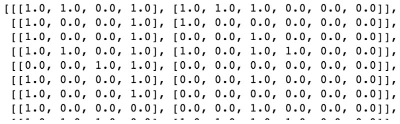
The implementation of this structure is further described in the following section.

It should be noted that the usefulness of this particular abstraction as a model of human diagnostic abilities or as an AI system to assist in diagnosis is likely very limited. First, a purely probabilistic system does not take into account categorical knowledge that a doctor will rely on, and demands a large amount of data to train before drawing any legitimate conclusions about correlation or causality (Szolovits & Pauker, 1978). Furthermore, the assumption that within themselves, the set of all symptoms and the set of all diseases are independent is difficult to justify, even in seemingly unrelated diseases. One could build a seemingly infinite net of all the causal factors and determinates of symptoms that affect every patient if full accuracy is desired.

Medical Bayes Net Construction

In the attached Python notebook, a Medical Bayes Net (MBN) class is established, alongside representations for diseases, symptoms, and patients. Sampling methods are also developed to generate patient data from these networks.

To create an MBN object, the program first generates Disease objects that have some randomly chosen prior probability. It then generates symptoms, which are represented simply as integer values. Then, the program uses the noisy-OR formula to create a "symptom dictionary" for each disease, where a random subset of symptoms is attributed to that disease assigned random p-value.



```
[[[1.0, 1.0, 0.0, 1.0], [1.0, 1.0, 1.0, 0.0, 0.0, 0.0]],
 [[1.0, 0.0, 0.0, 1.0], [1.0, 0.0, 0.0, 0.0, 0.0, 0.0]],
 [[1.0, 0.0, 0.0, 1.0], [0.0, 0.0, 1.0, 0.0, 0.0, 0.0]],
 [[1.0, 1.0, 0.0, 1.0], [1.0, 0.0, 1.0, 1.0, 0.0, 0.0]],
 [[0.0, 0.0, 1.0, 1.0], [0.0, 0.0, 0.0, 0.0, 0.0, 0.0]],
 [[1.0, 0.0, 0.0, 1.0], [0.0, 0.0, 1.0, 0.0, 0.0, 0.0]],
 [[1.0, 0.0, 0.0, 1.0], [0.0, 0.0, 0.0, 0.0, 0.0, 0.0]],
 [[1.0, 0.0, 0.0, 0.0], [0.0, 0.0, 1.0, 0.0, 0.0, 0.0]],
```

Figure 4: Generated samples from a 4 disease, 6 symptom net

To sample from an MBN, the program first samples from the distribution's priors to get a list of diseases that a patient has. It then takes that list and uses it as an input to the following formula, in determining the probability that a particular patient has a symptom, given that they have certain diseases. This expression is derived from the noisy-OR model, and assumes that each constituent disease is either present or absent in the patient. The patient is represented in a binary format, as [[diseases], [symptoms]], with the index of both corresponding to their index in the MBN from which they are derived. For example, the patient [[0, 1], [0,0,0,0,1]] has Disease 0 but not Disease 1, and only shows Symptom 4 out of the 5 possible symptoms in the network.

Structure Learning

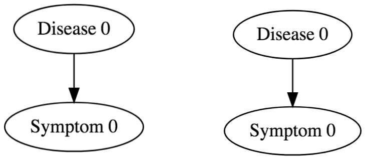
From the data generated by an MBN, we are able to attempt to infer the structure of a Bayesian network as an untrained human might by using the statistical parameters ΔP*,* causal power, and mutual information to estimate the causality between each (disease, symptom) pair, and add a structural link when our estimate exceeds a certain threshold. For mutual information, the cutoff was roughly determined to be most accurate at epsilon = 1/n, where n is the number of samples. So, if the MI value for a pair of values is below epsilon, the variables are assumed to be independent.

Results

MBN Structure Determination

Note: please see the Results section at the end of the attached python notebook if you would like to validate this section to yourself, or see the supplementary data files from different trials to see a more comprehensive analysis of the data presented here.

Data tables storing information about the causal estimator performances compared to the actual performance were created, and the overall % correctness of the structure evaluated. Structural graphics and the original data are also shown. In all cases, an MI link or CP link was concluded when the value of the MI or causal power of the data was greater than epsilon = 1/(number of samples).



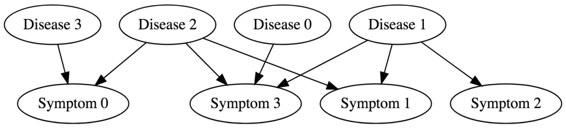
	D	S	Original p_val	MI	Causal Power	deltaP	MI Link	CP Link
0	0	0	0.22605	0.058189	0.221001	0.221001	1	1

Figure 6: Original (L) and inferred (R) structures are identical. The MI, deltaP, causal power, and linkage values for the connection are given in the table.

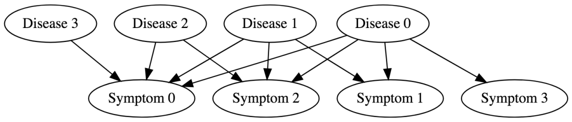
For simple models like in the figure above, Causal Power and Mutual Information are very good predictors of causal relationships, and the MI-derived inferred graph almost always showed the same structure as the original graph. The MI performance was largely consistent in randomly-generated graphs equal to or smaller than 2 diseases x 2 symptoms, and across samples sizes 100, 1000, and 10000. The 1000 sample size is the one used to generate the figures in this paper.

However, when we tried more complex graphs (i.e. with more nodes), the accuracy is much less than 100% most of the time. See the example below of a randomly-generated 4 diseases, 4 symptom graph. The structure-learning accuracy is 50%.

Original



MI-Inferred



	D	S	Original p_val	MI	Causal Power	deltaP	MI Link	CP Link
0	0	0	NA	8.116139e-03	0.019108	0.019108	1	1
1	0	1	NA	8.980117e-02	0.195860	0.195860	1	1
2	0	2	NA	8.116139e-03	0.019108	0.019108	1	1
3	0	3	0.799245	5.916749e-01	0.834395	0.834395	1	1
4	1	0	NA	1.642124e-02	0.030691	0.030691	1	1
5	1	1	0.428565	1.866537e-01	0.314578	0.314578	1	1
6	1	2	0.0225962	1.642124e-02	0.030691	0.030691	1	1
7	1	3	0.0858054	1.010047e-04	-0.494071	-0.159847	0	0
8	2	0	0.773601	4.952252e-02	0.190476	0.190476	1	1
9	2	1	0.308278	7.076074e-06	-0.004770	-0.004145	0	0
10	2	2	NA	1.133591e-03	-0.012295	-0.012146	1	0
11	2	3	0.698064	7.055361e-09	16.111111	-0.507703	0	1
12	3	0	0.0182105	2.386579e-03	-0.012295	-0.012146	1	0
13	3	1	NA	9.274325e-05	0.013588	0.011946	0	1
14	3	2	NA	1.754587e-04	-0.003358	-0.003321	0	0
15	3	3	NA	6.311295e-04	-7.677083	-0.387080	0	0

Figure 5: Original and MI-inferred structure of a 4x4 network, and summary table of (D,S) pairs.

Evidently, it is more difficult for the MI structure learning algorithm to accurately determine cause and effect relationships when there are many of these relationships at play. This holds true across many sample sizes and graph complexity-levels.

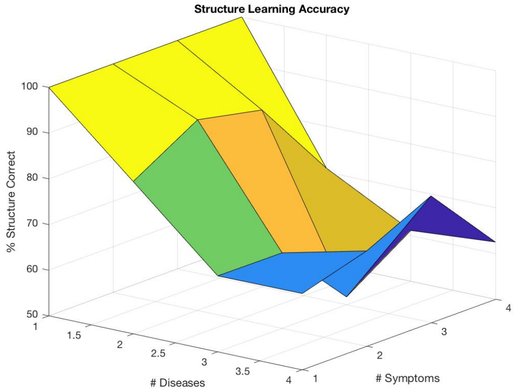


Figure 7: Graph of structure-learning accuracy averaged over 3 random graphs of each size, up to 4x4

Discussion

Human Perceptions of Causality

If we consider the MI structure learning algorithm to be a reasonable approximation of how humans determine causality from data alone, then this project shows that it is very difficult for humans to separate coincidence from causality in cases where there are many different causes and effects. In cases where there are much fewer possible causes and effects (e.g., 1 disease, 2 symptoms), a human is much more likely to be able to infer all of the correct causal relationships (and lack thereof). This is especially true in cases when there are more symptoms then diseases (see previous figure). The results may suggest statistical evidence, in a sense, that it is very difficult to determine independence (aka the lack of a causal link), from data alone. If we assume the underlying structure of the world is Bayesian, it would be difficult to conclude independence of any particular variable with another. Much like the causal inference program, humans may mistake coincidence or correlation with causation, and predict causal links when they are not there. Such links are only broken through rational thought – can we explain the causal link as a rational/categorical

Future Work

Ideally, I would have liked to perform maximum log likelihood estimation using the noisy-OR parameterization generated from the different frequencies of data observed. Given data from a particular MBN with known structure, it would be interesting to observe the most likely weightings in the following cases (i) assuming a fully-connected graph, (ii) assuming a graph that is inferred the way a human might (i.e., as predicted in this experiment), and (iii) in a graph that is accurate (i.e. graphically equivalent to the original). I would also like to apply this structure learning framework to real medical data, instead of data synthesized from a Bayes Net, and also to see if people perform similarly to the structure learning algorithm provided here if given the same task (with much less data, of course). Unfortunately, timing did not permit this exploration.

Unrelated Application: Visual Diagnostic Tool

If we can achieve accurate inference of the causal structure of specific disease and symptom networks, we may perhaps be able to develop a visual reference tool for doctors to use in order to aid in diagnosis. The causal structure may be best estimated, in reality, with either an active learning method, or by integrating over a number of different weights over a number of structures and selecting the one that is most likely. (Drton & Maathuis, 2016)

This idea is inspired by the intuition that deadly misdiagnoses (for example, misdiagnosing Osteosarcoma as growing pains), occurs when a doctor doesn't consider all possible causes of a disease and instead relies on categorical judgment entirely. By simply visually suggesting that the doctor consider several diseases as possible diagnoses for a symptom, we can ensure that the doctor considers the possibility that the patient has a different disease. Additionally, by color-coding the graph with severity, and making the sizes of the nodes match the relative prevalence of the disease, we can make a visual tool that best encompasses our current understanding of medicine. If this program is sufficiently dynamic, perhaps it could capture the various dependencies between a large number of symptoms and diseases and continuously update its assumptions of causality based on new findings from publications.

Conclusions

Overall, the findings suggest a statistical basis for superstition and misattribution, akin human nature. If we base the structure of our Bayesian models of the world purely on the interpretation of data (rather than a combination of logic and data), then we risk observing associations that do not exist, even with vast amounts of data. Determining and representing causal structure is medicine is a complex but promising task.

Appendix

Please see the following links for a more extensive summary of the data used in this paper:

All Data:
https://docs.google.com/document/d/1Wm3nzMPfbdUMyZ533fLWS1MWDd-Hfm2ypdvcURE_4Mg/edit?usp=sharing

% Correct:
https://docs.google.com/spreadsheets/d/1pEu4NwQtPih5Sko5-ioDlICsO9mjAHDECJmiYWmQqCc/edit?usp=sharing

One may also wish to use the attached Medical Bayes Net notebook to conduct experiments of their own, and add to these appendixes.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15, 147–149.

Allan, L. G. (1993). Human contingency judgments: *Rule based or associative? Psychological Bulletin,* 114, 435–448.

Cheng, P. (1997). *From covariation to causation: A causal power theory. Psychological Review*, 104, 367–405.

Drton, M., & Maathuis, M. H. (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4, 365-393.

Friedman, N., & Koller, D. (2000). Being Bayesian about network structure. In *Proceedings of the 16th annual conference on uncertainty in AI*, Stanford, CA (pp. 201–210).

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334-384.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, 79.

N. D. Goodman, J. B. Tenenbaum, and The ProbMods Contributors (2016). *Probabilistic Models of Cognition* (2nd ed.). Retrieved 2018-12-16 from https://probmods.org/

Pearl, J. (1988). Probabilistic reasoning in intelligent systems. San Francisco, CA: Morgan Kaufmann.

Pearl, J. (2000). Causality: Models, reasoning and inference. Cambridge, UK: Cambridge University Press

Szolovits, P., & Pauker, S. G. (1978). Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, 11(1-2), 115-144.

Tong, S., & Koller, D. (2001, August). Active learning for structure in Bayesian networks. *International joint conference on artificial intelligence* (Vol. 17, No. 1, pp. 863-869). Lawrence Erlbaum Associates ltd.