figure out theoremstle is undefined

Learning how to act: making good decisions with machine learning

Finnian Lattimore

April 29, 2017

Notes on papers

from association to causation via regression (Freedman) very clear description of the goal of causal inference, very critical causal modelling due to the impossibility of empirically assessing the assumptions.

What is causality and why do we care?

The meaning of causality remains widely debated within science and philosophy ...

For this thesis I define a causal model to be any model that aims to predict the outcome of an action or intervention in some system. From this viewpoint, most real problems require causal inference. What use is a model if you are not going to use it to guide any decisions? How then can machine learning be integrating itself into so many key aspects of our lives, while causal inferene remains a small subfield mostly of interest to economists? The answer is two-fold. For some problems, the effect of a decision is obvious and does not impact the system on which it was based. In these cases, the causal aspect of the problem can be safely ignored. Secondly it is likely that machine learners and data scientits, bouyed by successes in the former problems and the excitement of a world awash with data, computing power and non-linear models are overlooking the fact that their predictions may no longer hold if people act on them. Consider the following problems, do they require causal inference?

notes: We don't always need to predict the outcome of an action. In somes

The goal may be more limitied than predicting the outcome of an intervention. To decide between actions, we only assumptions or estimates as to which is better.

For example, in the speech recognition, the immediate result of the action may be known and deterministic (Siri does something). The full consequences of that action are not modelled, (we assume the desired outcome is that Siri)

- Speech recognition (for systems like Siri or Google)
- Machine translation
- Image classification
- Forecasting the weather
- Playing Go
- Identiying spam emails

- Automated essay marking
- Predicting the risk of death in patients with pnumonia.
- Predicting who will re-offend on release from prison
- Predicting which customers will cease to be your customers
- Demand prediction for inventory control
- Predicting who will click on an ad
- Financial trading
- Recommending movies
- Online search
- Self driving cars

The above problems are not posed with enough detail to know if causality is an important consideration. In particular, I failed to specify what actions the might be taken in response to model.

Consider speech recognition. You say something, which causes to sound waves, which are converted to a digital signal which Siri maps to words. Whatever action Siri takes is unlikely to change the distribution of words you use, and even less likely to change the function that maps sound waves to text (unless she sends you a dvd on elocution). A similar arguament could be made for many applications of machine translation and image classification. What about forecasting the weather? If you are using a short term forecast to decide whether to pack an umbrella it's clear causality can be ignored - your decicion will not effect if it actually rains. However, longer term climate forecasts might (theoretically) lead us to take action on emissions which would then change the weather system. For this we need a (causal) model that allows us to predict the outcome under various different interventions.

Identifing spam and automated essay marking systems are similar. The decision made by the algorithm is likely to change the relationship between the features used by the algorithm and the true label. Spammers and students will modify their writing in order to optimise thier results. These systems can only work if the resulting change is sufficiently gradual and fresh ground truth (probably human labelled) training data is provided. (What would the nature of the features have to be such that change did not occur? - they would have to be causes of the label.

What about predicting the risk of death in patients with pnumonia? Suppose we wish to use the model to decide who should be treated in hospital and who can be sent home with antibiotics. If we assume that in hospital treatment is more effective this seems like a straighforward prediction problem. It is not. Depending on how the decision to admit was previously made and what features are included (or ommited) in the model, the relationship between those features and the outcome may change if we start using the model to decide whome to admit. (xxx et al) found exactly this effect. Their model learnt that (among other things) people suffering asthma were LESS likely to die from pnemonia. They realised this was because doctors were treating such patints very agressively, thus acutally lowering their risk. There is no problem with this model if you want to predict who would be likely to die whilst maintaining the original admition and treatment protocols. However, using it to decide who to admit could kill. The key is understanding exactly what question you are asking. In this case we are care about what happens to patients with characteristics X if we treat them according to decision rule Z.

Predicting which customers will leave or who will re-offend if granted parole also fit within the category of problems where you wish to identify a group for which a problem will occur and target some treatment to them (loyalty reward, deny parole or more support whilst on parole, etc). For all these problems the assumptions required to treat them as pure prediction problems are;

1. The treatment is assumed to be effective (at least better than nothing)
2. Deciding who to treat on the based of the model predictions won't change the relationship between features and outcome

It is easy to see how similar issues to the pneumonia example could occur for the customer and parole cases. For example, if the customer retention model is not aware of existing customer loyalty schemes or the most risky criminals are currently only given parole in exceptional circumstances (not visable to the model) so rarely offend. If people are aware of the model they may also try to game it (as in the spam/essay example) so as to recive loyalty reward or early release. TODO talk about the interesting long-term vs short term dilema introduced by the parole example. (maybe best to discuss it further on)

Demand prediction seems relatively straightforward. These models use features such as location, pricing, marketing, time of year, weather, etc to forecast the demand for a product. It seems unlikely that using the model to ensure stock is available will itself change demand. However there is a potential data sensoring issue. If demand is modelled by the number of sales, then if a product is out of stock demand will appear to be zero. Changing availability does then change demand.

Playing Go (and other games) is another case with some subleties. At every turn, the AI agent has a number of actions available. The board state following each action is deterministic and given by the rules of the game. The agent can apply supervised machine learning based on millions of previous games to estimate the probability that each of these board states will lead to a win. ... this is interesting maybe come back to it ... an alternate approach would be to try to forecast the probability of a win given each action given the current board state as context ... One approach to causal inference is indeed to learn about actions from taking actions (or observing the actions that other have taken). When can we learn from the actions others have taken? When there is no confounding. And does this hold with Go? Probably because the board state encapsulates everything that should determine what move is played. Learning directly from actions and trying to generalize (can in some instances reduce the problem to standard ML)

I hope these examples gave you a feel for the richness and subtlties of causal inference. We will return to some of them in more detail once we have established some more concrete language and tools to approach them with.

I do not see the distinction between explaination and (causal) prediction. Explaination is all about the ability to compress and to generalize. The more a model can do this, the more we view as providing an understanding of the why.

Does this definition include mediation?

Sometimes the causal component is obvious.

Relationship to generalization. Variables causally directly causally related to the outcome (either causes or effects) should be more stable predictors over time. The assumption is there are less places for change to come in.

If a feature is a cause of an outcome then changing the input distribution over that feature won't break the model. If its an effect it could.

The direct causes (and effects) of a variable of interest make up a sufficient set for prediction (is this true)? This may be a reason for using structure learning type algorithms even if you are simply doing prediction.

The role of assumptions in causal inference

Fundamental challenges. How do you cross-validate or compare causal inference models? Lack of real world data on which to compare algorithms.

Does predictive accuracy indicate a good causal model?

Assumptions must be recognised. Without assumptions - only description is possible. Recognising the assumption (and associated risk) means that we understand we should still attempt to experiment.

Generalizing the results of one experiment to another (for example, dropping rocks to dropping people - with reference to the cross over trial for parashutes.) (This I think is the long term key to successful causal inference, learning from experimental data with representations that generalize). (and given sufficient generalizability, it may not matter if there is an underlying confounder - as this confounder is clearly not changing much)

Is causal inference possible (are the conclusions ever valid)?

Only from randomized experiment ... if people can do it without recource to constant randomized experiment then an algorithm exits.

The challenges and limitations of infering causal effects from observation data are numerous. However, there are problems we need to solve. The questions is are these problems so serious as to warrent discregarding all non-experimental data. I think answer is a clear no. Need ongoing validation, with assumptions. We can't validate the assumptions using the observational data alone (otherwise we would not have been forced to make them).

Mediation

Relationship to interpretability

A desire for interpretability indicates that something has been left out of the loss function.

Are causal/interpretable models more reliable or more likely to generalize well?

One form of interpretability gives people insight into what the features are that the model is relying on.

If we know the training and test data will be sampled from different distributions, knowing what the features that the model is looking at are, allows people use their background understanding of the world to evaluate whether or not those features are likely to be transferable to the test domain.

Specifically, people can

- rule out many possible features as highly unlikely to be relevent to a problem

People have access to a lot of detailed prior knowledge.

Relationship to transfer learning

find a feature representation in which $P(Y|X)$ is the same in many different domains (or stable over time). Causal models predict the outcome of actions. We could directly take these actions and learn $P(Y|a, X)$ for every (a, X) but, in reality, no two situations (or actions) are exactly alike. So we have to make representations such that things are stable.

This is tightly related to generalizability. If we take a person undergoing a medical test, we might describe the situation by the year and location, the person's age, gender, heart rate, medical condition and test results. We don't include , the color of the doctors shirt, the size of the room, ...

For example, in the advertising setting, we want to know how our on expenditure on paid search ads is linked to sales. However, this relationship may be very unstable over time because the ad slots are sold at auction. The amount we have to pay to obtain a given position for keyword depends crucially on the amount our competitors are bidding for that keyword. However, the relationship between displaying the ad at a particular position and the probability that someone clicks it an then makes a purchase may be much more consistent.

The two approaches to causality

What is the role of randomization? How do bandits algorithms work despite being only partially randomized? What else can you do to improve randomized studies (variance reduction, lower regret).

What are the issues with standard randomized experiments?

Causal Inference

Models of causality and intervention

DAGs

Although seemingly simplistic, the notion of hard interventions is surprisingly powerful.

A complaint leveled against this view point of causality is that the 'surgery' is too precise and that, in the real world, any intervention will effect many variables (eg Cartwright 2007). However,

Counterfactuals

Structural Equation Models

A translator from graphical independence to counter factual statements.

Identifiability

The notion of identifiability is binary. How could it be softened. There are two obvious approaches. 1) Look at the finite time convergence for identifiable queries. 2) For non identifiable

queries, look at what bounds can be achieved or what additional assumptions would be required to make them identifiable.

Question - why does everyone use the backdoor criterion. Is it much more frequent in the world. Easier to estimate?

Defifinitions

Interactions

Causal effects

Estimation of causal effects

Relationship to covariate shift

Discovering causal structure

Discovery based on conditional independence

Discovery with functional models

Multi-armed Bandits

Unifying the frameworks

Interesting open questions

Cycles - a huge issue. Not covered by Pearl,Rubin etc.

Places to look, statistical control theory, etc. any interesting papers along these lines?

Data for testing causal models

Simulators. Open competitions. Converting other data sets. Existing data sets.