Learning how to act: making good decisions with machine learning

Finnian Lattimore

April 13, 2017

0.1 This Thesis

- clarify where causal inference fits within machine learing, for which problems it is required and ...

Uncategorized

http://www.news.com.au/finance/money/costs/insurance-companies-secrets-spilt/news-story/f6ef17ae73e3a56
insurance decisions voodoo because of lack of transparancy and absense of obvious causal link.
Claim: People are more comfortable with desicion making on the basis of factors they believe to be causally relevet.

Introduction (3500 words)

Motivating summary

What is causality and why do we care? (2000 words)

Defining causality

- widely debated in science and philosophy (FIND SOME REFERENCES)
- what is explaination?
- any model that aims to predict the outcome of an action or intervention in a system
- 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.
- mediation?

Identifying when we have a causal problem

Examples of typical machine learning probems. Are they causal?

- 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

What aspects of a problem determine if causal inference is required?

(When is pure prediction useful?)

- To decide between actions we only need to rank them (not estimate their actual effect).
- The predicted outcome in the absence of an intervention provides a single point. We can use this to find which problems are most serious if left alone - and priortise those for modelling changes.
- Any decision we take does not significanly impact the system from which the data was drawn to make it (for repeat decision making)
- Does acting on the result of the prediction change the predictive distribution p(y—x)? Ie change people's behaviour.

Approaches to causality (1000 words)

Two broad approaches

- Build a model to map the natural behaviour of the system to what will happen for some action
- Take the action and see what happens

The first is causal inference

The second is reinforcement learning

Both generalize from randomized experiment Reinforcment learning to sequential decisions, causal inference to non-experimental conditions

Both approaches involve assumptions the latter that we can group context and actions.

Limitations of causal inference

Limitations of experimets What are the issues with standard randomized experiments?

Causal models (5000 words)

Causal graphical models

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

Comparing and unifying the models

A translator from graphical independence to counter factual statements

Two key questions (5000 words)

Causal Inference

You are willing to assume the causal graph

extremely widely appliled Implicitly accounts 10,000 of studies in phycology, medicine, business, etc.

Identifiability

Under what conditions in is the problem solveable

The Do Calculus

Open questions within identifiablity

Estimation

How well do we actually do with finite datasets?

Causal Discovery

You want to learn the graph

Equates to the aim of automating scientific discovery

Increadibly hard

Discovery with Conditional independence

Discovery with Functional Models

Multi-Armed Bandits (5000 words)

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).

Causal Bandits: Unifying the approaches (5000 words)

Causal Inference & Machine Learning (5000 words)

A more detailed discussion on where causal inference sits within machine learning and what it can offer.

Different approaches

The two cultures

Translating terminology

Economist -¿ ML

Challenges for the Machine learning approach

No cross validation

The challenge of model selection

Does predictive accuracy indicate a good causal model?

Less large, real world datasets

The dearth of experimental data

Data for testing causal models

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

Relation to other areas of ML

Covariate shift

Generalizability

Invariants

more stable 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.

change input distribution 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.

feature selection 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.

Interpretabilty

interpretable models as proxies for causal models Let the human do the work. 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.

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

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

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.

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.

Fair Machine Learning

Structural Equation Models

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

Conclusions (1000 words)

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?