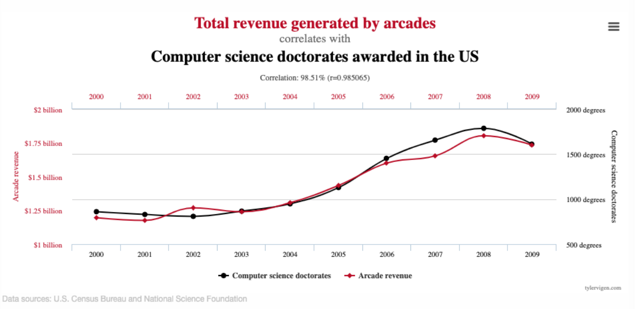
Causal Inference in NLP

Sensitivity Analysis : adjusting for unobserved confounders

Rémy Deshayes

9/1/2020

Correlation is not causation



Some other examples from *Causal Inference in Statistics*, J. Pearl :

- "Data show that people who hurry tend to be late to their meetings. Don't hurry, or you'll be late"
- "Data show that as the number of fires increase, so does the number of fire fighters. Therefore, to cut down on fires, you should reduce the number of fire fighters"

"Lucky is he who has been able to understand the causes of things", Virgil 29BC

- How effective is a given treatment in preventing a disease ?
- Could government actions have prevented the financial crisis of 2008 ?
- Can hiring records prove an employer guilty of gender discrimination ?

What's interesting is that those questions cannot be answered or articulated in the traditional language of statistics and that's why causal inference drew a lot of attention those past 15 to 20 years

Why is it important to study causal inference ?

Causal inference answers two very important questions :

- Do changes in one variable cause changes in another,and if so, how much change do they cause ?
- How a causal link can be established ?

Simpson's Paradox - 1951

A group of 700 sick patients are given the option to try a new drug. 350 in control group and 350 in treatment group

	Drug	No Drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Table 1 – Results of a study into a new drug, with gender taken into account

Simpson's interpretation :

- What if estrogen had a negative effect on recovery regardless of the drug ?
- If one select a user at random the person is more likely to be a woman
- Being a woman is a common cause of both drug taking and failure to recover

A few letters and easy formulas

- A population with n individuals
- **Features** of individual i is O_i
- **Treatment** status of individual i is T_i
- **Outcome** for individual i is Y_i
- Propensity score for individual i is $P(T_i|O_i)$

- Treatment Effect is : $[Y|T=1]-[Y|T=0]$
- and Average Treatment Effect is $\frac{1}{n}\sum_{i=1}^{n}\{[Y_i|T_i=1]-[Y_i|T_i=0]\}$

The fundamental problem

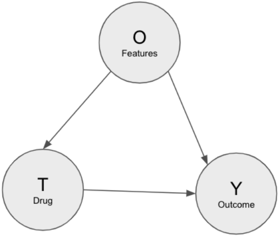
We never have both the treated and untreated "version" of each person !

How can we possibly know both $[Y_i | T_i = 1]$ and $[Y_i | T_i = 0]$?

Observational Studies

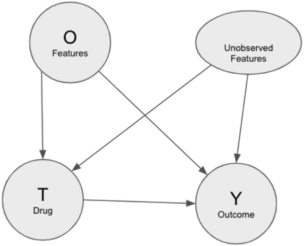
Groups are designed by the researcher - Goal is comparability

Introducing the confounders in a causality graph



A lot of different techniques on how to compares groups exist but here we think about what if there is something else ?

What is sensitivity analysis ?



The presence of unobserved confounders in observational studies casts doubt on each and every causal inference result they can produce

Sensitivity analysis is trying to assess how important those unobserved confounders could be when looking at some causal inference results

2 approaches for sensitivity analysis

1. **Statistical approach** : Finding the thresholds of the association(s) between the unobserved confounder and the outcome / observed confounders that would render the test statistics of the study inference insignificant
2. **Epidemiological approach** : Quantify the unobserved confounder to know how much our results are biased and therefore be able to get the true ATE

What are our goals

Context : Evaluate how a user responded to some compassionate comments of other users - reddit mental support thread

Initially, We wanted to answer 2 questions in this specific context :

1. How worried do we need to be about our results in general ? How quickly do the results breaks apart ?
2. How different models interact with sensitivity analysis ? Where do models disagree ?

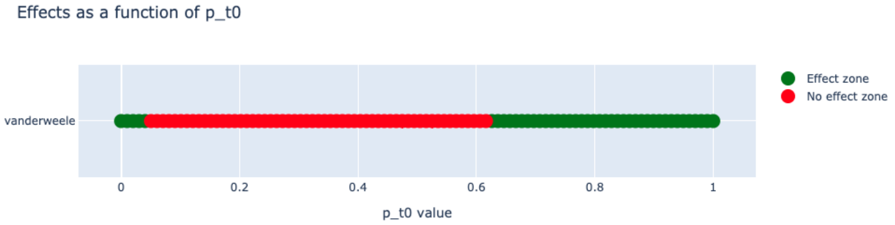
What are the challenges

- Sensitivity analysis is rarely if ever used in the CS/NLP community compared to the statisticians community because models are often very difficult to use on real data
- Sensitivity analysis often rely on a lot of different assumptions that need domain understanding and that are complicated to implement in an efficient way
- Unclear how to apply sensitivity analysis to high dimensional unstructured text data sets

What we did

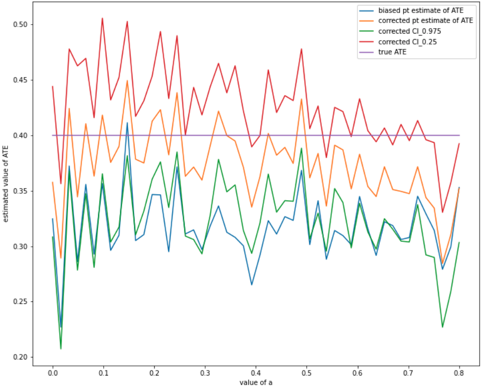
We generated synthetic and semi-synthetic data and implemented different sensitivity analysis techniques in an oracle infused setting

An example of how we dealt with one of the challenges : methods' prerequisites, namely assumptions on different parameters



Other outputs

	upp_p_value	test_significance
Rosenbaum	0.720969	No



Where we are now

Along the way, new questions came up :

1. How can we come up with a good heuristic of an unobserved confounder ? It is unobserved and unknown after all !
2. How to make an efficient trade-off between complexity of the model (some models can go up to thousands of inputs) and truthfulness of the results