

Combining Causal Discovery with CausalNLP

Praxis der Forschung WS 2021/22

Scope

1 – 4 students. The content can be extended if more participants.

Background

Causal discovery is a branch of *causal inference*, a subfield shared by many areas, such as medicine, epidemiology, social science, economics. Its goal is to derive causal relationships between observed variables in a form of causal graph, when experiment is infeasible. Since causal discovery methods are mostly unsupervised, experts are required to evaluate the resulting causal graph, especially when the matter is complex. This creates a bottleneck in model selection and evaluation.

Causal NLP involves inferring cause and effect from text. One of its subfield aims to extract knowledge in a form of causal graphs. Although there is still a research gap in dealing with implicit causality, the methods can potentially be used for laymen to evaluate causal graphs derived from observational data for complex matter. Users can verify the causal relationship by reading the relevant text.

It is to be expected that causal graph derived from causal discovery methods and from causal NLP methods will not coincide. From the causal discovery side, this could be due to hidden variables, unmeasurable variables, wrong assumptions on relationship type, leading to inadequate method selections, or noisy data. From causal NLP side, it could be due to false interpretation of implicit causality, text quality or disagreement among experts. It could also be the case that experts have always been wrong because they have never seen the data as a whole. Therefore, strategies must be developed to reliably resolve these conflicts without removing too many interesting hypotheses worth further investigation.

Goals

The goal of this project is to explore and test strategies to reliably combine resulting causal graphs from causal discovery and causal NLP.

Tasks

1. Literature reviews in combining graphs from different data sources and online machine learning
2. Search and create data sets that will allow for experiments to evaluate strategies
3. Implement the developed strategies
4. Design the experiments and evaluate the strategies

Contact

Ployplearn Ravivanpong, ployplearn.ravivanpong@kit.edu