Name: Bharghav Srikhakollu

CS549 – ST/Scalable Querying & ML – Spring 2022

Topic: Casual Relational Learning

https://dl.acm.org/doi/pdf/10.1145/3318464.3389759

What is the problem discussed in the paper?

For scientific discovery and informed decision making, casual inference is very important in natural and social sciences. The key factor in casual inference is to perform randomized controlled trials but this is not possible due to ethical, legal, or cost constraints. Alternative for this problem is to use methodologies in casual inference from observational data. Since these methods are dependent on data being homogeneous, but in many real-world problems the actual data is heterogeneous with complex relational structure where data is present across multiple tables.

Why is it important?

Since causality goes beyond correlation and predictive analysis it is important to understand on heterogeneous data. The approach adds on to the existing casual inference literature by relaxing the unit-homogeneity assumption and allowing the confounders, treatment units and outcome units to be of different kinds.

The approach is easy to use for the applications in industry and academia. Also, the approach is successfully able to recover the treatment effects for complex casual queries that may require multiple joins and aggregates.

What are the main ideas of the proposed solution for the problem?

A framework called CaRL is built which is a Casual Inference framework for Relational Data. In general, the two main approaches to Casual Inference are Rubin's potential outcome model and pearl's graphical casual model. The principle behind both the approaches is Interventions. The casual inference is a missing data problem and the possible under some assumptions is the background knowledge. But they need homogeneous units in a single flat table.

The framework works on a relational database, background knowledge and build a casual query. Using these single flat data table is created which do the casual effect estimates. Background knowledge syntax is like relational casual rules. The process steps in the methodology are as below.

Skeleton Traversal → Grounding → Confounder Identification → Summary Functions → Casual Inference.

Reference citation: https://dl.acm.org/doi/10.1145/3318464.3389759