- Chairs have requested users to enter domain conflicts. Please click here (/SIGMOD2020/User/ConflictDomains) to enter your conflict domains.

View Reviews

Paper ID
603

Paper Title
Causal Relational Learning

Track Name
Research Paper Second-Round

Reviewer #1

Questions

1. Overall evaluation
Weak Accept

2. Reviewer's confidence
Some Familiarity

3. Novelty
Medium

4. Importance: select all that apply:
SIGMOD attendees will learn something interesting from the paper
The paper is likely to influence other research in the community

5. Summary of contribution (in a few sentences)
This work presents a formal framework for causal inference from populations with homogeneous units, which is best represented as relational data. The framework includes a declarative language (CaRL) for specifying causal queries over such data, and an efficient algorithm for answering causal queries over the relational data. The semantics of the causal queries are provided by a non-trivial extension of Pearl's do-operator.

6. List 3 or more strong points, labeled S1, S2, S3, etc.
S1: Overall, the paper is very well-written and was logically organized.

S2: The experimental results on the real review data are very interesting.

S3: The paper did a good job of providing examples to help elaborate on the very technical material.

7. List 3 or more weak points, labeled W1, W2, W3, etc.
W1: One of the main contributions of this work is "a non-trivial extension to Pearl's do-operator to give semantics to complex causal queries." The actual extension is not called out explicitly anywhere in Section 4, or at least explicit enough for a reader with limited familiarity of Pearl's do-operator to immediately appreciate.

W2: Section 5.1 is extremely hard to follow, even with Example 5.3 and Observation 5.1. Particularly Equations 29 and 30 could benefit from some elaboration of the notation used.

W3: For Section 5.2, the statement "Now all the causal queries in Section 4 can be estimated using standard statistical techniques for estimating (34) using the unit table D". Could you please elaborate on those, and whether other assumptions made elsewhere in the framework impact the validity of applying these techniques.

W4: The very high variability of the CATE values for the embeddings considered requires a little more elaboration, specifically regarding what would the practical implications be for interpreting results.

8. Detailed evaluation. Number the paragraphs (D1, D2, D3, etc.)
D1: Figures 4, 5 and 6 would be easier to follow if they had the same structure with the common nodes in the same location in the figure. For example, to go from Figure 4 to Figure 5, keep the nodes from Figure 4 in the same location but add the nodes and edges to obtain Figure 5. Same thing for going from Figure 5 to Figure 6. That would make examining the differences between the 3 figures easier.

D2: "Eva" appears to have became "Eve" in the text at some point. Also, Figure 10 caption has typo "sensitivity of of the"

D3: In Section 4.1, "where the intervention replaces the NSE associated with...", NSE is not defined anywhere. Is NSE a typo for NDE (natural direct effect)?

D4: In Example 5.3, should $\Phi^{score}_{qualifications}$ be $\Phi^{quality}_{qualifications}$. If not, then please elaborate where $\Phi^{score}_{qualifications}$ comes from. It isn't included in the other examples anywhere (i.e., Figure 6).

9. Candidate for a revision?
Yes

10. Required changes for a revision, if appropriate (labeled R1, R2, R3, etc.).
R1: Please address W1 - W4 and D1 - D4. There is a full page available in the current submission, so additional material could be added without revising the current submission for length.

Reviewer #2

Questions

1. Overall evaluation
Weak Reject

2. Reviewer's confidence
No Familiarity

3. Novelty
Medium

4. Importance: select all that apply:
SIGMOD attendees will learn something interesting from the paper

5. Summary of contribution (in a few sentences)
The main contribution of the paper is a declarative language to specify models/queries for casual inference. The paper defines the syntax of the language as well as its semantics. It gives an experimental study on an academic

dataset ReviewData and a synthetic dataset of similar structure as ReviewData

6. List 3 or more strong points, labeled S1, S2, S3, etc.
S1. Casual inference has seen a resurgence in the last few years with the publication of "The Book of Why," perhaps the most notable advocate of it is Judea Pearl. From this perspective, the paper tackles a timely topic.

S2. The paper is theoretical exercise. Some parts are easier to follow than others.

S3. The proposed language appears to allow inference over heterogeneous units with different types of relationships.

7. List 3 or more weak points, labeled W1, W2, W3, etc.
W1. It is unclear who the target audience of this language is.

W2. The Review Data, while an nice case study, is not nearly convincing.

W3. There is a great deal of effort put into the theoretical underpins of the proposed language, but there is little effort put into characterizing its properties: learning curve, efficiency, transparency, ease of communication.

8. Detailed evaluation. Number the paragraphs (D1, D2, D3, etc.)
D1-W1. It would be nice to have a discussion about the target audience of the proposed tool and whether it can be integrated with existing tools.

D2-W2. Most of the work on casual inference is in health/medical environment. While the authors mention such scenarios, none is considered. The review data is nice example, but not nearly convincing. It is unclear the motivation for the synthetic data since the data mirrors real life one. Perhaps, including other datasets, unrelated to review data, would help understand the values of this language.

I feel that a relevant, but missing, study is a comparison with a "baseline" of some sort. While reading the paper and trying to understand all the theoretical nuances, I could not by think of: How are things done nowadays?, How does CaRL help a bioinformatician or medical researcher?, What tools do those people use?, Would their tasks be improved with CaRL? But, I could not answer any of these questions while reading the paper.

While the topic is the bread and butter of statisticians/data miners/ML, there is little interest in the DB community.

9. Candidate for a revision?
Yes

10. Required changes for a revision, if appropriate (labeled R1, R2, R3, etc.).
Please see W1-W3 and the comments in the detailed evaluation.

Reviewer #4

Questions

1. Overall evaluation
Weak Reject

2. Reviewer's confidence
Some Familiarity

3. Novelty
Medium

4. Importance: select all that apply:

None of the above

5. Summary of contribution (in a few sentences)

This paper presents a framework to perform causal inference over relational data. Their framework is composed of a query language CaRL that define queries as rules with conditions and an inference engine that answers such queries. They also propose an extension to the do- operator for complex queries. The full paper is driven by a motivational example in the academic publication domain

6. List 3 or more strong points, labeled S1, S2, S3, etc.

S1. The causal relational learning problem is interesting.

S2. The proposed framework is sound and coherent.

S3. One running example to explain the concepts is very useful.

7. List 3 or more weak points, labeled W1, W2, W3, etc.

w1. The paper can be structured better. E.g section 3 and 4 should be merged such that formal and informal definitions help each other. The explanation of different embedding should be also out of the experiment section and into the explanation parts.

W2. Poor experiment section. a) Only three queries are evaluated, one for each case. There authors should improve the number of queries used as well as a variety in complexity of the different queries used. b) different experiments use constraint and non-constrained settings without a valid explanation on why is one used over the other. c) There is no good comparison with other state of the art relational causal inference methods that are cited in the related work. for instance what datasets do they use? how does this method compare to others in their datasets? how do the other methods function on your new dataset?

W3. Adaptability to different domains. the papers starts with a motivation on the social sciences, but the method, dataset, examples and tests are all based on publication review process. The authors lack of explanation on how is this would apply in other domains or experiments such as using a different domain dataset. For instance the medical domain using MIMIC dataset of ICU patients would be an interesting relational dataset that is large and would really showcase the ideas exposes.

8. Detailed evaluation. Number the paragraphs (D1, D2, D3, etc.)

D1. Figure 1 is unnecessary.
D2. Example 1, Delete Significantly.
D3. Related work and Background analysis are not clear enough about what is the difference between other causal models from relational data and this work other than previous works being based on Statistical Relational Learning.

D4. Range(A) in 3.1 ?
D5. Are attributes and Attribute functions the same? after some paragraphs A's are referred to as Attributes only

D6. ATE is defined in section 2. and then again mentioned in section 3.3 without being explicit about the difference from related work and this work.

D7. In Page 6 the paragraph that explains relational queries is difficult to read. I don't understand the part "computes both isolated and relational effects of Prestige on
reviewer scores if all coauthors from prestigious institutions." Is not clear the difference between relational and isolated. Is the condition the difference? how so? query (15) computes both? This question is resolved in section (4) with further explanation. This is why the structure should be adjusted.

D8. 4.2 how does this work when the intervention variable is not binary? for example prestige is a [0,1]?

D9. Definition 4.1 has redundancy.

D10. (30) is an equation not a theorem and Theorem 5.2 is not proved

D11. The Embedding concept and methods are fuzzy before reading the experimental section (6). It's mentioned in section (3) that it will be further explained in section 6 (experiments?) it should be discussed before experiment.

D12. In the paragraph "By mistakenly interpreting the naive correlation as causation, one might arrive at the conclusion that double blinding is not effective in reducing bias " where is this stated or proved by actually applying correlation instead of causation?

D13. section 6. Q1: Is there a theorem 4.3? you mean proposition 4.3?

D14. why is experiment Q1(b) on synthetic and not real data?

D15. why figure 7 is on CATE and the explanation is on ATE?

9. Candidate for a revision?
No