"Learning" Causal Reasoning

Rohit Saikrishnan Ramesh
Computer Science, UIC
Chicago, USA
rrames7@uic.edu

Lingfang Hu
Computer Science, UIC
Chicago, USA
lhu7@uic.edu

KEYWORDS

meta-learning, reinforcement learning, causal inference, neural models

1 ABSTRACT

Causal Reasoning is a very interesting field that has gained attention over the recent years. Reasoning is an important trait of intelligence that helps humans make decisions quicker, faster and more importantly accurate. In the project, we explore tools and techniques that combine both Artificial Intelligence and Causal Reasoning literature together. Further, we evaluate these techniques on how robust they are in their claims of defining "learnability" with these techniques.

We explored two papers that exploit machine learning and AI techniques for learning causal graphs[4][1]. [1] introduces a MetaRL framework for estimating the causal models. We explore the MetaRL on dependant bandit problems to see the influence of MetaRL in learning.[4] uses neural model for determing causal models from unknown intervention. Neural models have shown huge success in getting accurate results in prediction and regression. We explore these in a causal model setting and finally discuss the details of the architecture if it matches with what we define as "learnability".

We observe that MetaRL frameworks help in faster learning algorithms by choosing right parameters for exploration and exploitation. The Neural Models are successful in determining the interventions in causal models. The architectural design puts serious question for generality of learning.

The dataset design for conducting the experiment in [1] seems inadequate for the claim of "learning". We discuss in detail the pros and cons of both approaches and suggest modified approaches to our knowledge can get us closer to claiming "learning" in causal reasoning.

2 INTRODUCTION

Artificial intelligence has attracted more and more attentions of researchers, which aims to design intelligent agents who can learn to complete many complicated tasks like human beings. Causal reasoning may be one of the most essential components of human intelligence. We often want to ask questions such as "What if I put more efforts in studying one course?" and "Will taking the new drug improve the chance of recovery for patients?". To answer such types of questions, intelligent agents are required to be able to reason about causes and effects.

Although some formal causal reasoning algorithms has been established previously, they have pre-conceived strong model assumptions, which are model-dependent [6]. In many cases, the model assumptions are not well-matched to the reality. Thus, we want to achieve such reasoning tasks by model-free meta-learning, which is to learn the learning/inference procedure directly from data[2]. Via end-to-end meta-learning, the algorithm will have the potential to mine the causal structure representations which are adaptive to specific tasks. Besides, reinforcement learning enables the agents to learn the causal reasoning not only from passive observations but also from active interactions with the environment [5]. So finally we will adopt the "meta-reinforcement learning" method introduced previously [2] to help us study the topic.

Another well established technique in the field of Artificial Intelligence is the neural network models [7], These techniques have shown immense success in the prediction and regression tasks. It has become a common notion in the community that these neural network architecture are an analogous to the information processing that happens in the human brain that enables us to think intelligently. Interventions in the causal graphs are an important concept that helps us understand treatment effects on the causal structure. Determining these interventions through a learning paradigm like neural networks is the right step to understand intelligence in generality.

3 RELATED WORK

On the one hand, there is a rich literature giving formal definition of causal reasoning and performing causal reasoning with formal approaches. Pearl et. al give an comprehensive introduction to causal models, causal reasoning, and causal inference [6]. Generally, we use a acyclic directed graph (DAG) to represent a causal model, where each node is a random variable and a directed edge directing from a parent node to a child node encodes a direct causal relationship. For those DAGs with some specific structures, they develop effective tools to estimate the causal effects between two nodes. In these formal approaches, strong parametric assumptions are imposed on the model structures.

On the other hand, in machine learning and artificial intelligence, deep learning and reinforcement learning algorithms become more and more popular and achieve surprisingly good performance when there is enough data. Both of deep learning and reinforcement learning are end-to-end and thus model-free learning algorithms, where the learning process is directly driven by the target task without strong parametric model assumptions. There is a major limitation for traditional deep reinforcement leaning: the demand for massive amounts of training data. Thus, some researchers propose the meta-learning mechanism, which can adapt rapidly to new tasks. Previous work has shown that recurrent neural networks (RNN) can support meta-learning in a fully supervised context. Wang et al. extend the RNN framework to the RL setting and then propose the deep meta-reinforcement learning approach [8], whose recurrent dynamics implement a second, quite separate RL procedure. After training on a large family of structured tasks, the algorithm can easily generalize to new tasks drawn from a similar distribution.

The successful implementation of [8] has pushed the researchers to explore these techniques for causal reasoning tasks. Dasguta et. al train a RL based agent on DAGs of N=6 and further train a MetaRL based engine which allows for faster learning in graph of N=5. They segregate the graphs based on the structure and define a test set for evaluating their approach. The agent performed better than a random agent in choosing the right nodes to seek useful information for building the causal model.

Apart for these methods to estimate the causal graph structure. Researchers in the recent past have looked to model the causal structure in graphs from dataset of multivariate distribution. These techniques are popularly known as causal structural learning[3]. LINGAM and BACKSHIFT are two algorithm which show high performance estimating the causal structure. This techniques however, do not use Artificial Intelligence techniques to come up with the results. Although these techniques are very useful their architecture do not allow us to mimic intelligent agent and give us understanding about intelligent agents.

Rosemary Ke Et. Al[4] come up with neural networks based Multi Layer Perceptron model that takes huge data as input and learns the causal structure existing within the data. The learned model is then applied on test set to predict for the edge weights in the causal graph structure. This method is closely tied to Artificial Intelligence techniques. Successful results from this techniques gets us to one step closer in understanding how intelligent agents model causality structure while making their decisions. They also describe methods to predict interventions in the structure by learning representations from MLP models. As discussed earlier estimating interventions through a neural networks framework is a good model for intelligence.

4 PROBLEM DESCRIPTION

4.1 Meta Reinforcement Learning for Causal Reasoning

The authors[1] want to design agents who can do causal reasoning in three distinct data settings: observational, interventional, counterfactual. In each kind of setting, the agent need to do different reasoning task. For example, in observational setting, the agent only has access to observational data P(X) and $P(X \setminus X_i | X_i = x)$ and is expected to infer $P(X \setminus X_i | do(X_i = x))$. The authors want to answer two big questions based on the experimental results:

- Do the agents learn to perform cause-effect reasoning using the available data?
- Do the agent learn to select useful information

Technically, the authors use the recurrent RNN-based metareinforcement learning framework to learn from causal Bayesian Networks (CBNs) structural data. In each CBN, there are N nodes and edges between nodes are set randomly to generate different types of causal graphs. For evaluation, they train and test different types of agents in each kind of data setting: Active Agents V.S. Optimal Associate Baseline Agents for answering (Qa), and Active Agents V.S. Random Agents for answering question (Qb).

As a result, on the one hand, the authors claim that if the selected node X_i as parents, then the Active agent will outperform the Baseline agent. They think that such comparison result demonstrates

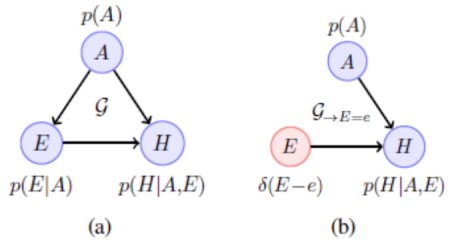

Figure 1: An example of causal graphs in MetaRL setting

the Active agent having causal reasoning ability, that is, the answer to question (Qa) is positive.

On the other hand, Active Agents achieve better performance than Random Agents, which seems to show that the answer to question (Qb) is also positive.

	$P(X \setminus X_{i0} \mid X_{i0})$	P(X)	$P(X \setminus X_{i0} \mid \forall X_{i0})$
Active Agent	Y		
Baseline Agent		Y	
Random Agent			Y

Table 1: Training data settings for different agents. Here, $P(X \setminus X_{i0} | X_{i0})$ means that the node is chose by the meta-learning mechanism at each iteration. $\forall X_{i0}$) means that the node is chose randomly.

4.2 Causal Neural Models for Unknown Interventions

The authors look to design a neural network model that works well in generality across broad class of Structural Causal Models. But for this experiment setup. Authors restrict the model for smaller class of Structural Causal Models to evaluate the results of the proposed approach.

Authors of the paper make the following assumption in their experimental set up:

- **Data is discrete valued:** By this assumption, we restrict values for all random variables in the graph to discrete values. This reduces the class of SCM for study
- **Data is fully observed:** The data for all the random variable for the graph is available is not hidden.
- **Interventions are Sparse:** Interventions affect only one single random variable. The random variable that gets affected is not known to us. This assumption according to authors is valid since, for any particular agent it is implausible to have broad subset of coordinated interventions.
- **Interventions are soft:** This ensures that interventions do not cause the values to be pinned to a single value as a necessity.
- **Interventions do not stack:** What this means is that before a new interventions occurs the earlier intervention is retracted and then the new intervention is studied

- **No control over interventions:** This ensures that the black box algorithm has no control over the target variable or the nature of the intervention.

With the above mentioned assumption, the problem set is skewed. We look to see if the proposed neural network solution works best in determining intervention and estimating the structural causal models.

5 SOLUTIONS

5.1 Meta Reinforcement Learning for Causal Reasoning

The meta-reinforcement learning paper [8] introduces techniques to better learn policies in reinforcement learning that ensures that optimal policy is learned within few episodes of the new task. Figure 2 gives a pictorial representation of the setup

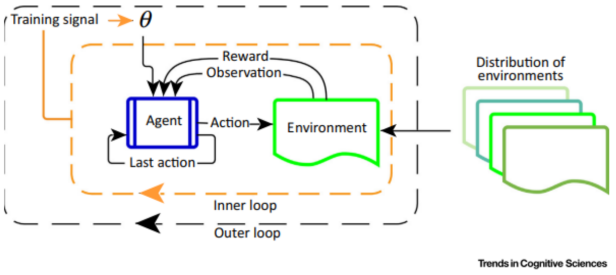

Figure 2: Description of Meta Reinforcement Learning setup

The techniques proposed in the paper look to better model the A3C reinforcement learning algorithm by introducing initial RNN layers that better make the decision of exploitation and exploration.

Firstly, we look in depth the A3C reinforcement learning algorithm. The A3C Reinforcement Learning algorithm [2] consists of components that bind to make the learning more robust. The components are described below.

- **Actor-Critic Network:** This component consists of the main Actor-Critic Policy. Unlike the policy gradient methods like Q-Learning. Actor-Critic is more efficient in intelligently update the weights which lead to a robust policy. This is achieved by forming a network of layers in the component where critic part reward the actor based on its choices.
- **Worker Nodes:** Unlike the Q-learning, where a single agent makes runs in one environment. In A3C, Each worker node makes a copy of the network as well as the environment and then feeds back the update weights to the master node of AC Network. This makes the working of the algorithm more efficient.

Figure 3 gives a detailed pictorial view of the A3C algorithm setup and how it works. This will help us in better understanding the algorithm.

The major component of the set up is the MetaRL units. The most important decision in a reinforcement learning setting is the exploration vs exploitation tradeoff. The RNN units are installed on top of the A3C setting which enables better choice for this tradeoff. Figure 4 gives us a better picture of the units.

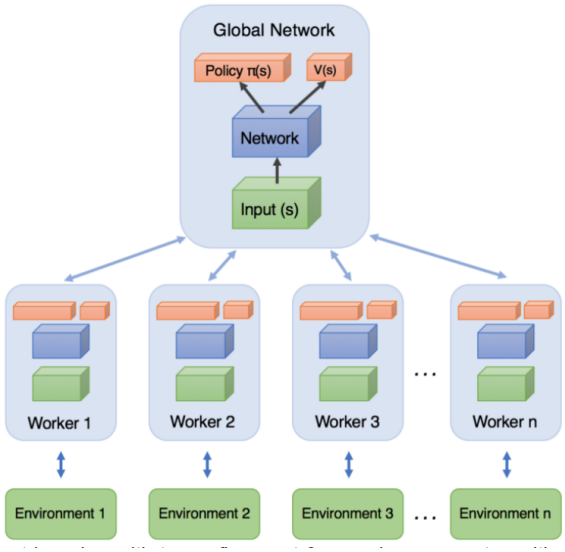

Figure 3: Pictorial Depiction of A3C Algorithm.

As an initial setup for this approach we test the approach on dependant bandit problem. Where we see the differences in performances for non-MetaRL and MetaRL approaches.

Further, we look to train causal graphs as MDP as next step and perform testing on a new dataset of causal graphs that are similar but are not trivially interpreted by other algorithm. Successful implementation of such a setup would be the solid step in claiming "learnability" in causal reasoning through meta-reinforcement learning.

5.2 Causal Neural Models for Unknown Interventions

The proposed neural model method is based on score, iterative and continuous optimization. The method is divided into phases. Wherein, in each phase certain training tasks are done that enable the model to estimate the causal structure in the underlying graphs. Figure 4 shows the iterative phase approach for the proposed algorithm.

Existence of prior knowledge based graph may speed up the model to learn the causal structure faster and efficiently. The model follows parameterized approach for beliefs during the graph generation for the synthetic dataset. There exists a prior which is responsible for generating the graphs intended for the ground truth synthetic data. We discuss each phase of the algorithm in detail here.

Phase 1: Graph fitting on observational data In this phase, the model learns to fit to the likelihood of having this particular graph generated from randomly generated graph with a belief prior. Ensemble of neural models are used to get a goof fit of causal model structure. The neural model is depicted in figure 6.

Phase 2: Graph Scoring on Interventional Data In this phase, we would like the model to learn interventions on causal model. This phase follows a 3 step approach where the in the first step

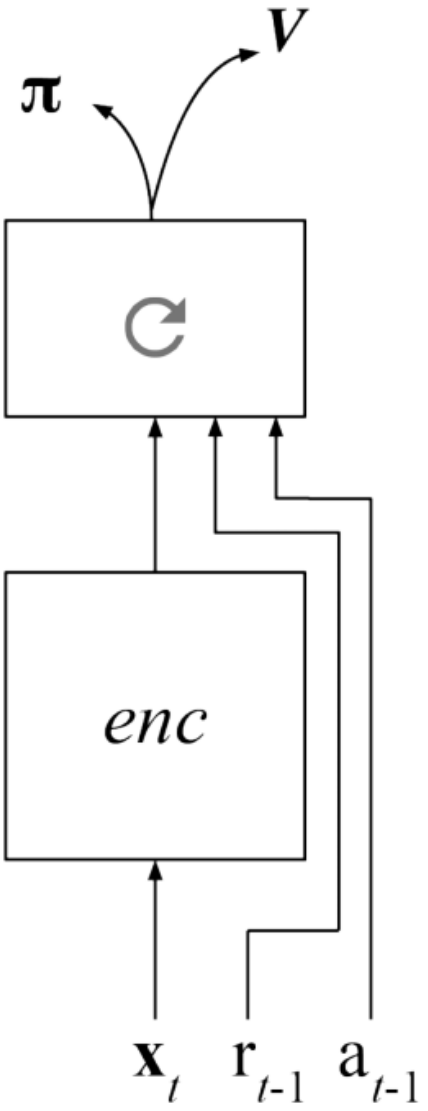

Figure 4: Pictorial Depiction of MetaRL units.

the interventions are chosen in random and the target variable is selected at random. in the next step, if intervention is not known then the intervention is predicted. in the third step, for the random graphs generated the interventions are predicted and then parameterized scoring is done to evaluate the model.

Phase 3: Credit assignment to structural parameters The most import part of the task is to predict the parameters of the structural causal models. these parameter when predicted need a good scoring rule for improving the model. This phase assigns credit based on the prediction of the structural parameters.

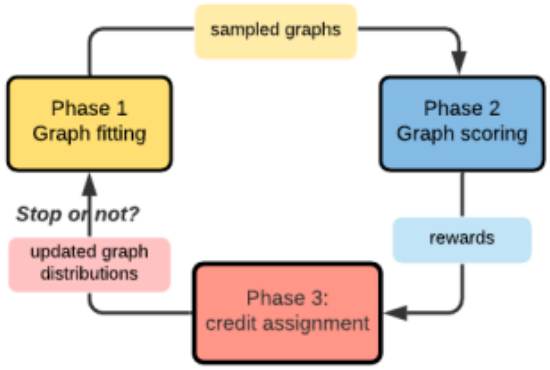

Figure 5: Pictorial Depiction of phases in the proposed SDI approach

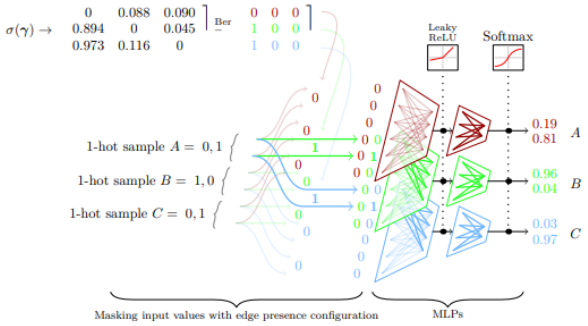

Figure 6: Ensemble Neural Models for SDI

6 DATASET DESCRIPTION

6.1 MetaRL Dataset

As a first step, we design a dataset to see the results for dependant bandit problem using MetaRL setting. For this purpose, we develop three types of arm bandits as follows:

- **[0.9,0.1] bandit:** here either of the arms have the probability of 0.9 for reward and the other as 0.1.
- **[0.6,0.4] bandit:** here either of the arms have the probability of 0.6 for reward and the other as 0.4.
- **[0.25,0.75] bandit:** here either of the arms have the probability of 0.9 for reward and the other as 0.1.

We explore our methods on this problem to see how well the algorithm performs for these tasks. Success here would help us to extrapolate to causal models.

The authors of the paper create synthesized dataset where the graphs have nodes N with N = 5. The edge weights of the graphs are {-1,0,1} distributed in random.

The graphs are then segregated into equivalence class that is the the graphs of one class are similar in structure but vary in the permutation of distribution of edge weights. 12 graphs are chosen

in random and all the graphs that belong to the equivalence class are chosen as the test set.

It is here that we intend to change the choice of the test set and analyze the results of the methodology. What truly claims as learning is to see performance gain when the test set is chosen as graphs with different structure or graphs with different number of nodes. For example, if the methodology when trained on N = 5 performed well when tested on graphs with N = 3 that would be a good claim of what we can define as learning. We intuit this might require a slight change in methodology design to include this proposed model of what we claim as learning and also that graphs are not manually fed.

6.2 Causal Models for Unknown Interventions

The graphs are represented as a one hot vector model with vertices only in the lower triangle to ensure that the dataset consists of only Directed Acyclic Graphs.

These graphs are generated based on a prior distribution on values of the random variable and also the edges. For training the interventions, The Graph dataset is such that each intervened variable has only one target variable. For our experiment 10000 graphs are sampled to have the best estimation in the neural models.

7 EXPERIMENTAL SETUP

7.1 Experimental setup(MetaRL)

As a first step towards building an MetaRL system we test the approaches in [8] on a dependant bandit problem and provide results of its implementation. These results shows the significance of MetaRL system in learning newer task and interpolating this approach to causal reasoning feels like the right best step forwards for our claim of "learning" in causal reasoning.

7.1.1 Modelling the Dependant Bandit as MDP.

- **State:** The history of all the actions taken in the previous episodes is defined as the present state. The initial state of the MDP is an empty set of all action taken.
- **Actions:** Action is defined as pulling the lever "Left" or "Right".
- **Transition Probability:** Transition Probability for our problem is the probability with which an agent chooses lever "Left" or lever "Right".
- **Reward Function:** At every episode one of the lever has a positive reward "1" and the other lever has a reward of "0". Reward Function return the reward for pulling that lever.

After modelling this as an MDP, we run the system wherein one of the model is MetaRL free and other uses MetaRL for optimization. We evaluate the performance of the system on both these settings.

7.2 Experimental Setup (Neural Models)

We first train our MLP based neural models on synthetic ground truth dataset. After training on variants of such models. we test that model on new synthetic dataset. We evaluate the model based on the hamming distance of values of variables which were different from the predicted.

Later, we use these models for predicting interventions. We generate 20 graph dataset as test set where each of the graph is intervened by a different variable. The correctness of the prediction is finally reported.

8 RESULTS

8.1 Results(MetaRL)

We compare the mean value of the experiments performed on bandit problem with a MetaRL system and without a MetaRL system. The Results show that MetaRL system performed better than the system without on a new task. The Table 2 describes the results of the experiment.

From the experiment we say significant boost in performance by introducing a MetaRL agent in the system. This serves as an inspiration than interpolating such a technique for causal reasoning is good step towards describing learning in causal reasoning techniques.

	Mean Values
MetaRL	1.034
Non-MetaRL	0.495

Table 2: Perfomance comparision of having a MetaRL agent vs not having a MetaRL agent in dependant bandit problem.

These results show a great promise that MetaRL would be a great direction forward in modeling intelligence.

8.2 Results(Causal Models)

We evaluate the performance of observation data by considering the hamming distance. We observe that when tested the averge hamming distance across the dataset was 4. Further we test our model in the ability to predict intervention.

The model when applied to interventional data. The accuracy of the model was 80% on test graphs.

	Hamming Distance	Accuracy
Observation Data	4	NA
Interventional Data	NA	80%

Table 3: Perfomance comparision of having a MetaRL agent vs not having a MetaRL agent in dependant bandit problem.

9 DISCUSSION AND IDEAS FOR NEXT STEPS

The work done on the project has given us a complete picture on potential directions we can take for claiming "learning" in causal reasoning environment.

Both methods show promising directions. However, the experimental setup is skewed that it lacks the ability to extrpolate the claim for general "learnability" of causal models.

These are interesting directions which were explored. An ensemble model of model-based and model-free strategies will help

us get more closer to the claim for "learnability" of causal models. Listed below are some steps we wish to take that can help us solve the problem.

(1) Replicate the Meta-Learning procedure for the causal inference setting as described in the paper
(2) Analyze the results and build intuitions for why the system performed the way it did. This shall help us better design the experimental setup or choice of dataset
(3) Bringing the intuitions into real experiment design to confirm our understanding
(4) Carefully choose the test dataset and evaluate its results to solidify our claim for "learning" by the algorithm or otherwise.

10 GITHUB LINK FOR CODES

The dataset generation and experimental setup is enclosed in this github link.

https://github.com/rohitsaikrishnan/-Learning-causal-reasoning

REFERENCES

[1] Ishita Dasgupta, Jane Wang, Silvia Chiappa, Jovana Mitrovic, Pedro Ortega, David Raposo, Edward Hughes, Peter Battaglia, Matthew Botvinick, and Zeb Kurth-Nelson. 2019. Causal reasoning from meta-reinforcement learning. *arXiv preprint arXiv:1901.08162* (2019).
[2] Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. 2016. RL^2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779* (2016).
[3] Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. 2018. Causal structure learning. *Annual Review of Statistics and Its Application* 5 (2018), 371–391.
[4] Nan Rosemary Ke, Olexa Bilaniuk, Anirudh Goyal, Stefan Bauer, Hugo Larochelle, Bernhard Schölkopf, Michael C Mozer, Chris Pal, and Yoshua Bengio. 2019. Learning neural causal models from unknown interventions. *arXiv preprint arXiv:1910.01075* (2019).
[5] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. Explainable reinforcement learning through a causal lens. *arXiv preprint arXiv:1905.10958* (2019).
[6] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer.* John Wiley & Sons.
[7] Marcel Van Gerven and Sander Bohte. 2017. Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience* 11 (2017), 114.
[8] Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. 2016. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763* (2016).