# Towards Continual Learning of Causal Models

**Anonymous submission**


### Abstract

A common assumption in causal modelling is that the relations between variables are fixed mechanisms. But in reality, these mechanisms often change over time and new data might not fit the original model as well. But is it reasonable to regularly train new models or can we update a single model continually instead? We propose utilizing the field of continual learning to help keep causal models updated over time.


## Introduction

Causal models (Pearl 2009; Peters, Janzing, and Schölkopf 2017) can be useful in a variety of applications (Koch, Eisinger, and Gebharter 2017; Carriger and Barron 2020; Fenton et al. 2020). Various algorithms exist for causal discovery, i. e. learning the causal graph from data (Glymour, Zhang, and Spirtes 2019). With the causal graph known, causal model can be trained, enabling a causal perspective on the respective problem. Due to a lack of knowledge about unobserved variables, the relations between effect(s) and cause(s) are usually not deterministic, instead relying on probabilities for making predictions.

Continual learning is another research direction which in recent years has attracted the attention of an increasing number of people (based on the number of publications (Mundt et al. 2022)). While we are not aware of any universally accepted definition, we would describe continual learning roughly as updating a model over time, given new information and without losing the *knowledge* encoded in the original model (Mundt et al. 2022; Chen and Liu 2018).

The general idea behind this paper is that continual learning can help causal models stay up-to-date over time.

Imagine there is a causal model trained on some data and we assume that this is the best possible model that can be obtained using known techniques and the provided data. What if something in the problem changes or we get new data describing only some parts of the true underlying model?

For example, consider a company trying out different products to sell. Every day, they put a different type of product (A) up for sale and at the end of the day, they obtain information about the amount of money earned (B). In this example, the company is still trying to find out which products are selling well and which do not, so it is safe to assume that the amount of money earned here does not causally influence the type of product sold. On the other hand, which products are put up for sale certainly determines the profit at the end of the day. Therefore, we have the small causal graph A $\rightarrow$ B and, given enough data, it can be calculated how much profit is to be expected depending on the type of product sold. Most likely, in a real-world scenario, there would be more variables included which have an influence on the model, for example the number of guests entering the store or the day of the year. We exclude most of these in our example but it is reasonable to assume that some of those would be included and known, while other variables which influence the model are not included, either because the actual values are unknown (e. g. wealth of the customers that day), they might not have been thought of, or it was impossible to include them for other reasons. Now, imagine one such variable suddenly becomes available. For example, the company could sell its wares in several places and one day the data scientists get access to new data including the location of the store (C) which now is another cause of the amount of money earned (new causal graph: A $\rightarrow$ B $\leftarrow$ C). How can the existing model be updated to account for the newly acquired data without having to retrain the entire model?

But even if the general structure of the problem (causal graph) remains unchanged, we could benefit from continual learning. Staying with the previous example, imagine that the popularity of a product type changes. We would assume that much of the relations between the causes A and C and the effect B remains the same but certainly some relations including that specific product type change.[1] Is it possible to retain the knowledge which still holds and at the same time update the model to represent the new situation?

In this paper, we mainly focus on Neural Causal Models (NCMs) (Xia et al. 2021) as one example for causal models. Here, continual learning provides tools which can help keep NCMs running and up-to-date.

First, we look in more detail at the problem to be solved.

## Problem

The problem addressed in this position paper is concerned with updating causal models given new information (data). Here, it can be distinguished between two general problems:

[1] Arguably, this is not a change of relations but a change of unknown variables not included in the model (lack of knowledge). Even so, it can be useful to think of it as a change of relationships.

**P1** Parts of the probability distributions of the model change (the causal graph remains unchanged).

**P2** The number of variables in the causal model changes.

Since **P2** usually includes **P1**, **P1** can be seen as a sub-task of **P2** and the overall easier problem.

But why is this an important problem? Is it not easy to simply add the new data to the original data or replace some parts of the original data and retrain the model? Depending on the application, this might be a valid possibility and, in that case, continual learning is not needed. However, just retraining the model has several possible downsides:

**A) Time and efficiency in general.** Training a new causal model can take a lot of time and resources. Especially if the data changes regularly, it could be unreasonable to train a model from scratch every time.

**B) Original data is unavailable.** Privacy aspects, storage constraints, and other reasons could make it impossible to keep data stored for a longer time. If the new data is not sufficiently large and complete, a retrained new model could end up significantly worse than the original model, while an updated model could benefit from both the information of the original model (indirectly the original data) and the newly acquired data.

**C) New data is incomplete.** The new data might not be complete and only contain some features, like a new variable or only the features for one cause-effect relationship which presumably changed. Here, retraining could simply be impossible.

## Proposed Solution Strategy

For the purpose of this position paper, we consider neural causal models as the type of causal model used. Here, parent-child relationships are modeled by neural networks (NNs). Applying the ideas on models other than NCMs might very well be possible but is not within the scope of this position paper.

First, it is worth mentioning that simply using a model such as an NCM inherently (to a degree) opens the causal model up to continual learning. Since the "mechanisms" (functions determining a variable based on the parent variables) are usually assumed to be independent of each other, they can also be updated separately. In other words, if it is known that only a certain subset of mechanisms changed, those can be updated while requiring data only for the features (variables) relevant for these mechanisms (the respective child and parent variables).

This can be very useful but is not an exciting new revelation so let us get back to the two aforementioned problems and discuss solution strategies. For the second, more complex problem, these strategies are less specific but might serve as first steps towards tackling that problem.

### Problem 1: Change of Probabilities

**Retrain.** Training a new model is a valid strategy in general but this approach also has various problems (refer to the previous parts of this paper).

**Continue Training.** One can continue training as before but using the new data. If what you train on now consists of data representative for the entire problem region, this should work out well. However, if the new data is very specific and does not capture the entire region, *catastrophic forgetting* (Robins 1995; Kirkpatrick et al. 2017) could become a problem, where predictions for data points not represented by the new data are incorrect, although they were correct for the original model.

**Continual Learning.** Continual learning can help a lot, depending on the specific problem formulation. Assume that we have discrete variables[2] and therefore, given a specific model, there is only a finite amount of probabilities this variable can obtain (one for each parent configuration). If an NCM is used and the new data only covers some of these parents configurations, continual learning methods can be used to avoid (or at least reduce) *catastrophic forgetting* of the other parent configurations. Continual learning methods for this purpose include elastic weight consolidation (Kirkpatrick et al. 2017) (keep neurons for useful input-output relationships fixed), knowledge distillation (Gou et al. 2021) (distill desired input-output relationships and train a new model), and (pseudo-)rehearsal (Robins 1995) (train on the new data but add artificial or actual data points representing the input-output relationships you want to keep).

### Problem 2: Change of Variables

Continual learning could also be a helpful tool towards updating the structure of a causal model. If a variable is added as a new parent to another variable, the existing NN could be extended by additional neurons to increase the expressivity of the NN and maybe even keep some useful connections and neurons within the NN which are still helpful (but they should not be fixed in case the relationship changed significantly). Another idea is that if a relationship between some variables is presumably complex but has two or more causes (child variables), maybe those could benefit from sharing a part of the NN architecture at the beginning and have different child variables represented by different output (task) specific layers at the end of the NN (Li and Hoiem 2017).

## Conclusion and Outlook

Continual learning and causality (NCMs in particular) have several goals in common, including but not limited to model adaptation given new information, invariance of unchanged knowledge, and efficient use of data. We postulate that causal models therefore can benefit from continual learning methods which are designed to update new or changed parts of a model while keeping other parts functional.

One can also think of further areas in which a continual perspective on causality could help. For example, one might even try to create some kind of "meta model" which operates on top of a causal model but, given the previous changes in that causal model, is tasked to predict how the causal model is expected to change in the future.

[2]The idea could also work for continuous variables but it requires a more sophisticated approach. For this position paper, we consider discrete variables as the simpler version of this problem.

## References


Carriger, J. F.; and Barron, M. G. 2020. A Bayesian network approach to refining ecological risk assessments: Mercury and the Florida panther (Puma concolor coryi). *Ecological modelling*, 418: 108911.

Chen, Z.; and Liu, B. 2018. Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 12(3): 1–207.

Fenton, N. E.; Neil, M.; Osman, M.; and McLachlan, S. 2020. COVID-19 infection and death rates: the need to incorporate causal explanations for the data and avoid bias in testing. *Journal of Risk Research*, 23(7-8): 862–865.

Glymour, C.; Zhang, K.; and Spirtes, P. 2019. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10: 524.

Gou, J.; Yu, B.; Maybank, S. J.; and Tao, D. 2021. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6): 1789–1819.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13): 3521–3526.

Koch, D.; Eisinger, R. S.; and Gebharter, A. 2017. A causal Bayesian network model of disease progression mechanisms in chronic myeloid leukemia. *Journal of theoretical biology*, 433: 94–105.

Li, Z.; and Hoiem, D. 2017. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12): 2935–2947.

Mundt, M.; Lang, S.; Delfosse, Q.; and Kersting, K. 2022. CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability. In *International Conference on Learning Representations*. https://arxiv.org/abs/2110.03331.

Pearl, J. 2009. *Causality*. Cambridge university press.

Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Robins, A. 1995. Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2): 123–146.

Xia, K.; Lee, K.-Z.; Bengio, Y.; and Bareinboim, E. 2021. The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 34: 10823–10836.