CONTINUAL LEARNING AND CAUSAL INFERENCE: A REVIEW

Neil Ashtekar
Department of Computer Science
The Pennsylvania State University
University Park, PA 16802, USA
nca5096@psu.edu

ABSTRACT

Continual learning and casual inference have received increased attention in the machine learning community over the past few years. Both fields enable new ways of learning and performing inference: continual learning methods enable models to be efficiently updated over time as new data arrives, and causal inference methods enable models to answer counterfactual questions. These fields are complimentary – ideas from one can be used to improve and extend methods in the other. In this review article, we summarize existing work at the intersection of 'continual causality' and outline potential new directions.

1 INTRODUCTION

Causal inference is a subset of statistics concerned with estimating cause-and-effect relationships from data. Continual learning is a subset of machine learning concerned with efficiently updating previously learned models given new data without forgetting prior knowledge. Both causal reasoning and continual learning are essential components of truly intelligent systems. These two fields are complimentary – ideas from causal inference can improve existing continual learning techniques, and ideas from continual learning can extend causal inference techniques to new applications.

Many desirable properties of machine learning systems such as generalization, interpretability, and robustness have deep connections to causality. For example, understanding the causal structure of a supervised learning problem enables strong generalization to unseen samples and avoids errors due to spurious correlations. Ideas from causal inference can be used to improve knowledge transfer and generalization ability of continual learning systems. Alternatively, ideas from continual learning can be used to build 'real-world' causal inference models which adapt as new data becomes available over time.

In this review article, we provide an overview of existing work at the intersection of causal inference and continual learning, outline open problems, and briefly discuss connections to neuroscience. We include prior work as well as work from the recent AAAI '23 "ContinualCausality" bridge program.

2 CAUSAL INFERENCE PROBLEMS IN THE CONTINUAL LEARNING SETTING

2.1 CAUSAL EFFECT ESTIMATION WITH INCREMENTAL DATA

Performing causal effect estimation as data arrives sequentially is perhaps the most straightforward combination of causal inference and continual learning. Chu et al. (2023) proposes an extension of representation balancing to estimate causal effects in the continual learning setting. The idea of representation balancing for causal effect estimation was first proposed in Shalit et al. (2017). This approach uses deep learning to construct a "treatment-agnostic representation" which enables unbiased estimation. The approach in Chu et al. (2023) combines representation balancing with replay – a continual learning technique in which important samples from previous tasks are stored in learned in addition to new data – to avoid catastrophic forgetting (see Figure 1). Note that the extension of representation balancing to the continual learning setting is nontrivial, as feature transformations are necessary to ensure that representations are comparable through time.

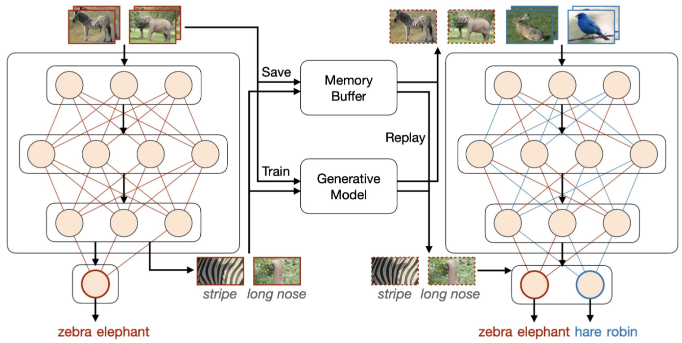


Figure 1: An illustration from Wang et al. (2023) of replay for continual learning. A subset of previous task data (images of zebras and elephants) is either stored or generated and learned jointly with current task data (images of hares and robins) in order to avoid forgetting.

2.2 Causal Abstractions

In the continual learning setting, performing various tasks may require reasoning at various levels of abstraction. Zečević et al. (2023) explores the idea of continual causal abstractions, motivated by two examples: a chemist performing experiments and reasoning using causal abstractions, and a dietitian giving health advice to clients over time aided by continual abstractions. Formally, Zečević et al. (2023) describes τ-abstractions between pairs of Structural Causal Models (SCMs) and outlines a plan to learn τ-abstractions as new data arrives over time.

2.3 Active Learning

The Never-Ending Language Learner (NELL) (Mitchell et al., 2018) is an online learning system that continuously reads the internet and learns relationships between entities. It is one of the most well-known continual learning systems, and is arguably the best example of a truly online learner (it ran continuously from 2010 until 2018). Motivated by NELL, Natarajan & Kersting (2023) proposes NERL – Never-Ending Reasoning and Learning – focused on causal knowledge. NERL describes a high-level approach for continual causal reasoning through learning from multimodal data and interacting with human experts. This is one example of active learning in which an automated system queries humans for labels to improve its performance. Active learning for experiment selection has shown promise for causal discovery (He & Geng, 2008; Hauser & Bühlmann, 2014). The sequential progression of active learning makes it an natural fit for the continual learning setting.

2.4 New Problem Statements

The authors of Chu et al. (2023) also submitted a short paper (Chu & Li, 2023) to the Continual Causality program summarizing their past work and introducing new problem statements. For example: previously unobserved confounders may be observed as new data is collected over time, the typical Rubin framework assumptions may need to be revisited in the continual learning setting, and there may be interesting new problems combining causal inference with “big data” and federated learning.

Busch et al. (2023) describes two problem statements and potential solutions for continual causal modelling within the Pearl Structural Causal Model (SCM) framework. In the first problem statement (1), the structure of the underlying causal graph remains constant over time, though the probability distributions of the variables change. In the second problem statement (2), the structure of the causal graph changes over time. To address (1), continual learning approaches such as replay, regularization, and knowledge distillation could be applied. To address (2), transfer learning and/or architecture-based continual learning approaches are potential solutions. These problem statements bring up interesting philosophical questions about whether true causal relationships ever actually change – while causal relationships may remain constant at the atomic level, most causal models are abstractions and therefore may change over time.

3 Improving Continual Learning with Causality

3.1 A Causal Interpretation of Catastrophic Forgetting

Instead of addressing causal inference problems in the continual learning setting, several papers have looked at standard formulations of continual learning through the lens of causality. In the context of computer vision, Hu et al. (2021) attempts to explain catastrophic forgetting in continual learning using the Pearl framework of SCMs and do-calculus. First, the authors construct a causal graph indicating the relationships between new and old data, learned representations, and predicted labels (see Figure 2). Next, they discuss how existing continual learning strategies (data replay, feature and label distillation) mitigate forgetting by introducing new edges in the graph. Finally, they use do-calculus to derive a new continual learning strategy. This strategy works by conditioning on colliders (previously learned representations, $\boldsymbol{X}_o$ in Figure 2) to introduce 'good' confounding and maintain the causal effect of old data on the new model. This approach is based on knowledge distillation and attempts to replicate the net effect of data replay without a memory buffer.

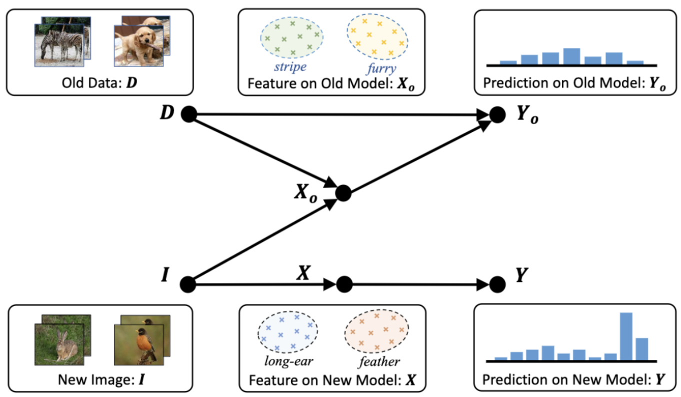


Figure 2: Illustration of the SCM described in Hu et al. (2021) for the continual learning setting.

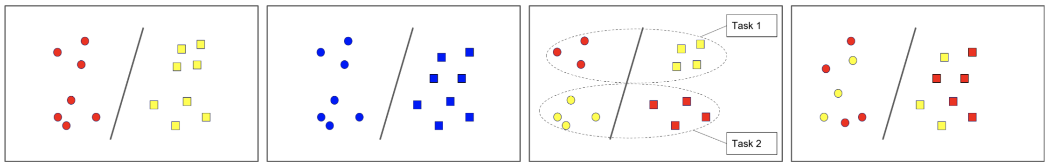
3.2 Causal Replay

A couple of papers accepted to the ContinualCausality program investigate the use of replay (sometimes called rehearsal) to mitigate forgetting of previously learned causal effects. See Figure 1 for an illustration of replay. Finding samples to replay (i.e. exemplars) is the central challenge in this family of approaches. Some approaches do not store exemplars from the training set, and instead learn a generator to produce previous task data – these approaches are known as "psuedo-rehearsal". Churamani et al. (2023) discusses interesting ways to choose or learn exemplars for replay by leveraging causal knowledge. Specifically, score-based causal discovery techniques can be used to prioritize samples for rehearsal. Alternatively, learned generators can be queried to produce *counterfactual* data used in psuedo-rehearsal. Rehearsal is also described in Busch et al. (2023) as a solution to deal with probability distributions which change over time in a SCM.

3.3 Addressing Spurious Features and Dataset Bias

Several works have considered ways to address spurious features in continual learning. Spurious features are predictive of labels during training but do not generalize beyond the training set (Lesort, 2022). The cow-camel classification problem is a famous example of spurious features – a computer vision classifier trained on images of the two animals learns to associate cows with grass and camels with sand, and performs poorly when classifying animals in unusual locations (Beery et al., 2018). "Local spurious features" present a challenge specific to continual learning (Lesort, 2022; 2023) – these are features *within* a task that correlate with the task's labels but do no generalize to other tasks (see Figure 3). Lee et al. (2023) discusses essentially the same problem – continual learning from biased data – empirically showing that existing methods transfer bias across tasks and that the naive application of debiasing techniques induces forgetting.

Previous work (Javed et al., 2020) attempts to detect and remove spurious features in the online setting in order to indirectly learn a causal model. (Javed et al., 2020) has noted that the temporal structure of the data is essential when detecting spurious features online – this is an interesting observation and potential upside of the order-dependence of many continual learning algorithms.



(a) Train (Spurious) (b) Test (Spurious) (c) Train (local spurious) (d) Test (local spurious)

Figure 3: Illustration of spurious features from Lesort (2022). True class labels are represented by shape (squares and circles), while spurious features are represented by color (red and yellow).

3.4 Meta-Learning

Meta-learning is one family of approaches for continual learning (Javed & White, 2019; Riemer et al., 2019). Seng et al. (2023) from the ContinualCausality program takes an interesting approach to meta-continual learning using causal inference. Namely, they treat continual learning as a treatment effect estimation problem. The treatment corresponds to learning a specific task, and the effect corresponds to the difference in model performance before and after the task is learned. This approach can be used to predict whether learning task B will improve a model's performance on previously learned task A prior to actually learning task B. Based on these predictions, specific tasks can be prioritized or de-prioritized given their expected effect on model performance and overall importance.

3.5 Learning with Causal Assumptions

Auxiliary information can often be used to improve the performance of machine learning systems. For example, in the supervised learning setting, methods have been proposed to leverage pairwise relationships between data to reduce sample complexity (Xu et al., 2019), make use of additional unlabeled data to improve classification performance (Hady & Schwenker, 2013), and take advantage of shape priors in computer vision to make predictions from otherwise insufficient information (Mittal et al., 2023). Causal assumptions are another form of auxiliary information that could be used to improve learning. This idea is relatively unexplored, with existing work focused on the use of partial causal knowledge to aid learning (Mahajan et al., 2019; Berrevoets et al., 2023). From the ContinualCausality program, Willig et al. (2023) proposes the use of SCMs to guide open-world concept discovery. They give the example of visual concept discovery using images – machine learning is used to identify concepts (i.e. a match and matchbox in an image) and a SCM is used to infer new concepts (i.e. striking a match against a matchbox creates a flame). This work is motivated by the observation that humans rarely 'learn from scratch' and typically leverage existing causal knowledge when learning new things.

4 Open Problems

Almost all problems at the intersection of continual learning and causality remain open. While there have been many papers investigating methods for 'vanilla' continual learning (Wang et al., 2023), there are only a few papers looking at causal inference problems in the continual learning setting. Javed et al. (2020) and Chu et al. (2023) appear to be the only papers addressing this setting with fully-developed solutions. Further, preliminary work such as Chu et al. (2023) relies on a straightforward (and perhaps naive) application of continual learning techniques to the causal inference setting. There may be more sophisticated ways to combine ideas from both fields to improve performance, such as causal rehearsal (Churamani et al., 2023). In addition, recent work has only just begun to *define* continual causality problem statements, let alone propose solutions (Zečević et al., 2023; Chu & Li, 2023; Busch et al., 2023).

Connections between other ideas in causality and continual learning could be further strengthened in order to create better solutions. For example, it would be interesting to revisit causal transportability (Pearl & Bareinboim, 2011; Rojas-Carulla et al., 2018) in the continual learning setting to better understand how to transfer knowledge between tasks. Finally, the cross-pollination of ideas between causal inference and machine learning (specifically deep learning) is relatively new, and there is likely still room for innovation.

5 Connections to Neuroscience

As an aside, a couple of cognitive neuroscience papers have relevant connections to continual causality. First, Gong et al. (2023) performs experiments to understand how humans learn causal structures over time through both observation and intervention. They find that subjects intervene in ways which maximize information gain while minimizing the computational cost of inference. This is consistent with past work in economics on "bounded rationality" of human decision making (Simon, 1990) and may have implications for resource-constrained automated causal discovery. Second, Shams & Beierholm (2022) proposes Bayesian causal inference as a potential unifying neuroscience theory. Bayesian inference is a natural fit for continual learning as priors can be updated over time as new evidence is observed.

6 Conclusion

Many interesting problems exist in the realm of continual causality. Ideas from continual learning can extend causal inference methods to more flexible, realistic settings with incrementally available data and changing causal relationships. Ideas from causal inference can improve continual learning methods through a better understanding forgetting and incorporation of causal knowledge. Problems at the intersection of these two fields remain largely open, with high potential for innovative new ways of learning.

References

Sara Beery, Grant van Horn, and Pietro Perona. Recognition in terra incognita, 2018.

Jeroen Berrevoets, Krzysztof Kacprzyk, Zhaozhi Qian, and Mihaela van der Schaar. Causal deep learning, 2023.

Florian Peter Busch, Jonas Seng, Moritz Willig, and Matej Zečević. Continually updating neural causal models. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=dwIudqXzBCL`.

Zhixuan Chu and Sheng Li. Continual treatment effect estimation: Challenges and opportunities. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=kn6rBT9gotr`.

Zhixuan Chu, Ruopeng Li, Stephen Rathbun, and Sheng Li. Continual causal inference with incremental observational data, 2023.

Nikhil Churamani, Jiaee Cheong, Sinan Kalkan, and Hatice Gunes. Towards causal replay for knowledge rehearsal in continual learning. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=t6HPdtyIaXB`.

Tianwei Gong, Tobias Gerstenberg, Ralf Mayrhofer, and Neil R. Bramley. Active causal structure learning in continuous time. *Cognitive Psychology*, 140:101542, 2023. ISSN 0010-0285. doi: https://doi.org/10.1016/j.cogpsych.2022.101542. URL `https://www.sciencedirect.com/science/article/pii/S0010028522000780`.

Mohamed Farouk Abdel Hady and Friedhelm Schwenker. Semi-supervised learning. *Handbook on Neural Information Processing*, pp. 215–239, 2013.

Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4):926–939, 2014.

Yang-Bo He and Zhi Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547, 2008.

Xinting Hu, Kaihua Tang, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. Distilling causal effect of data in class-incremental learning, 2021.

Khurram Javed and Martha White. Meta-learning representations for continual learning, 2019.

Khurram Javed, Martha White, and Yoshua Bengio. Learning causal models online, 2020.

Donggyu Lee, Sangwon Jung, and Taesup Moon. Issues for continual learning in the presence of dataset bias. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=iWLbLoleZMN`.

Timothée Lesort. Spurious features in continual learning. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=7_QHeKujaj`.

Timothée Lesort. Continual feature selection: Spurious features in continual learning, 2022.

Divyat Mahajan, Chenhao Tan, and Amit Sharma. Preserving causal constraints in counterfactual explanations for machine learning classifiers. *arXiv preprint arXiv:1912.03277*, 2019.

T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling. Never-ending learning. *Commun. ACM*, 61(5):103–115, apr 2018. ISSN 0001-0782. doi: 10.1145/3191513. URL `https://doi.org/10.1145/3191513`.

Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. Autosdf: Shape priors for 3d completion, reconstruction and generation, 2023.

Sriraam Natarajan and Kristian Kersting. Never ending reasoning and learning: Opportunities and challenges. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=vhQsMW7L60`.

Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25, pp. 247–254, 2011.

Matthew Riemer, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro. Learning to learn without forgetting by maximizing transfer and minimizing interference, 2019.

Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. Invariant models for causal transfer learning. *The Journal of Machine Learning Research*, 19(1):1309–1342, 2018.

Jonas Seng, Florian Peter Busch, Matej Zečević, and Moritz Willig. Treatment effect estimation to guide model optimization in continual learning. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=hsdU13XxklX`.

Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms, 2017.

Ladan Shams and Ulrik Beierholm. Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, 137:104619, 2022. ISSN 0149-7634. doi: https://doi.org/10.1016/j.neubiorev.2022.104619. URL `https://www.sciencedirect.com/science/article/pii/S0149763422001087`.

Herbert A Simon. Bounded rationality. *Utility and probability*, pp. 15–18, 1990.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: Theory, method and application, 2023.

Moritz Willig, Matej Zečević, Jonas Seng, and Florian Peter Busch. Causal concept identification in open world environments. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=1M6fV3HQ3RE`.

Yichong Xu, Sivaraman Balakrishnan, Aarti Singh, and Artur Dubrawski. Regression with comparisons: Escaping the curse of dimensionality with ordinal information, 2019.

Matej Zečević, Moritz Willig, Florian Peter Busch, and Jonas Seng. Continual causal abstractions. In *Continual Causality Bridge Program at AAAI23*, 2023. URL `https://openreview.net/forum?id=UC-7gWM77ZY`.