Towards Causal Replay for Knowledge Rehearsal in Continual Learning

Anonymous submission

Abstract

Given the challenges associated with the real-world deployment of Machine Learning (ML) models, especially towards efficiently integrating novel information *on-the-go*, both Continual Learning (CL) and Causality have been proposed and investigated individually as potent solutions. Despite their complimentary nature, the *bridge* between them is still largely unexplored. In this work, we focus on causality to improve the learning and knowledge preservation capabilities of CL models. In particular, positing *Causal Replay* for knowledge rehearsal, we discuss how CL-based models can benefit from *causal interventions* towards improving their ability to replay past knowledge in order to mitigate forgetting.

1 Introduction

Real-world application of Machine Learning (ML) solutions require models to dynamically learn and adapt with streams of *incrementally* acquired data, while preserving past knowledge. Conventional ML-based methods are ill-fated to meet these challenges as they work under a pivotal assumption that all data is available a priori under relatively stationary data distributions (Graffieti, Borghi, and Maltoni 2022). This *stationarity* ensures that training samples are *identically and independently drawn* (i.i.d), allowing models to learn in batches of representative distributions. The real-world, however, is not stationary and changes continuously (Hadsell et al. 2020). As models continually encounter novel information, violating this i.i.d assumption, their ability to remember previously learnt tasks progressively deteriorates, resulting in forgetting (McCloskey and Cohen 1989).

Continual Learning (CL) (Parisi et al. 2019; Hadsell et al. 2020) aims to address adaptability in ML-based models by enabling them to *continually* learn and adapt, balancing *incremental learning* of novel information with the *preservation* of past knowledge. CL focuses on learning with continuous streams of data acquired from non-stationary or changing distributions (Hadsell et al. 2020). This may be achieved by *regulating* model updates to control plasticity or *rehearsing* past knowledge by storing and *replaying* already seen information to simulate *i.i.d* learning settings.

Given the above, Causality (Pearl 2009), especially addressing adaptability and causal discovery, can complement lifelong learning of information by helping understand the causal structure of the data or task and 're-adjust' model learning to cope with changing data distributions (Pearl 2019). Furthermore, it has been posited that the increasingly apparent challenges in ML (such as robustness, generalisation, bias, transparency) are due to conventional ML methods learning *correlation*-based patterns and relationships (Schölkopf et al. 2021). *Causal reasoning* tools can contribute towards addressing some of these challenges (Cheng et al. 2022).

In this position paper, we focus on *knowledge rehearsal* as an effective tool for CL-based models to preserve past knowledge particularly using *causal interventions* to understand and update data distributions such that only the most relevant data samples (for *rehearsal*) or features (for *pseudo-rehearsal*) can be used by the model to preserve past knowledge. Such *Causal Replay* can help improve the efficiency of knowledge rehearsal for continual learning of information.

1.1 Knowledge Rehearsal to Mitigate Forgetting

Efficient rehearsal of past knowledge can be achieved by physically storing samples from previous tasks in memory buffers and regularly sampling from them (*rehearsal* – Robins 1993) mixing it with new data. The simplest strategy to achieve this is to fix the size of the memory buffer to be '*large enough*' and *randomly* maintain a fraction of previously seen samples from each task in the buffer for periodic rehearsal (Hsu et al. 2018). However, as the number of tasks increases, fewer samples are available for rehearsal per task. Other *sophisticated* rehearsal methods focus on prioritising replay following certain heuristics such as feature or classification margins (Hu, Zhang, and Zhu 2021), or storing exemplars for each task that best approximate task means (Rebuffi et al. 2017). Despite such 'intelligent' sampling, high dimensionality of data and a large number of tasks require a huge amount of memory, making their real-world application inefficient (Kwon et al. 2021).

Alternatively, generative models may be used, along with the learning model, that learn the inherent data statistics, enabling models to draw *pseudo-samples* to be replayed (*pseudo-rehearsal* – Robins 1995) along with novel data. Recent advances in generative models (Goodfellow et al. 2014; Kingma and Welling 2013), particularly in their ability to generate high-quality samples, have greatly enhanced the potential of pseudo-rehearsal methods (Shin et al. 2017; Churamani and Gunes 2020). More recent methods focus on

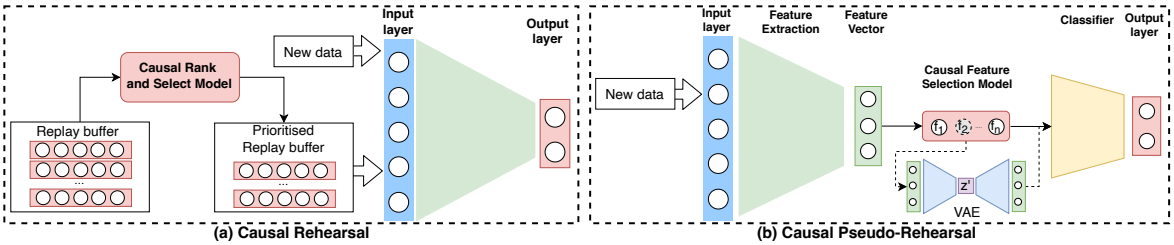

Figure 1: Causal Replay for (a) Prioritised Rehearsal and efficient (b) Pseudo-rehearsal of past knowledge.

generative feature replay (van de Ven, Siegelmann, and Tolias 2020; Stoychev, Churamani, and Gunes 2023), alleviating the need to optimise generators for reconstructing high-dimensional samples. However, as the number of tasks increases, they face *capacity saturation* and are not able to efficiently learn task-discriminative representations. Furthermore, the *generators* become harder to train, resulting in an inefficient rehearsal of past knowledge. We believe causality can offer significant improvements in this regard. To date, there is minimal work that explores the synergies between CL and causality (Chu, Rathbun, and Li 2021).

1.2 Causality

The study of causality entails a range of tools such as graphical models, the *do*-operator, counterfactuals as well as structural equations (Pearl 2009). Using these tools, conventional causal research has mostly focused on causal pattern recognition (Vowels, Camgoz, and Bowden 2021) and causal distribution estimation (Yao et al. 2021), Here, we focus on methods to merge conventional causal research with ML to address the existing gaps. Recent works in causal interpretability (Moraffah et al. 2020) and causal fairness (Makhlouf, Zhioua, and Palamidessi 2020) have proven such an approach to be promising. Here, we leverage two main themes: *Causal Interventions* and *Causal Structure Discovery*.

Following Pearl's notation (Pearl 2009) for a Structural Causal Model (SCM), we have a set of variables V and a set of functions F that encode the causal relations between each variable. Using this framework, *causal interventions* can be either be 'structural' or 'parametric' (Spirtes et al. 2000) representing a continuum of 'harder' to 'softer' interventions. A 'hard' intervention can be understood as a forcible removal of an edge such that the function encoding $V_i \leftarrow f_{V_i}$ is modified such that another variable V_j is no longer a parent of variable V_i. 'Soft' interventions, on the other hand, simply modify the conditional probability distributions of the *intervened* variable V_i. Depending on the task, we can combine the most appropriate form of causal interventions with CL-based models to preserve past knowledge and update the model using only the relevant features. In addition, we also propose to leverage existing causal discovery methods (Vowels, Camgoz, and Bowden 2021) that can be utilised to discover causal relations within the observational data. We propose to impart the discovered causal knowledge to CL-based methods in order to mitigate forgetting and to learn new relevant features.

2 Causal Replay for Knowledge Rehearsal

2.1 Causal Rehearsal

One strategy for augmenting CL with causality can be causality-driven rehearsal (see Figure 1 a). Firstly, we aim to understand the causal structure of task-specific data in order to *prioritise* samples for rehearsal. As neural networks are capable of representing the input features as well as their respective causal relations to each other within their parameters, we can learn the causal structure of the data during the training phase using a range of existing causal discovery methods (Vowels, Camgoz, and Bowden 2021). An online example of doing so is exemplified by Javed, White, and Bengio (2020). As we are only able to discover causal Directed Acyclic Graphs (DAGs) up to Markov equivalence, we can subsequently leverage causal-scoring methods (Glymour, Zhang, and Spirtes 2019) or causality-based feature selection methods (Yu et al. 2020) to determine which samples to prioritise for rehearsal. Subsequently, as we update the model with each new task, we *reprioritise* the samples to update the memory buffer as well as the learnt causal structure. As such, the causal model can then also be updated in a continual manner as more data becomes available.

2.2 Causal Pseudo-rehearsal

Another opportunity is that of causality-driven pseudo-rehearsal (see Figure 1 b). Here the goal is to use the learnt causal structure of the data to rehearse information in a principled manner. Attempts to remove unwanted causal relations has proven to be effective in the case of knowledge distillation (Deng and Zhang 2021). However, such an idea has yet to be fully explored in CL. Existing methods largely rely on pattern generation to simulate i.i.d. settings. However, this does not take into account the causal relations between variables. One way of addressing this is to make use interventions (both 'hard' and 'soft') such that we generate samples from the updated distribution which has been 'intervened' upon. Such an approach has proven to be effective in the domain of disentangled representation learning using Variational Autoencoders (VAEs) (Yang et al. 2021). Instead of simply generating pseudo-samples, we can intervene by updating the parameters of the generative model based on the causal effect estimated or parameterised by the learnt causal structure of the data. These parameters can also be continually updated given new information. By conducting pseudo-rehearsal in this manner, we are able to adapt to the changes in new data whilst preserving old information.

References

Cheng, L.; Guo, R.; Moraffah, R.; Sheth, P.; Candan, K. S.; and Liu, H. 2022. Evaluation methods and measures for causal learning algorithms. *IEEE Transactions on Artificial Intelligence*.

Chu, Z.; Rathbun, S.; and Li, S. 2021. Continual Lifelong Causal Effect Inference with Real World Evidence.

Churamani, N.; and Gunes, H. 2020. CLIFER: Continual Learning with Imagination for Facial Expression Recognition. In *15th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 322–328. IEEE.

Deng, X.; and Zhang, Z. 2021. Comprehensive knowledge distillation with causal intervention. *Advances in Neural Information Processing Systems*, 34: 22158–22170.

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

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.; and Weinberger, K. Q., eds., *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, volume 27 of *NIPS'14*, 2672–2680. MIT Press, Cambridge, MA, USA: MIT Press.

Graffieti, G.; Borghi, G.; and Maltoni, D. 2022. Continual Learning in Real-Life Applications. *IEEE Robotics and Automation Letters*, 7(3): 6195–6202.

Hadsell, R.; Rao, D.; Rusu, A. A.; and Pascanu, R. 2020. Embracing Change: Continual Learning in Deep Neural Networks. *Trends in Cognitive Sciences*, 24(12): 1028–1040.

Hsu, Y.-C.; Liu, Y.-C.; Ramasamy, A.; and Kira, Z. 2018. Re-evaluating Continual Learning Scenarios: A Categorization and Case for Strong Baselines. In *NeurIPS Continual learning Workshop*.

Hu, G.; Zhang, W.; and Zhu, W. 2021. Prioritized Experience Replay for Continual Learning. In *2021 6th International Conference on Computational Intelligence and Applications (ICCIA)*, 16–20.

Javed, K.; White, M.; and Bengio, Y. 2020. Learning causal models online. *arXiv preprint arXiv:2006.07461*.

Kingma, D. P.; and Welling, M. 2013. Auto-Encoding Variational Bayes. *CoRR*, abs/1312.6114.

Kwon, Y. D.; Chauhan, J.; Kumar, A.; Hui, P.; and Mascolo, C. 2021. Exploring System Performance of Continual Learning for Mobile and Embedded Sensing Applications. *CoRR*, abs/2110.13290.

Makhlouf, K.; Zhioua, S.; and Palamidessi, C. 2020. Survey on causal-based machine learning fairness notions. *arXiv preprint arXiv:2010.09553*.

McCloskey, M.; and Cohen, N. J. 1989. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In Bower, G. H., ed., *Psychology of Learning and Motivation*, volume 24 of *Psychology of Learning and Motivation*, 109–165. Academic Press.

Moraffah, R.; Karami, M.; Guo, R.; Raglin, A.; and Liu, H. 2020. Causal interpretability for machine learning-problems, methods and evaluation. *ACM SIGKDD Explorations Newsletter*, 22(1): 18–33.

Parisi, G. I.; Kemker, R.; Part, J. L.; Kanan, C.; and Wermter, S. 2019. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113: 54–71.

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

Pearl, J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3): 54–60.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. iCaRL: Incremental Classifier and Representation Learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2001–2010.

Robins, A. 1993. Catastrophic forgetting in neural networks: the role of rehearsal mechanisms. In *Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, 65–68.

Robins, A. 1995. Catastrophic Forgetting, Rehearsal and Pseudorehearsal. *Connection Science*, 7(2): 123–146.

Schölkopf, B.; Locatello, F.; Bauer, S.; Ke, N. R.; Kalchbrenner, N.; Goyal, A.; and Bengio, Y. 2021. Toward causal representation learning. *Proceedings of the IEEE*, 109(5): 612–634.

Shin, H.; Lee, J. K.; Kim, J.; and Kim, J. 2017. Continual Learning with Deep Generative Replay. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*, 2990–2999. Curran Associates, Inc.

Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. *Causation, prediction, and search*. MIT press.

Stoychev, S.; Churamani, N.; and Gunes, H. 2023. Latent Generative Replay for Resource-efficient Continual Learning of Facial Expressions. In *17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE.

van de Ven, G. M.; Siegelmann, H. T.; and Tolias, A. S. 2020. Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, 11(1).

Vowels, M. J.; Camgoz, N. C.; and Bowden, R. 2021. D'ya like DAGs? A survey on structure learning and causal discovery. *ACM Computing Surveys (CSUR)*.

Yang, M.; Liu, F.; Chen, Z.; Shen, X.; Hao, J.; and Wang, J. 2021. CausalVAE: Disentangled representation learning via neural structural causal models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9593–9602.

Yao, L.; Chu, Z.; Li, S.; Li, Y.; Gao, J.; and Zhang, A. 2021. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5): 1–46.

Yu, K.; Guo, X.; Liu, L.; Li, J.; Wang, H.; Ling, Z.; and Wu, X. 2020. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys (CSUR)*, 53(5): 1–36.