Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Doctoral Consortium Track

Information Injection to Deep Learning Solutions in Knowledge Transfer

Paulina Tomaszewska
Warsaw University of Technology,
Faculty of Mathematics and Information Science, Warsaw, Poland
paulina.tomaszewska3.dokt@pw.edu.pl

Abstract

Nowadays, with the thrive of AutoML techniques, it is expected that Machine Learning algorithms will work well without any human intervention. This is why the main focus is on introducing new neural network architectures, especially those capable of learning to learn. Recently, the Data-Centric AI Challenge was proposed by Andrew Ng whose goal was to change the paradigm and instead of having a fixed dataset and modifying the model, now the model is fixed and the data is preprocessed so that the model results in the best performance. In my thesis, I would like to focus on another approach, where I would not modify the given data nor introduce new architectures, instead, I would like to propose new ways of injecting additional information into *knowledge transfer* models to increase their performance.

1 Introduction

In my dissertation, I would like to combine two aspects: *knowledge transfer* models and the injection of additional knowledge into the system in order to boost the performance.

1.1 Knowledge Transfer

Knowledge transfer is understood in a broader sense than *transfer learning* (TL). It includes algorithms as diverse as *domain adaptation*, *knowledge distillation*, *multi-task learning*, *continual learning* etc. They vary in the type, source, and target of the transferred knowledge (for more details, see the survey [Lu *et al.*, 2020]). Another algorithm that could be interpreted as *knowledge transfer* is *multiple-instance learning* [Ilse *et al.*, 2018], where the information from multiple instances is aggregated to represent the entire bag of instances in a suitable form to perform the final task. This is the bottom-up approach to *knowledge transfer*. Moreover, in the case of Computer Vision applications, the information is often extracted from the instances using pretrained neural networks.

1.2 Information Injection

The injection of additional information into the Deep Learning system could be defined in various ways. In my definition, knowledge injection means providing additional information to an intelligent system. It could be expert knowledge or other information that could be derived by a human based on the data itself. However, the goal is not to replicate the information that can be discovered by the neural network on its own (such as edges in convolutional neural networks). Expert knowledge can be incorporated into DL models in the form of neuro-symbolic AI (i.e., integration of rule-based AI and neural networks). This research direction is of utmost importance, especially in healthcare [Yousefirizi *et al.*, 2022].

To make the definition of information injection less fuzzy, positive and negative examples are provided. Positive examples include adding positional encoding in Transformer architecture, injecting information about the class in Conditional Generative Adversarial Networks (cGAN) [Mirza and Osindero, 2014] and incorporating information about data distributions in the Instance Weighting Strategy in *continual learning* [Huang *et al.*, 2006]. The information injection can be applied at different stages of the network, from an additional information added to the input (positional encoding), additional input to intermediate layers (cGAN) to the formulation of cost function (weighting). What all these solutions have in common is that the injected information is not trainable, unlike the models themselves. However, the injection is not a *multimodal learning* in which additional knowledge is inserted into the system as a new modality. In this case, in order to solve the task, all the modalities are necessary, unlike in proposed information injection concept, where it is optional but may improve the performance of the system. Moreover, *active learning* [Settles, 2009], where a human labels key data samples, is not an example of information injection. This is caused by the fact that this additional information is provided outside the model.

2 Proposed Contribution

One of the planned projects on the topic of information injection into *knowledge transfer* models refers to the area of *multiple-instance learning* algorithm for histopathological data analysis. Histopathological images are of large, varying resolution e.g. 60000x100000. This makes them difficult to be processed using classical convolutional neural networks. One of the interesting works in the field of histopathological data analysis that deals with the problem of large input size is described in [Lu *et al.*, 2021a] (later referred to as

CLAM for brevity). The general flow of this solution is as follows. First, features are extracted from the patches of images, and then they are passed to the attention module that returns scores. The scores are used as weights in the weighted mean that aggregates information from all patches. The result of the aggregation is supposed to be a proper representation of the overall histopathological image to perform a classification task (distinguish healthy vs. cancerous or different types of cancer). In addition, an auxiliary task (clustering) is applied where pseudolabels are created based on attention scores.

The authors stated in their conclusions that their solution lacks information about the context. This additional information is believed to improve the performance of the system by providing valuable insights that are difficult for the network to recognize.

In CLAM, each patch is processed separately and then aggregated using only the attention scores with respect to the analyzed classes. Therefore, there is no spatial information about the surrounding patches, unlike in the case of convolutional neural networks, where the spatial relations are preserved by default. Consequently, in my work, the main focus is on the injection of context into the CLAM solution.

The idea is to provide additional information in the form of a similarity measure between the extracted features from the patches. Given a patch, the similarity metric (such as cosine similarity or correlation) is calculated with each of the patches in Queen's case. Later, the mean of the values is returned. The advantage of the mean is that, in the case of border patches, the score is not biased. The context could be later incorporated into the aggregation function. The idea behind such a context formulation is the result of a preliminary study. Below, due to limited space, I present only results where a similarity metric in a form of Pearson correlation was used to process Figure 1 into context information at Figure 2.

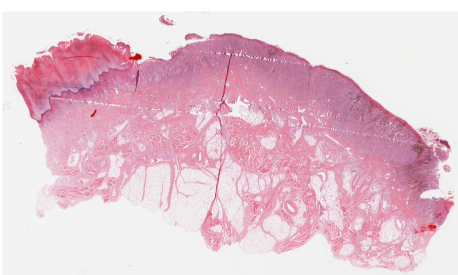

Figure 1: Original histopathological image of a skin lesion.

It has been observed that by computing correlation between the features of neighboring patches, the original tissue structure (shown in Figure 1) can be highly restored, which is promising from the perspective of information injection into the CLAM model. Another idea for providing context could be to incorporate a variant of a positional encoding.

Note that there is an extension of the CLAM solution in [Lu *et al.*, 2021b], where information injection is limited to the concatenation of an intermediate data representation within a neural network with information about the patient's sex. Such an approach is common where an image and a small amount of tabular data are required to make a prediction, however, it does not provide context about the spatial relationships of the patches.

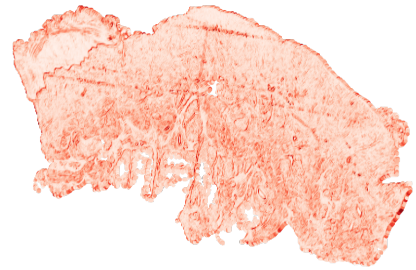

Figure 2: Neighbourhood correlation of every patch visualized by color intensity.

Acknowledgments

I would like to thank my supervisors: Professor Przemysław Biecek and Doctor Elżbieta Sienkiewicz from Warsaw University of Technology for their support.

References

[Huang *et al.*, 2006] Jiayuan Huang, Arthur Gretton, Karsten Borgwardt, Bernhard Schölkopf, and Alex Smola. Correcting Sample Selection Bias by Unlabeled Data. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006.

[Ilse *et al.*, 2018] Maximilian Ilse, Jakub M Tomczak, and Max Welling. Attention-based Deep Multiple Instance Learning. *arXiv preprint arXiv:1802.04712*, 2018.

[Lu *et al.*, 2020] Ying Lu, Lingkun Luo, Di Huang, Yunhong Wang, and Liming Chen. Knowledge Transfer in Vision Recognition: A Survey. *ACM Comput. Surv.*, 53(2), apr 2020.

[Lu *et al.*, 2021a] Ming Lu, Drew Williamson, Tiffany Chen, Richard Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5:1–16, 06 2021.

[Lu *et al.*, 2021b] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, and Faisal Mahmood. AI-based pathology predicts origins for cancers of unknown primary. *Nature*, page 1–25, 2021.

[Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. *arXiv preprint arXiv:1411.1784*, 2014.

[Settles, 2009] Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[Yousefirizi *et al.*, 2022] Fereshteh Yousefirizi, Pierre Decazes, Amine Amyar, Su Ruan, Babak Saboury, and Arman Rahmim. AI-Based Detection, Classification and Prediction/Prognosis in Medical Imaging: Towards Radiophenomics. *PET Clinics*, 17(1):183–212, 2022. Artificial Intelligence and PET Imaging, Part II.