SCAI 2025 doctoral projects

Robust transfer learning, application to deep probabilistic models

S. Destercke, B. Quost, P.-H. Wuillemin

1 Gradual domain adaptation

1.1 Context

In classical machine learning, the purpose is to estimate a model, for instance via a parameter $\theta \in \Theta$, using a training set $\mathcal{L} = \{(x_i, y_i), i = 1, \ldots, n_0\}$ of input/output pairs. This can be done by minimizing an empirical risk:

$$\theta^* = \arg\min_{\theta \in \Theta} R(\theta, \mathcal{L}),$$
$$R(\theta, \mathcal{L}) = \sum_{(x_i, y_i) \in \mathcal{L}} \ell(h(x_i; \theta), y_i),$$

where $\ell(h(x_i; \theta), y_i)$ is the loss incurred when predicting $h(x_i; \theta)$ when y_i is the true label. This approach relies on the assumption that all data are distributed according to the same distribution $P(X, Y)$.

In domain adaptation [4], the initial data supposedly follow some *source* distribution $P_0(X, Y)$. The aim is to train a model able to process data from a *target* distribution $P_T(X, Y)$, possibly far away from P_0. However, generally, only unlabelled target data $\mathcal{U}_T = \{x_i, i = 1, \ldots, n_T\}$ with $x_i \sim P_T(X)$ are available: this impedes computing the best estimate θ_T^* for $P_T(X, Y)$ in a supervised way, and θ_0^* is generally suboptimal for $P_T(X, Y)$.

1.2 Gradual domain adaptation

In gradual domain adaptation [2], in addition to the labeled dataset $\mathcal{L}$ and the unlabeled target instances $\mathcal{U}_T$, subsets of "intermediate data" $\mathcal{U}_t = \{x_{ti}, i = 1, \ldots, n_t\}$ are available, with $x_{ti} \sim P_t(X)$: the underlying distribution "gradually" evolves from $P_0(X)$ to $P_T(X)$, after having gone through each P_t, $t \geq 1$.

The idea of this PhD proposal is to investigate leveraging these intermediate data to obtain gradual parameter estimates θ_t^*, $t = 0, \ldots, T$, so that a better estimate of θ_T^* can eventually be reached.

2 Research directions

2.1 Gradual domain adaptation

Our purpose is to propose to use *data imprecisiation* and *weakening* as a way to leverage the data in the previous steps (and in particular at step $t - 1$) leading to more robust estimates θ_t^*. This arguably requires a reliable self-supervised labelling process, to associate every unlabeled intermediate instance x_{ti} with a *pseudo-label* $h(x_{ti}; \theta_{t-1}^*)$, except for those associated with too ambiguous predictions, or those who are too far away from the distribution $P_{t-1}(X)$.

For instance, a simple way to exploit previous data, explored previously in another setting [7], would be to use the estimated model $h(\cdot; \theta_{t-1}^*)$ to label the instances x_{ti}, but to allow for *partial* (or *vague*) label predictions. Such partial predictions can be used to account for two kinds of uncertainty [3]: ambiguity (instances close to decision boundaries), and poor knowledge (outliers in low-density regions).

In order to leverage data from previous steps, we may allow them to "drift" towards the new distribution P_t (unfortunately known only up to $\mathcal{U}_t$). We propose to account for such a drift by "imprecisiating" past data into regions whose sizes increase with time. For instance, for some $x_{i,t-u}$ observed u periods before the current one, we may the set $X_{i,t-u} = x_{i,t-u} \pm u \cdot \delta$, i.e. an hyper-cube whose size increases as times passes by.

2.2 Possible application to probabilistic circuits

Probabilistic Circuits (PCs) are prominent class of tractable generative probabilistic models, among which sum-product networks [6, 8]. Complex probability distributions are modelled by combining simple distributions using products (via product nodes) and mixtures (via sum nodes)—Figure 1 shows a simple example. Training (and notably structure inference) remains an open problem. Arguably the most classical strategy, LearnSPN [1] recursively partitions the instances (to create sum nodes) or the variables (product nodes).

PCs are a powerful approach to probabilistic learning, since they can approximate a wide variety of distributions, and even arbitrarily complex ones by making the network "deep" [6]. A robust variant has been proposed, in which *sets of weights*, rather than single weights, are associated to the sum nodes [5].

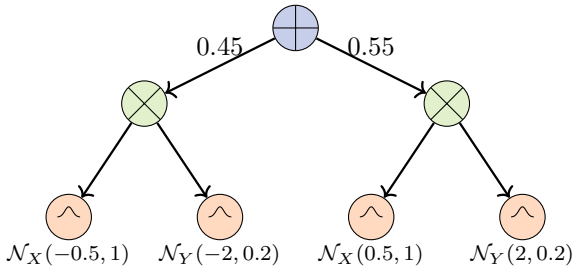

Figure 1: Example of SPN: the sum node (in blue) corresponds to a mixture (with weights 0.45 and 0.55) of the product nodes (in green), which themselves multiply the base distributions in the leaves (in salmon).

Due to their versatility, their ease of use, and their possible extension to robust learning, SPNs may be considered as a suitable candidate to the gradual domain adaptation setting considered in the PhD proposal. We may more particularly investigate

1. aligning the "robustification" of the PCs to the data relevance (determined by their drift and quality),
2. designing PCs with calibration guarantees, following previous research [7], so as to provide additional robustness guarantees,
3. developing interpretation mechanisms offering insights into the drift process, allowing for reducing (part of) the uncertainty in the final model.

3 Relation to PostGenAI@Paris

The PhD proposal addresses transfer learning through the prism of robust inference, with a focus on deep generative probabilistic models. Its ultimate purpose is to tailor trustful, robust and understandable models offering tractability guarantees.

Such models are particularly well-suited to high-stake applications, such as healthcare. As such, the PhD proposal perfectly aligns with the topics brought forward in the scientific project of PostGenAI@Paris.

4 Desired candidate profile

- solid background in probability, statistics, and machine learning; good knowledge in optimization;
- programming skills (preferably Python);
- autonomy, curiosity, keen interest for new topics.

5 Supervising team

This PhD proposal reunites two main participants of SCAI and PostGenAI@Paris, namely the université de technologie de Compiègne (UTC) and Sorbonne université (SU). The supervising team is composed of

- Benjamin Quost, professor, UTC, Heudiasyc laboratory;
- Sébastien Destercke, senior researcher, CNRS, Heudiasyc laboratory;
- Pierre-Henri Wuillemin, associate professor, SU, LIP6.

References

[1] R. Gens and P. Domingos. Learning the structure of sum-product networks. In S. Dasgupta and D. McAllester, editors, *Proceedings of ICML 2013*, volume 28 of *PMLR*, pages 873–880, Atlanta, Georgia, USA, 06 2013. PMLR.

[2] Y. He, H. Wang, B. Li, and H. Zhao. Gradual domain adaptation: theory and algorithms. arXiv preprint arXiv:2310.13852, 2023.

[3] E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.

[4] W. M. Kouw and M. Loog. A review of domain adaptation without target labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):766–785, 2019.

[5] D. D. Mauá, D. Conaty, F. G. Cozman, K. Poppenhaeger, and C. P. de Campos. Robustifying sum-product networks. *International Journal of Approximate Reasoning*, 101:163–180, 2018.

[6] H. Poon and P. Domingos. Sum-product networks: A new deep architecture. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 337–346, 2011.

[7] C. Rodriguez, V. M. Bordini, S. Destercke, and B. Quost. Self learning using Venn-Abers predictors. In *Proceedings of the 12th Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2023)*, volume 204, 2023.

[8] R. Sánchez-Cauce, I. París, and F. J. Díez. Sum-product networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(07):3821–3839, 07 2022.