Table of Contents

Paulus Stephanus

218047576

Assignment 4 – TAI911S (Trends in Artificial Intelligence and Machine Learning)

Paper Reviewed [24] Tai, Jin yang, and Yi ke Guo. 2024. "Hierarchical Linear Symbolized Tree-Structured Neural Processes." In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2818–29. KDD '24. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3637528.3671861.

1. Introduction

The advancement of deep learning has led to remarkable improvements in few-shot learning and probabilistic modeling, with Neural Processes (NPs) emerging as a key framework for learning distributions over functions. However, standard NP architectures often struggle with capturing structured uncertainty and hierarchical dependencies, especially in complex tasks that require multi-level abstraction. In response to these limitations, the research paper *"Hierarchical Linear Symbolized Tree-Structured Neural Processes"* by Tai and Guo (2024) proposes an enhanced NP variant HLNPs that incorporates symbolic tree structures to address these challenges. This review critically evaluates the paper's contributions, contextualizes it within existing literature, and discusses its experimental effectiveness and limitations.

2. Summary of the paper

2.1 Problem being addressed

Tai and Guo (2024) identify that NPs, although powerful in few-shot learning, struggle with modeling data that exhibit hierarchical or multi-level characteristics. They argue that traditional methods do not fully leverage positional or structural information within datasets. As a result, current NP models often underperform in tasks requiring layered abstraction, structured uncertainty modeling, and causal reasoning.

2.2 Main contribution of the work

Traditional Neural Processes (NPs) and their variants are limited in their ability to model complex data distributions due to their flat representation structure. They primarily learn from context points without incorporating multi-level or hierarchical information, which restricts their capacity to capture deep structural relationships and uncertainty in data.

2.3 Experimental/theoretical results

This paper proposes Hierarchical Linear Symbolized Tree-structured Neural Processes (HLNPs) a novel architecture that integrates hierarchical and symbolic tree structures into the Neural Processes framework. The core innovation is the use of a hierarchical linear symbolized tree to enrich representation learning and enhance the model's capacity to approximate complex latent distributions.

In their evaluation, Tai and Guo (2024) demonstrate HLNPs outperform standard NPs on various benchmarks including 1D function regression, Bayesian optimization, and 2D function approximation.

1D Function Regression: HLNPs demonstrated superior performance in estimating smooth, non-linear functions compared to standard NPs.

1. Bayesian Optimization**:** The model showed stronger exploration-exploitation tradeoffs, attributed to improved uncertainty modeling.
2. 2D Function Approximation: HLNPs achieved better generalization on sparse data distributions, benefiting from symbolic hierarchical decomposition.

The experimental results highlight the model's robustness and its ability to scale across tasks of varying complexity.

3. Literature review

The development of hierarchical and probabilistic models in deep learning has been influenced by several foundational works. One major contribution is the Variational Autoencoder (VAE), introduced by Kingma and Welling (2013), which combines variational inference with deep neural networks to learn latent variable models efficiently. The VAE framework approximates intractable posteriors and facilitates generative modeling through a reparameterization trick, laying the groundwork for scalable and expressive probabilistic systems. Complementing this, Sum-Product Networks (SPNs) by Poon and Domingos (2011) introduced a tractable, hierarchical probabilistic model capable of exact inference by organizing variables in a tree-like structure of sum and product operations. SPNs have proven useful in capturing complex distributions while preserving computational efficiency an inspiration for tree-structured neural processes.

The integration of symbolic reasoning and deep learning is further advanced by Deep Neural Decision Forests, proposed by Kontschieder et al. (2015), which merge decision trees with neural networks to provide both interpretability and end-to-end trainability. This hybrid structure is particularly useful in scenarios where model transparency and structured decision-making are crucial. Another notable architectural innovation is the Sparsely-Gated Mixture-of-Experts (MoE) model by Shazeer et al. (2017), which introduces a gating mechanism to activate only a subset of expert neural sub-networks for a given input. This approach significantly improves scalability and efficiency, especially in large-scale systems, and supports the kind of modular and compositional modeling that hierarchical neural processes aim to achieve.

Together, these models contribute core ideas probabilistic reasoning, hierarchical structuring, interpretability, and efficient modularity that underpin modern architectures like Hierarchical Linear Symbolized Tree-structured Neural Processes (HLNPs**)**. SPNs and VAEs offer complementary strengths in uncertainty modeling and latent variable learning, while neural decision forests and MoEs bring structure and scalability into focus. The interplay between symbolic structure (from SPNs and decision forests), probabilistic encoding (from VAEs), and selective computation (from MoEs) illustrates the multifaceted evolution of deep learning models, laying a robust theoretical and architectural foundation for the development of interpretable and expressive hierarchical neural frameworks.

Sum-Product Networks (SPNs) were introduced as a novel deep probabilistic architecture designed to enable efficient and exact inference in high-dimensional models. SPNs are defined as directed acyclic graphs with variables as leaves and a hierarchical arrangement of sum and product operations as internal nodes. A key insight of Poon and Domingos (2011) is that if an SPN is both *complete* and *consistent*, it can correctly compute the partition function and all marginals of a graphical model. This enables the architecture to support a wide range of probabilistic inferences with tractable complexity. The authors show that SPNs can generalize existing tractable models and that their structure naturally supports backpropagation and Expectation-Maximization (EM) learning algorithms.

In comparison to traditional graphical models and deep networks, SPNs offer significant advantages in efficiency and expressiveness. Unlike graphical models, which often suffer from exponential inference and learning time due to complex partition functions, SPNs leverage the distributive law to perform marginal computations using a polynomial number of operations. Furthermore, SPNs can represent many distributions more compactly than hierarchical mixture models or junction trees, thanks to their reuse of substructures and support for context-specific independence. The authors demonstrate that SPNs outperform existing deep models such as Deep Belief Networks (DBNs) and Deep Boltzmann Machines (DBMs) in tasks like image completion, offering both faster inference and superior accuracy (Poon & Domingos, 2011).

Attentive Neural Processes (ANPs) extend the foundational work of Neural Processes (NPs) by addressing a critical shortcoming: the tendency of NPs to underfit their own context data. This limitation arises because the standard NP architecture aggregates context representations into a fixed-length vector via mean-pooling, which ignores the relevance of specific context points to different queries. ANPs resolve this by incorporating attention mechanisms into both the deterministic and latent paths of the model, allowing each target input to dynamically attend to relevant context points. This results in enhanced predictive accuracy, particularly at the context locations, and improves model expressiveness without sacrificing the permutation invariance property of NPs (Kim et al., 2019).

In their experiments, the authors demonstrate that ANPs significantly outperform NPs in tasks such as 1D regression, 2D image inpainting, and Bayesian optimization. The introduction of crossattention and multihead self-attention mechanisms enables the model to learn both local and global dependencies within the context data. These enhancements allow the ANP to generalize more effectively across function families, speed up training time, and deliver sharper, more coherent predictions in high-dimensional visual data. The paper positions ANPs as a strong bridge between Gaussian Processes and deep neural networks, offering the flexibility of learned similarity functions with improved uncertainty modeling and representation learning (Kim et al., 2019).

Recent advancements in attention mechanisms, particularly the Transformer architecture (Vaswani et al., 2017), have inspired new directions in hierarchical probabilistic modeling by demonstrating how self-attention can replace traditional recurrence or convolution to capture global dependencies. This innovation has been integrated into probabilistic frameworks such as Variational Transformers and Attentive Neural Processes, facilitating finer control over latent structure and more expressive generative capabilities. These methods showcase how modular and interpretable structures can be built through the fusion of attention-driven architectures with probabilistic reasoning, allowing for flexible modeling of variable-sized input-output relationships while maintaining tractable inference.

Moreover, the evolution of representation learning (Bengio et al., 2013) has strengthened the theoretical underpinnings of hierarchical models by providing a framework for disentangling latent factors in data. Probabilistic approaches like VAEs and stochastic backpropagation (Rezende et al., 2014) further enrich this landscape by introducing principled ways to learn uncertainty-aware, hierarchical embeddings. These developments collectively suggest that the intersection of symbolic structure, probabilistic theory, and deep representation learning forms a powerful triad for building interpretable, scalable, and generalizable machine learning systems—an ethos central to modern architectures like HLNPs.

4. Limitations

Despite its promising design, HLNPs face several limitations. First, scalability remains a challenge, as the tree-based structure can result in computational inefficiencies when handling large datasets or high-dimensional tasks. Second, the process of learning symbolic functions is complex, lacking full interpretability and automation, thereby requiring additional methods to guide it effectively. Third, the evaluation scope is limited, with the absence of empirical validation on real-world datasets such as image or text data, which restricts the generalizability of the findings. Lastly, reproducibility is a concern due to insufficient details provided on implementation, hyperparameters, and ablation studies.

5. Conclusion

The paper by Tai and Guo (2024) introduces a significant advancement in the Neural Processes domain by embedding hierarchical symbolic representations into the modeling process. Their proposed HLNP architecture demonstrates superior performance across multiple benchmark tasks, particularly in function approximation and uncertainty modeling. Through a comprehensive analysis of its theoretical foundations and related work, it is evident that HLNPs stand on a solid conceptual lineage drawn from probabilistic models and structured deep learning architectures. Nonetheless, challenges remain in scaling the model, improving interpretability, and validating results on real-world datasets. Overall, this work represents a meaningful step forward in designing interpretable and scalable probabilistic learning systems, with ample opportunities for further research and application.

6. References

[24] Tai, Jin yang, and Yi ke Guo. 2024. "Hierarchical Linear Symbolized Tree-Structured Neural Processes." In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2818–29. KDD '24. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3637528.3671861

Poon, H., & Domingos, P. (2011). Sum-product networks: A new deep architecture. *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI-11)*, 337–346. Retrieved from https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/poon11.pdf

Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, S. M. A., Rosenbaum, D., Vinyals, O., & Teh, Y. W. (2019). Attentive neural processes. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1901.05761

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint* arXiv:1312.6114.

Kontschieder, P., Fiterau, M., Criminisi, A., & Bulo, S. R. (2015). Deep neural decision forests. *Proceedings of the IEEE International Conference on Computer Vision*, 1467–1475.

Poon, H., & Domingos, P. (2011). Sum-product networks: A new deep architecture. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 689–690.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint* arXiv:1701.06538.

Bengio, Y., Courville, A., & Vincent, P. (2013). *Representation learning: A review and new perspectives.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). *Stochastic backpropagation and approximate inference in deep generative models.* Proceedings of the 31st International Conference on Machine Learning (ICML), 1278–1286. https://arxiv.org/abs/1401.4082

Ghahramani, Z. (2015). *Probabilistic machine learning and artificial intelligence.* Nature, 521(7553), 452–459. https://doi.org/10.1038/nature14541

Gens, R., & Domingos, P. (2012). *Discriminative learning of sum-product networks.* Advances in Neural Information Processing Systems, 25. https://proceedings.neurips.cc/paper_files/paper/2012/file/beb1e79e0f994f74f4e7e382d6baf07d-Paper.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need.* Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762