Statement of Purpose

Peijia Qin *October 18, 2024*

Machine Learning (ML) and Deep Learning (DL) have become indispensable tools for addressing complex real-world challenges. During my undergraduate studies at the Southern University of Science and Technology (SUSTech) and my internship at the University of California, San Diego (UCSD), I engaged in research that spanned a broad spectrum of ML and its applications. My work encompassed data stream mining [1, 2], natural language processing [3], computer vision [4], and dynamic systems.

ML for Data Stream Mining My research began with online learning with varying feature space (VFS) in data stream mining under the supervision of Prof. Xin Yao and Prof. Liyan Song. Unlike offline learning, we focus on dynamic data streams where data samples arrive sequentially and are discarded after training. VFS problem, where the feature dimension of the sample changes over time, had not been explored in our group at that time. I conducted a comprehensive literature review and categorized existing works into three types based on feature variation patterns. Despite existing efforts, I identified a long-ignored issue: whether the presence of a feature at each time can provide information about the class label. To bridge this gap, I designed an algorithm that splits the data stream into two: the first stream includes all features with unobserved features padded to zero, and the second stream encodes feature variations using a binary bit for each feature. My approach allows existing methods to utilize feature variation information and was published as a conference paper [1].

During the literature review in my initial research, I gradually developed an interest in a related topic called online class imbalance learning. While extensively studied offline, the streaming nature introduces challenges as the imbalance status evolves dynamically. Existing works address the imbalance by making heuristic assumptions on tracking and utilizing imbalance status, which lack theoretical justification and often lead to sub-optimal results. I realized that optimizing models directly for performance metrics using evolutionary algorithms (EAs) could overcome this issue, and I began to work under Prof. Siang Yew Chong's supervision. Being the first to apply EA to online class imbalance, I reformulated the original problem into a bi-level optimization problem. The lower level trains the model without addressing class imbalance, while the upper level uses a cost vector to reweight prediction probabilities. An evolutionary algorithm with a dynamic objective optimizes this cost vector. This work was published in the TMLR journal [2].

Parallel to my work [2], I also independently explored applying Bayes filter methods to accurately estimate the class imbalance ratio between classes to make it more explainable. Common practice relies on a decay factor to exponentially diminish past knowledge of class ratios. However, this approach may be sub-optimal as the decay factor must be predefined. I proposed a hierarchical Bayesian approach to infer the most likely decay factor value through MAP estimation on the fly, controlling the forgetting rate in the Bayes filter. Experimental results demonstrate improved tracking accuracy. This work is still ongoing now, focusing on enhancing the efficiency of posterior inference, which is essential for rapid streaming scenarios.

ML for Natural Language Processing During my internship under the supervision of Prof. Pengtao Xie at UCSD, I gradually realized that the bi-level optimization method extends far beyond the applications of my previous work. In fact, a substantial body of literature has focused on gradient-based bi-level optimization frameworks. My initial application of this gradient-based framework was to enhance a parameter-efficient fine-tuning method for a large language model named DoRA. I discovered that while DoRA enhances the well-known LoRA by explicitly reformulating parameters into two components, it restricts learning capacity because the two components are still trained in a coupled and joint manner. During my experiments, I found that training the two parts of DoRA in distinct layers on different subsets significantly improves learning capacity and resistance to overfitting. This work is currently under review at ICLR 2025 [3].

ML for Computer Vision Another successful application of bi-level optimization occurred when I was working in Prof. Jianguo Zhang's group at SUSTech. During that time, I focused on vector-quantized networks to learn discrete representations within a codebook, specifically targeting image generation. Training the codebook has long been challenging due to the non-differentiable nature of the quantization layer. I discovered that bi-level optimization can significantly enhance sparse gradient guidance. Inspired by the well-established meta-learning approach, I proposed treating the codebook as hyperparameters while other parts of the network remain as parameters. I observed that applying hypergradients to the codebook induces a nontrivial and useful gradient pattern unattainable by existing methods. Moreover, I conducted a detailed gradient analysis, demonstrating that our method resolves three key issues of the original approach within a cohesive framework. This work is also under review at ICLR 2025 [4].

ML for Dynamic Systems Complementary to my other research interests, I also engaged in machine learning for dynamic systems. I started an internship under the supervision of Prof. Yuanyuan Shi at UCSD, focusing specifically on the time delay control problem. Real-time control of complex systems with delayed input is time-consuming because the common practice involves repeatedly computing temporal integration of the system dynamics. Initially, we replaced temporal integration with a neural operator trained in a data-driven manner. While effective in some cases, I discovered that the model was particularly vulnerable to out-of-distribution scenarios in certain experiments. This vulnerability is especially detrimental for online control, as errors at each time step can accumulate and lead to system collapse. To address this, I applied adaptive conformal prediction, a lightweight uncertainty quantification method, to assess the reliability of the model's predictions. This approach transforms the system into a switching system with two subsystems, where the conformal prediction determines whether to use the model or switch to the slower but more accurate numerical method. However, theoretical justification for the stability of the switching system is still needed, and this work is currently in progress.

Future Plans In the long term, I aim to become a leading researcher in the field of machine learning. I intend to work on classic problems in machine learning and deep learning, including data curation (data selection, reweighting, and cleansing), AutoML (meta-learning, NAS), knowledge transfer (transfer learning, domain adaptation, multi-task learning), and deep probabilistic models. I am also enthusiastic about applying these methods in various fields, including but not limited to those mentioned in my previous research experiences.

References

[1] Peijia Qin and Liyan Song. Online learning in varying feature spaces with informative variation. In *13th International Conference on Intelligent Information Processing*, 2023.

[2] Peijia Qin, Shuxian Li, Xiaoqun Liu, Zubin Zheng, and Siang Yew Chong. Threshold moving for online class imbalance learning with dynamic evolutionary cost vector. *Transactions on Machine Learning Research*, 2024.

[3] Peijia Qin, Ruiyi Zhang, and Pengtao Xie. Bidora: Bi-level optimization-based weight-decomposed low-rank adaptation. In *Under review at ICLR 2025*, 2024.

[4] Peijia Qin and Jianguo Zhang. Mq-vae: Training vector-quantized networks via meta learning. In *Under review at ICLR 2025*, 2024.