Automated machine learning for neural network verification

König, H.M.T.

Summary

In an era marked by the widespread deployment of neural networks, the criticality of ensuring their reliability and safety cannot be overstated. This thesis delves into the sphere of neural network verification, a crucial yet challenging task aimed at proving the correctness of neural network models. Moreover, this thesis seeks to introduce novel techniques and strategies that could significantly improve the efficiency of neural network verification algorithms, thereby contributing to the development of more secure and reliable deep learning applications.

In Chapter 3, the thesis presents a detailed examination of the current landscape in neural network verification, specifically focusing on local robustness verification. The diversity within the field, both in terms of verification techniques and neural network architectures, presents a complex challenge for practitioners aiming to ensure the safety and reliability of these systems. We conduct a thorough empirical analysis of several prominent verification algorithms, revealing the fragmented nature of the state of the art. The findings suggest that there is no single dominant algorithm that excels across all types of verification instances. Instead, the performance of these algorithms is highly complementary, indicating the potential benefits of employing algorithm portfolios to enhance verification efficiency. This nuanced view highlights the importance of considering a range of methods and techniques to address the varied challenges posed by different neural network configurations and verification properties.

Following this, Chapter 4 delves into the realm of mixed integer programming (MIP)-based neural network verification, with a particular focus on speeding up the verification process through automated algorithm configuration. The chapter introduces novel approaches to harness algorithm portfolios, demonstrating how the performance of MIP-based verification systems can be significantly improved. By adopting automated configuration techniques, we present a systematic approach to tailor verification methods more closely to the specific characteristics of a given neural network, thereby reducing

computational costs and enhancing the efficiency of the verification task.

Chapter 5 introduces a novel perspective on verification by exploring the predictability of the running time of various verification algorithms for specific problem instances. This approach seeks to allocate computational resources more judiciously, focusing efforts on instances with a higher likelihood of being solvable within reasonable time budgets. The integration of running time predictions into the verification process represents a strategic shift towards more resource-aware methodologies, potentially transforming the efficiency and applicability of neural network verification in real-world scenarios.

In the final thematic chapter, Chapter 6, the research focuses on the task of robust model selection within the domain of adversarial robustness. We propose a sophisticated racing algorithm that leverages simple yet novel heuristics to efficiently determine the most robust neural network model from a given set. This method not only streamlines the model selection process but also significantly reduces the computational overhead associated with evaluating multiple candidate models.

Therefore, we have demonstrated how automated machine learning, or meta-algorithmic approaches in general, can improve the performance and practical efficiency of neural network verification systems, thereby contributing to a safer and more reliable usage of deep neural networks in the evolving landscape of artificial intelligence applications.