

Abstract

In recent years, the rapid growth of artificial intelligence systems and increasingly larger Language Models has led to significant advances in a wide range of natural language processing applications. Despite their strong performance, these approaches are often associated with substantial computational demands, extensive training times, and a heavy reliance on large volumes of annotated data, which limits their practicality in low- and mid-resource environments. The prevailing trend toward ever larger model architectures further amplifies these limitations and raises concerns regarding efficiency, accessibility, and sustainability.

This thesis examines contemporary approaches to improving the efficiency and robustness of model fine-tuning under constrained data and resource conditions. It focuses on strategies that emphasize informed data selection, adaptive training procedures, and the combination of complementary methodologies rather than sole reliance on architectural scaling. A unified framework is proposed that integrates uncertainty-aware sample selection, structure-based data organization, and dynamic scheduling mechanisms, with optional extensions for enhanced uncertainty modeling.

Through a series of empirical evaluations across multiple languages and settings, the work demonstrates that carefully designed combinations of data-efficient learning techniques can substantially reduce annotation requirements while maintaining or improving predictive performance and training stability. The findings suggest that synergistic integration of modern learning strategies represents a promising direction for developing more efficient, accessible, and reliable artificial intelligence systems in resource-constrained scenarios.

Keywords: Active Learning, low-resource NLP, uncertainty sampling, Epistemic Neural Networks, Agglomerative clustering, dynamic scheduling, Sentiment Analysis, Named Entity Recognition, BERT