Generative Transformers for Structural Transformation in Language and Logic

Master's Thesis – Abstract

Máté Gedeon

2025

The widespread adoption of pre-trained transformer models has established transfer learning as the standard approach for achieving state-of-the-art performance across a wide range of machine learning domains, including but not limited to natural language processing. This thesis investigates whether custom transformer architectures, trained from scratch on task-specific data, can perform competitively against the fine-tuned versions of established pre-trained models.

We conduct experiments in two distinct domains: morphological tagging—a classic NLP task where pre-trained models typically excel due to their broad linguistic knowledge—and CNF (Conjunctive Normal Form) conversion—a symbolic reasoning task that demands logical transformation capabilities. Our evaluation compares three transformer-based models: a fine-tuned T5, a sequence-to-sequence-adapted RoBERTa, and a custom transformer trained from scratch.

For morphological tagging, we built a dataset of over 55,000 English words with automated annotations from the EMOR tool—framing it as a sequence-to-sequence task—mapping from words to morphological tags. In the CNF experiments, we used Boolean expressions from the SemVec project, training models to convert arbitrary logical formulas into CNF directly or through structured, step-by-step transformations.

Results show that custom architectures can achieve highly competitive performance without the benefits of pre-training. In morphological tagging, the custom transformer reached 98.7% accuracy—1.3% below the 100% scores of pre-trained models. Remarkably, in CNF conversion, the custom model outperformed both pre-trained alternatives, achieving 83.0% accuracy versus 73.6% for T5 and 71.7% for RoBERTa. Moreover, introducing structured, intermediate steps consistently improved performance across all models, with gains ranging from 4.3% to 15.1%.

These findings suggest that model choice should be driven by task characteristics rather than an assumed superiority of pre-trained models. For specialized tasks—particularly those requiring symbolic reasoning—custom architectures trained from scratch can not only match but sometimes exceed the performance of their pre-trained counterparts, while offering greater interpretability and control.