

Hierarchical Text Classification using Language Models with Global Label-wise Attention Mechanisms

Jaco du Toit^{1,2}[0000–0001–6782–2381] and Marcel Dunaiski^{1,2}[0000–0003–1957–3979]

¹ Department of Mathematical Science, Computer Science Division
Stellenbosch University, Stellenbosch, South Africa

² School for Data Science and Computational Thinking
Stellenbosch University, Stellenbosch, South Africa
{jacodutoit,marceldunaiski}@sun.ac.za

Abstract. Hierarchical text classification (HTC) is a natural language processing task with the objective of categorising text documents into a set of classes from a structured class hierarchy. Recent HTC approaches combine encodings of the class hierarchy with the language understanding capabilities of pre-trained language models (PLMs) to improve classification performance. Furthermore, label-wise attention mechanisms have been shown to improve performance in HTC tasks by placing larger weight on more important parts of the document for each class individually to obtain label-specific representations of the document. However, using label-wise attention mechanisms to fine-tune PLMs for downstream HTC tasks has not been comprehensively investigated in previous work. In this paper, we evaluate the performance of a HTC approach which adds label-wise attention mechanisms, along with a label-wise classification layer, to a PLM which is subsequently fine-tuned on the downstream HTC dataset. We evaluate several existing label-wise attention mechanisms and propose an adaptation to one of the approaches which separates the attention mechanisms for the different levels of the hierarchy. Our proposed approach allows the prediction task at a certain level to leverage the information gained from the predictions performed at all ancestor levels. We compare the different label-wise attention mechanisms on three HTC benchmark datasets and show that our proposed approach generally outperforms the other label-wise attention mechanisms. Furthermore, we show that without using the complex techniques proposed in recent HTC approaches, our relatively simple approach outperforms state-of-the-art approaches on two of the three benchmark datasets.

Keywords: Hierarchical Text Classification · Large Language Models · Attention Mechanisms