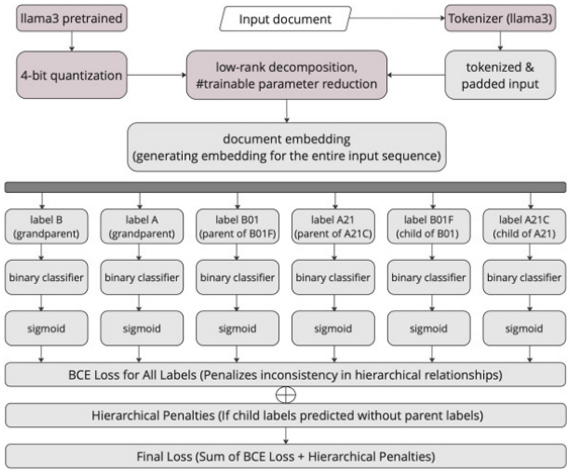
ViRVIG
Visualització, Realitat Virtual i Interacció Gràfica

www.virvig.eu
C. Llorens i Artigas 4-6
Edifici U, 1er pis
E- 08028- Barcelona

Improved multi-label hierarchical patent classification using LLMs

Rafieian, Bardia; Vazquez, Pere-Pau



Classifying multi-label documents has always been a challenging task, especially when the labels follow a hierarchical structure. This complexity increases the difficulty of accurately predicting multiple interrelated labels, often limiting the overall classification performance. To address these challenges and improve the accuracy of hierarchical multi-label classification systems, we introduce a novel pipeline leveraging large language models (LLMs). By incorporating quantization techniques and optimizing weight updates through smaller matrices, we improve computational efficiency and scalability. Our approach demonstrates an improvement in accuracy compared to current state-of-the-art models. In this work, we apply this pipeline to patent documents, focusing on the multi-label hierarchical text classification problem using a transformer-based architecture. The hierarchical structure of the CPC (Cooperative Patent Classification) labels is preserved through a graph-based taxonomy, which enables more effective processing of patent categories. Our model is trained and evaluated on the USPTO-70k dataset, and we achieve substantial improvements across various metrics, including precision, recall, F1-score, and AUC.

http://dx.doi.org/10.1016/j.wpi.2025.102356

UPC