Código: PPGEC-MESTRADO_2026_2_BLDB3

Universidade de Pernambuco
Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Inteligência Computacional

Título: Uso de Grafos Multimodais e Transformers para Extração de Informações em Documentos

Orientador – Byron Leite Dantas Bezerra (byron.leite@upe.br)

Descrição

A utilização de modelos de Inteligência Artificial em Processamento de Linguagem Natural tem permitido avanços importantes na extração automática de informações em diferentes tipos de documentos. Tarefas como reconhecimento de entidades nomeadas, extração de relações e classificação de documentos têm sido amplamente investigadas nos últimos anos, principalmente a partir do uso de arquiteturas baseadas em Transformers, como BERT e modelos derivados [1].

Apesar dos resultados obtidos, muitos problemas reais apresentam limitações que não podem ser resolvidas apenas por informações textuais. Diversos documentos possuem informações distribuídas em múltiplas fontes como texto, imagens, elementos visuais, estrutura espacial e contexto semântico. Contratos, formulários, documentos históricos digitalizados, currículos e prontuários médicos são exemplos em que a organização visual e a relação entre os elementos podem ser tão importantes quanto o conteúdo textual. Nesse contexto, modelos voltados à compreensão de documentos visualmente ricos passaram a integrar texto, layout e imagem em arquiteturas multimodais baseadas em Transformers, como LayoutLM, LayoutLMv2 e LayoutLMv3 [2], [3], [4]. Esses modelos demonstram que a incorporação de informações espaciais e visuais pode melhorar o desempenho em tarefas de entendimento de documentos.

Além dessas abordagens, estudos recentes têm explorado o uso de grafos para representar relações estruturais presentes em documentos. Nessa abordagem, diferentes elementos podem ser modelados como nós, incluindo palavras, regiões visuais, componentes do layout e conceitos semânticos externos. As conexões entre esses elementos permitem representar relações espaciais, semânticas e contextuais, tornando possível incorporar informações adicionais ao processo de aprendizado. Trabalhos como GraphIE e PICK exploram o uso de grafos para extração de informações, modelando relações não sequenciais entre elementos textuais e estruturais dos documentos [5], [6]. Mais recentemente, abordagens como GraphDoc e DocGraphLM passaram a combinar representações multimodais, mecanismos de atenção em grafos e modelos de linguagem pré-treinados para tarefas de compreensão de documentos [7], [8]. Esses trabalhos indicam que a combinação entre Transformers e grafos pode ser promissora para capturar relações que não são adequadamente modeladas por sequências lineares ou apenas por embeddings posicionais.

Dessa forma, surge a seguinte questão de pesquisa que se pretende responder neste projeto: como modelos neurais baseados em Transformers enriquecidos com grafos multimodais podem melhorar o desempenho em tarefas de extração automática de informações em documentos quando comparados a abordagens puramente textuais ou apenas multimodais? É neste contexto que reside o projeto de mestrado aqui proposto. A proposta envolve uma equipe multidisciplinar e faz parte do projeto de pesquisa e inovação "***Algoritmos e Modelos de Inteligência Artificial e Visão Computacional para Processamento Inteligente de Documentos***" fomentado pelo CNPQ, e em parceria com a empresa Di2Win (www.di2win.com). Para conhecer mais sobre o orientador e seus temas de pesquisa, convido a assistir a entrevista aqui.

Referências Bibliográficas

[1] DEVLIN, Jacob et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL, 2019.

[2] XU, Yiheng et al. LayoutLM: Pre-training of Text and Layout for Document Image Understanding. KDD, 2020.

[3] XU, Yang et al. LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding. ACL, 2021.

[4] HUANG, Yupan et al. LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. ACM MM, 2022.

[5] QIAN, Yujie et al. GraphIE: A Graph-Based Framework for Information Extraction. NAACL, 2019.

[6] YU, Wenwen et al. PICK: Processing Key Information Extraction from Documents using Improved Graph Learning-Convolutional Networks. ICPR, 2020.

[7] ZHANG, Zilong et al. GraphDoc: Multimodal Pre-training Based on Graph Attention Network for Document Understanding. IEEE Transactions on Multimedia, 2023.

[8] YE, Junwen et al. DocGraphLM: Documental Graph Language Model for Information Extraction. SIGIR, 2023.