World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:11, 2022

Neural Graph Matching for Modification Similarity Applied to Electronic Document Comparison

Authors : Po-Fang Hsu, Chiching Wei

Abstract : In this paper, we present a novel neural graph matching approach applied to document comparison. Document comparison is a common task in the legal and financial industries. In some cases, the most important differences may be the addition or omission of words, sentences, clauses, or paragraphs. However, it is a challenging task without recording or tracing the whole edited process. Under many temporal uncertainties, we explore the potentiality of our approach to proximate the accurate comparison to make sure which element blocks have a relation of edition with others. In the beginning, we apply a document layout analysis that combines traditional and modern technics to segment layouts in blocks of various types appropriately. Then we transform this issue into a problem of layout graph matching with textual awareness. Regarding graph matching, it is a long-studied problem with a broad range of applications. However, different from previous works focusing on visual images or structural layout, we also bring textual features into our model for adapting this domain. Specifically, based on the electronic document, we introduce an encoder to deal with the visual presentation decoding from PDF. Additionally, because the modifications can cause the inconsistency of document layout analysis between modified documents and the blocks can be merged and split, Sinkhorn divergence is adopted in our neural graph approach, which tries to overcome both these issues with many-to-many block matching. We demonstrate this on two categories of layouts, as follows., legal agreement and scientific articles, collected from our real-case datasets.

Keywords : document comparison, graph matching, graph neural network, modification similarity, multi-modal
Conference Title : ICDAR 2022 : International Conference on Document Analysis and Recognition
Conference Location : Jerusalem, Israel
Conference Dates : November 29-30, 2022