A Transformer-Based Question Answering Framework for Software Contract Risk Assessment

Authors : Qisheng Hu, Jianglei Han, Yue Yang, My Hoa Ha
Abstract : When a company is considering purchasing software for commercial use, contract risk assessment is critical to identify risks to mitigate the potential adverse business impact, e.g., security, financial and regulatory risks. Contract risk assessment requires reviewers with specialized knowledge and time to evaluate the legal documents manually. Specifically, validating contracts for a software vendor requires the following steps: manual screening, interpreting legal documents, and extracting risk-prone segments. To automate the process, we proposed a framework to assist legal contract document risk identification, leveraging pre-trained deep learning models and natural language processing techniques. Given a set of pre-defined risk evaluation problems, our framework utilizes the pre-trained transformer-based models for question-answering to identify risk-prone sections in a contract. Furthermore, the question-answering model encodes the concatenated question-contract text and predicts the start and end position for clause extraction. Due to the limited labelled dataset for training, we leveraged transfer learning by fine-tuning the models with the CUAD dataset to enhance the model. On a dataset comprising 287 contract documents and 2000 labelled samples, our best model achieved an F1 score of 0.687.
Keywords : contract risk assessment, NLP, transfer learning, question answering
Conference Title : ICDAR 2022 : International Conference on Document Analysis and Recognition
Conference Location : Jerusalem, Israel
Conference Dates : November 29-30, 2022