Title	法的なテキスト処理における注意深いニューラルネットワークの改善に向けて
Author(s)	NGUYEN, HA THANH
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17789
Rights	
Description	Supervisor:NGUYEN,Minh Le, 先端科学技術研究科, 博士

氏　　　　　　　　名	NGUYEN, Ha Thanh		
学　位　の　種　類	博士（情報科学）		
学　位　記　番　号	博情第 466 号		
学 位 授 与 年 月 日	令和 4 年 3 月 24 日		
論　文　題　目	Toward Improving Attentive Neural Networks in Legal Text Processing		
論　文　審　査　委　員	Nguyen Minh Le (main examiner)	JAIST	Professor
	Satoshi Tojo	JAIST	Professor
	Kiyoaki Shirai	JAIST	Assoc. Prof.
	Ken Satoh	NII	Professor
	Shinobu Hasegawa	JAIST	Professor

論文の内容の要旨

In recent years, thanks to breakthroughs in neural network techniques especially attentive deep learning models, natural language processing has made many impressive achievements. However, automated legal word processing is still a difficult branch of natural language processing. Legal sentences are often long and contain complicated legal terminologies. Hence, models that work well on general documents still face challenges in dealing with legal documents. We have verified the existence of this problem with our experiments in this work. In this dissertation, we selectively present the main achievements in improving attentive neural networks in automatic legal document processing. Language models tend to grow larger and larger, though, without expert knowledge, these models can still fail in domain adaptation, especially for specialized fields like law.

This dissertation has three main tasks to achieve the goal of improving attentive models in legal document processing. First, we survey and verify the factors affecting the performance of the models when operating on a specific domain such as law. This investigation is to provide clearer insights to improve models in this domain. Second, as pretrained language models are recently the most well-known attentive approaches in natural language processing, we provide methods to create language models specific to the legal domain, producing state-of-the-art results on reliable datasets. These models are built on features from the data of legal documents, with the goal of overcoming the challenges found in our previous survey. Third, besides the approach to let the model learn completely from raw data, we propose and prove the effectiveness of using different knowledge sources to inject into the model in different ways to adjust their output. This approach not only increases explainability but also allows humans to control pretrained language models and take advantage of the knowledge resources available during the development of the field such as vocabulary, grammar, logic and law.

論文審査の結果の要旨

The thesis proposes a new method of legal analyzing with an attentive neural network. I would like to summarize the significant contributions of the thesis as follows.

(1) The first contribution is investigating how pre-trained models such as BERT can be applied for the legal domain. During the experiment with BERT, the candidate proposed using a parallel corpus to enhance the performance of the legal method based on deep learning.

(2) Typically, the legal sentence is long and complicated, making it difficult to exploit a transformer model. The proposed method can deal with encoding long and complex sentences using the encoding paragraph. The thesis presents a paraformer model which aims at encoding paragraphs in the legal domain. Experiments are conducted on benchmark data showed that the proposed method is effective. As a result, it leads to the best system in COLIEE 2021 competition.

(3) The candidate proposed HYDRA for training the transformer model's attention heads separately and then grafting them onto the original body for linguistic knowledge. To incorporate knowledge in legal structure, the candidate proposed using logical parts of legal which are injected into the different layers of the transformer model. This approach leads to cost-effectiveness in training and supporting users in terms of controllable the performance of legal applications.

Overall, the thesis creates a new idea for dealing with the legal domain using deep learning models. In addition, the proposed solutions can also apply to other language domains, such as for biomedical domain and manual documents.

This is an excellent dissertation, and we approve awarding a doctoral degree to Mr. Nguyen Ha Thanh.