

Title	法的テキストのためのテキスト含意認識及び構造解析にディープラーニングを使用する
Author(s)	Nguyen, Truong Son
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15526
Rights	
Description	Supervisor:NGUYEN, Minh Le, 情報科学研究科, 博士

氏名	NGUYEN, Truong Son		
学位の種類	博士(情報科学)		
学位記番号	博情第 394 号		
学位授与年月日	平成 30 年 9 月 21 日		
論文題目	Structure Analysis and Textual Entailment Recognition for Legal Text using Deep Learning		
論文審査委員	主査	Nguyen Minh Le	JAIST Assoc. Prof
		Satoshi Tojo	JAIST Professor
		Kiyooki Shirai	JAIST Assoc. Prof
		Akira Shimazu	JAIST Professor Emeritus
		Ken Satoh	NII Professor

論文の内容の要旨

Analyzing the structure of legal documents and recognizing textual entailment in legal texts are essential tasks to understand the meaning of legal documents. They benefit question answering, text summarization, information retrieval and other information systems in the legal domain. For example, recognizing textual entailment is an essential component in a legal question answering system which answers the correctness of user's statements, or a system which checks the contradiction and redundancy of a newly enacted legal article. Analyzing the structure of legal texts has broader applications because it is one of the preliminary and fundamental tasks which support other tasks. It can break down a legal document into small semantic parts so other systems can understand the meaning of the whole legal document easier. An information retrieval system can leverage a structure analysis component to build a better engine by allowing to search on specific regions instead of searching on the whole legal document.

In this dissertation, we study deep learning approaches for analyzing structures and recognizing textual entailment in legal texts. We also leverage the results of the structure analysis task to improve the performance of RTE task. Both of the results are integrated into a demonstrated system which is an end-to-end question answering system which can retrieve relevant articles and answer from a given yes/no question.

In the work on analyzing the structure of legal texts, we address the problem of recognizing requisite and effectuation (RRE) parts because RE parts are special characteristics of legal texts which different from texts in other domains. Firstly, we propose a deep-learning model based on BiLSTM-CRF, which can incorporate engineering features such as Part-of-Speech and other syntactic-based features to recognize non-overlapping RE parts. Secondly, we propose two unified models for recognizing overlapped RE parts including Multilayer-BiLSTM-CRF and Multilayer-BiLSTM-MLP-CRF. The advantages of proposed models are that

they possess a convenient design which can train only a unified model to recognize all overlapped RE parts. Besides, it can reduce the redundant parameters, so the training time and testing time are reduced significantly, but the performance is also competitive. We experimented our proposed models on two benchmark datasets including the Japanese National Pension Law RRE and Japanese Civil Code RRE which are written in Japanese and English, respectively. The experimental results demonstrate the advantages of our model. Our model achieves significant improvements compared to previous approaches on the same feature set. Our proposed model and its design can be extended to use other features easily without changing anything.

We then study the deep learning models for recognizing textual entailment (RTE) in legal texts. We encounter the lack of labeled data problem when applying deep learning models. Therefore, we proposed a semi-supervised learning approach with an unsupervised method for data augmentation which is based on syntactic structures and logical structures of legal sentences. The augmented dataset then is combined with the original dataset to train entailment classification models.

RTE in legal texts is also challenging because legal sentences are long and complex. Previous models use the single-sentence approach which considers related articles as a very long sentence, so it is difficult to identify important parts of legal texts to make the entailment decision. We then propose methods to decompose long sentences in related articles into simple units such as a list of simple sentences, or a list of RE structures and propose a novel deep learning model that can handle multiple sentences instead of single sentences. The proposed approaches achieve significant improvements compared to previous baselines on the COLIEE benchmark datasets.

We finally connect all components of structure analysis and recognizing textual entailment into a demonstration system which is a question answering system that can answer yes/no question in the legal domain on the Japanese Civil Code. Given a statement which a user needs to check whether or not it is correct, the demonstration system will retrieve relevant articles and classify whether the statement is entailed from its relevant articles. Building these systems can help ordinary people and law experts can exploit information in legal documents more effective.

Keywords: Recognizing textual entailment, Natural Language Inference, Legal Text Analysis, Legal Text Processing, Deep learning, Recurrent Neural Network, Recognizing Requisite and Effectuation.

論文審査の結果の要旨

The thesis aims at dealing with the challenging problem in analyzing the structure of legal text document and its application to textual entailment recognition and question and

answering system. The research consists of the major contributions which can be briefly summarized as follows.

(1) Analyzing the requisite part and effectuation parts in legal sentences

Various deep learning techniques are applied to solve the problem of analyzing legal sentences to requisite and effectuation parts. The significant method is the exploiting of the multi-layer Bi-LSTM-CRF model when dealing with the overlapping issues when parsing a legal sentence into requisite parts and effectuation parts structure.

(2) Textual entailment recognition in the legal domain

This contribution aims at solving the problem of lack of labeled data problem when applying deep learning models. The thesis proposed a semi-supervised learning approach with an unsupervised method for data augmentation using syntactic and logical structures of legal sentences. The proposed method significantly improves the performance of textual entailment recognition when applying to the legal domain. In addition to this, the thesis proposes a decomposition method for dealing with long sentences in textual entailment for the legal domain.

As a result, Nguyen Truong Son successfully defended his thesis with the agreement of the committee members. The committee concludes that the results contribute to Legal Engineering are valuable and significant. Besides, the candidate also shows his ability in working as an independent researcher for dealing with legal documents. He submitted two international journal articles, in which one paper is published. He also published many international conference papers. In conclusion, this is an excellent dissertation and we approve awarding a doctoral degree to Mr. Nguyen Truong Son.