Title	A Study of Deep Learning for Legal Question Answering Systems
Author(s)	Tran, DUC VU
Citation	
Issue Date	2016-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/13738
Rights	
Description	Supervisor: NGUYEN, Minh Le, ,

Abstract

Question answering systems have been developed with the goal to produce high quality and human-like answers. Especially, question answering systems on legal domain are really practical. However, building legal question answering systems with good quality is quite a hard challenge. Not only reading and understanding legal terms requires expertise knowledge, but also legal language is at different level of complication.

We aim to build an appropriate deep learning model for legal question answering systems and particularly target the document re-ranking phase of the systems. With deep learning is emerging as a powerful machine learning approach, we evaluate several deep learning architectures for modeling text into computational vector space so as to compare them with similarity measures.

While there are numbers of deep learning architectures, we are interested in Convolutional Neural Networks for the task. The models have strong performance in Computer Vision, especially object recognition. Models based on Convolutional Neural Networks are also implemented for Natural Language Processing and have been gaining more improvement so as more attention, though, it's probably quite strange that the models treat a text as an image. Despite of the nature of deep learning models, which require no manual feature extraction, we investigate the capability of deep learning models with additional features for further improvement.

We evaluate the models on three question answering corpora with various characteristics. One of them is on open domain and the other two are on legal domain. Though they are all question answering related tasks, the actual requirements are quite distinguishable. Interestingly, one of our experimental corpora is limited in data size which is a theoretically and practically huge disadvantage for deep learning models which are data-oriented and hungry for observable instances. We then use a strong feature that can be extracted from the corpus to enrich the models. While combining additional features is not new, we experiment in a different way. We set the additional features as starting point of learning for the models. The experiment results show that deep learning models with appropriate configuration can perform well on legal domain data.

In the future, we'd like to further investigate the structural information of complicated texts like legal documents and building a deep architecture which is capable of encoding structural information into its parameters for more robust document modeling.

Keywords: Deep Learning, Question Answering, Legal Domain, Document Re-ranking