Applying Deep Neural Network to Retrieve Relevant Civil Law Articles

Nga Tran Anh Hang

University of Evora / Rua Romao Ramalho 59

Evora, Portugal 7000-671 `m38310@alunos.uevora.pt`

Abstract

The paper aims to achieve the legal question answering information retrieval (IR) task at Competition on Legal Information Extraction/Entailment (COLIEE) 2017. Our proposal methodology for the task is to utilize deep neural network, natural language processing and word2vec. The system was evaluated using training and testing data from the competition on legal information extraction/entailment (COLIEE). Our system mainly focuses on giving relevant civil law articles for given bar exams. The corpus of legal questions is drawn from Japanese Legal Bar exam queries. We implemented a combined deep neural network with additional features NLP and word2vec to gain the corresponding civil law articles based on a given bar exam 'Yes/No' questions. This paper focuses on clustering words-with-relation in order to acquire relevant civil law articles. All evaluation processes were done on the COLIEE 2017 training and test data set. The experimental result shows a very promising result.

1 Credits

This approach was used to participate in COL-IEE, an annual competition for legal Information Extraction (IE) and Retrieval. In 2017, the International Conference on Artificial Intelligence and Law (ICAIL) invited participation in a competition on legal information extraction/entailment. The COLIEE competition held 3 competitions (COLIEE 2014-2016) on legal data collection with the JURISIN workshop, and they helped establish a major experimental effort in the legal information extraction/retrieval field. We will hold the fourth competition (COLIEE-2017) in 2017 with the ICAIL conference, and the motivation for the competition is to help create a research community of practice for the capture and use of legal information. In the context of the competition on legal information extraction/entailment (COLIEE), the main approach is to build a system to address legal question answering

and relevant civil law articles.

2 Introduction

The main goal of the project is to find the relevant civil law article for a given yes/no legal question to a Japanese exam bar query. Legal Question Answering is a form of textual entailment problem. The major achievement is to capture the relationship between legal question and civil law articles. To process the system, firstly, we have to identify informative sentences/keywords from a given question by applying part of speech tagging and name entities recognition. Secondly, we retrieve related documents based on extracted sentences/keywords. Eventually, the system has to compare the semantic connection between questions and relevant sentences/keywords and make a decision as to whether they have a real relationship or not in order to give the final answer.

3 Task & Materials

Japanese civil law articles (English translation besides Japanese) will be provided, and training data consists of pairs of a query and relevant articles. The process of executing the queries over the articles and generating the experimental runs should be entirely automatic. Test data will include only queries but no relevant articles. We have 600 training corpus and 73 test corpus in total.

3.1 Information Retrieval Task

The task investigates the performance of systems that search a static set of civil law articles using previously-unseen queries. The goal of the task

is to return relevant articles in the collection to a query. We call an article "Relevant" to a query if the query sentence can be answered Yes/No, entailed from the meaning of the article. If combining the meanings of more than one article (e.g., "A", "B", and "C") can answer a query sentence, then all the articles ("A", "B", and "C") are considered "Relevant". If a query can be answered by both an article "D" and another article "E" independently, we also consider both of the articles

"D" and "E" to be "Relevant". This task requires the retrieval of all the articles that are relevant to answering a query.

3.2 Corpus Structure

The structure of the test corpora is derived from a general XML representation developed for use in RITEVAL, one of the tasks of the NII Testbeds and Community for Information access Research (NTCIR) project. The RITEVAL format was developed for the general sharing of information retrieval on a variety of domains.

3.2.1 Example

$< pairlabel = "Y"id = "H18 - 1 - 2" >$
$< t1 >$
(Seller's Warranty in cases of Superficialities or Other Rights)Article 566 (1)In cases where the subject matter of the sale is encumbered with for the purpose of a superficialities, an emphysematous, an easement, a right of retention or a pledge, if the buyer does not know the same and cannot achieve the purpose of the contract on account thereof, the buyer may cancel the contract. In such cases, if the contract cannot be canceled, the buyer may only demand compensation for damages. (2)The provisions of the preceding paragraph shall apply mutatis mutandis in cases where an easement that was referred to as being in existence for the benefit of immovable property that is the subject matter of a sale, does not exist, and in cases where a leasehold is registered with respect to the immovable property.(3)In the cases set forth in the preceding two paragraphs, the cancellation of the contract or claim for damages must be made within one year from the time when the buyer comes to know the facts. (Seller's Warranty in cases of Mortgage or Other Rights)Article 567(1)If the buyer loses his/her ownership of immovable property that is the object of a sale because of the exercise of an existing statutory lien or mortgage, the buyer may cancel the contract.(2)If

the buyer preserves his/her ownership by incurring expenditure for costs, he/she may claim reimbursement of those costs from the seller.(3)In the cases set forth in the preceding two paragraphs, the buyer may claim compensation if he/she suffered loss.
$< /t1 >$
$< t2 >$
There is a limitation period on pursuance of warranty if there is restriction due to superficialities on the subject matter, but there is no restriction on pursuance of warranty if the seller's rights were revoked due to execution of the mortgage.
$< /t2 >$
$< /pair >$

The above is an example where query id **"H18-1-2"** is confirmed to be answerable from article numbers 566 and 567.

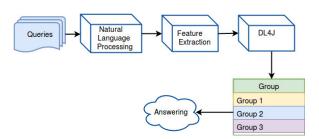## 4 Proposed Method

4.1 Proposed Model Overview

Figure 1: This is the proposed model overview

Figure 1 shows the proposed model overview. The initial idea is to try to cluster words in the training data to find the relation between them in order to retrieve relevant civil law articles.

4.2 Model Description

Each training query-article was represented as a feature vector by using Vector Representations of Words - word2vec model. These vectors were used to train neural network model - DL4J to map inputs to outputs. Deep learning is known as *universal approximator* because it can learn to approximate the function f(x) = y between any input x and any output y, assuming they are related through correlation or causation at all.

In the testing phase, given a query q, the model extracts its features and computes the relevance score corresponding to each article by using the

DL4J model. A higher score means the article is more relevant.

5 Result and Discussion

In our model, we focus on precision more than recall. For example, for each legal bar question the model only selects the one with the highest confidence score. Despite the fact that, there are more than one relevant document for the given question.

Noticed that one of the key properties of legal civil law articles is the reference and/or citation among documents. In other words, an article could refer to the whole of the other articles or to their paragraphs. If an article has a reference to other articles, the authors expanded it with words of referential ones. However, based on our system experiment, this approach confused the rank articles and eventually lead to a worse performance. Therefore, we ignored the reference and only took into account individual articles themselves. The results of the legal information retrieval task for data language using English is shown in the following table.

Team ID	Precision	Recall	F-measure
NOR17	0.462185	0.500000	0.480349
UA-LM	0.602564	0.427273	0.500000
Our model	0.430556	0.281818	0.340659

Table 1: COLIEE-2017 Evaluation Results

In conclusion, the main focus of our project is on precision not recall. Therefore, we approached this result. Compared to other work, this is not the best performance but the project has a very high potential and we strongly believe that, in the near future we will achieve a better approach by increasing the size of our training data, set such as utilizing dictionary of criminal law terminology, eurovoc to name a few.

Acknowledgements

References

Cabrio, E., Cojan, J., Aprosio, A.P., Magnini, B., Lavelli, A. and Gandon, F., 2012. November. QAKiS: an open domain QA system based on relational patterns. In Proceedings of the 2012th International Conference on Posters & Demonstrations Track-Volume 914 (pp. 9-12). CEUR-WS. org. Vancouver.

Luisa Bentivogli, Peter Clark, Ido Dagan, Hoa Dang, and Danilo Giampiccolo, 2011. The seventh pascal recognizing textual entailment challenge, volume 9 (pp. 9-12). Proceedings of TAC.

Leif Azzopardi, Mark Girolami, and CJ Van Rijsbergen, 2008. Concept and Context in Legal Information Retrieval volume-10 (pp. 3281-3286). JURIX

Leif Azzopardi, Mark Girolami, and CJ Van Rijsbergen, 2004. Topic based language models for ad hoc information retrieval. In Neural Networks, volume-4., Proceedings 2004 IEEE International Joint Conference on IEEE