Neural Attention and Morphological Word Embedding for Contract Element Extraction

Jose Luis Velasquez Sosa and Gerasimos Spanakis

Maastricht University, Maastricht, 6200MD, Netherlands
j.velasquezsosa@student.maastrichtuniversity.nl,
jerry.spanakis@maastrichtuniversity.nl

1 Introduction

As organizations try to automate their internal and external processes as much as possible, the need to keep track of legal obligations, rights and limitations has become an important task. In this paper neural attention mechanisms and morphological embedding are proposed to enhance the BiLSTM-CRF and BiLSTM-LSTM-LR architecture to explore the effects of neural attention mechanism and the morphologically-aware word embeddings.

The dataset used to evaluate this task consists of 2461 annotated and encrypted contracts with 11 different types of element extraction[1]. We focus on extractors of five of the elements: Parties (CNP), Start Date (STD), Effective Date (EFD), Termination Date (TED) and Governing Law (GOV).

2 Proposed Extractors

Each extractor is in itself a stacked deep learning architecture using a combination of simpler architectures. Each extractor extracts one contract element.

BiLSTM-LSTM-LR (bll) consists of a BiLSTM layer with an LSTM stacked on top of it. On top of the LSTM there is a fully connected layer with one output. **BiLSTM-CRF (bc)** consists of a BiLSTM network with a CRF, modeled on an RNN, on top of it. The output and evaluation are presented in the same format as the one from the extractor above. However, these probabilities are com- puted using the CRF layer. The biggest difference in this architecture is that all modules in the network are recurrent, which may prove to be an advantage over the BiLSTM-LSTM-LR. The BiLSTM-CRF model has been shown to perform better over all classes in the con- tract element extraction dataset [2]. The **Att-BiLSTM-CRF (abc)** extractor is the same as the BiLSTM-CRF extractor, however, an attention mechanism is used in order to weight the output of the BiLSTM before being used by the CRF layer to compute the probabilities.

Morphological embedding acts as a prefix architecture that can be added to all previous models. It consists of filtering the two words before and after along the one being classified to create a new embedded vector, using one vector for

[1] http://nlp.cs.aueb.gr/software_and_datasets/CONTRACTS_ICAIL2017/

every feature. Therefore, it should be possible to train all the extractors above with a special embedding layer (namely **Conv-BiLSTM-LR (cbll), Conv-BiLSTM-CRF (cbc) and Conv-Att-BiLSTM-CRF (cabc)**). The use of convolutions in textual sequence tagging has shown to increase the performance of models without a convolutional layer [1].

3 Experimental Results and Discussion

In Figure 1 the means of F-scores of each extractor for all elements can be seen. There seems to be a slightly better performance from the extractors with morphological embedding (confirmed by a paired t-test) and there seems to be no much of an effect with the attention mechanism.

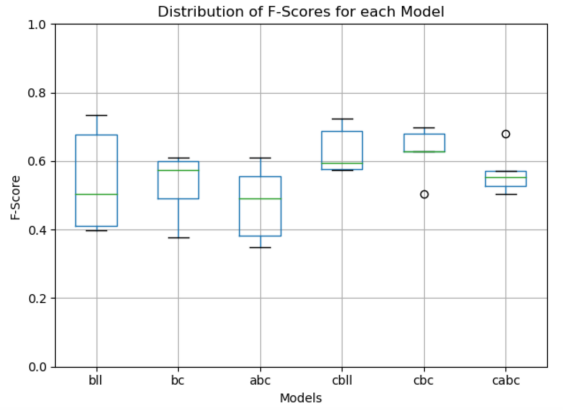

Fig. 1: Mean F-score per extractor

One of the limitations of this work is that the extractors trained in this study under-performed compared to the ones in previous work. This might be explained by the lack of shape information in the input data which could affect training, since for elements like Parties and Governmental Law, capitalization of certain characters might indicate their belonging in the class or not.

In order to fully explore the effects of attention mechanisms and morphological embedding in the BiLSTM-CRF extractor it would be necessary to replicate this experiment with different hyper-parameters and the addition of word-shape information, as well as extened the experimentation on a bigger dataset.

References

1. Ma, X., Hovy, E.: End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint arXiv:1603.01354 (2016)
2. Wyner, A., Casini, G.: A deep learning approach to contract element extraction. Legal Knowledge and Information Systems p. 155 (2017)