Learning Local and Global Contexts using a Convolutional Recurrent Network Model for Relation Classification in Biomedical Text

Desh Raj and **Sunil Kumar Sahu** and **Ashish Anand**

Department of Computer Science and Engineering, Indian Institute of Technology Guwahati

Objectives

Our objectives in this work are as follows:

- To extract structured knowledge from biomedical data such as journal articles, discharge summaries, and electronic health records.
- To learn a joint sentence and entity type embedding that represents local as well as global contexts, using a combination of CNNs and RNNs. We call this model a **CRNN**.
- To evaluate an attentive pooling strategy for this purpose.

Motivation

- The regression layer must see a **complete representation** of the sentence, i.e., both short and long-term dependencies must be represented in the sentence embedding.
- CNNs are capable of learning **local features** such as short phrases or recurring n-grams.
- RNNs utilize the word order in the sentence, and are also able to learn the **long-term dependencies**.

Datasets

We experimented with 2 data sets:

1. **i2b2-2010 relation extraction**: This dataset contains manually annotated discharge summaries collected from three different hospitals, for identifying problems, treatments and test entities, and 8 relation types among them.
2. **SemEval 2013 DDI extraction**: 4 kinds of interactions:
   - *Advice*: opinion related to the simultaneous use of the two drugs
   - *Effect*: effect of the DDI together with pharmacodynamic effect or mechanism of interaction
   - *Mechanism*: pharmacokinetic mechanism
   - *Int*: drug interaction without any other information

CRNN-Max

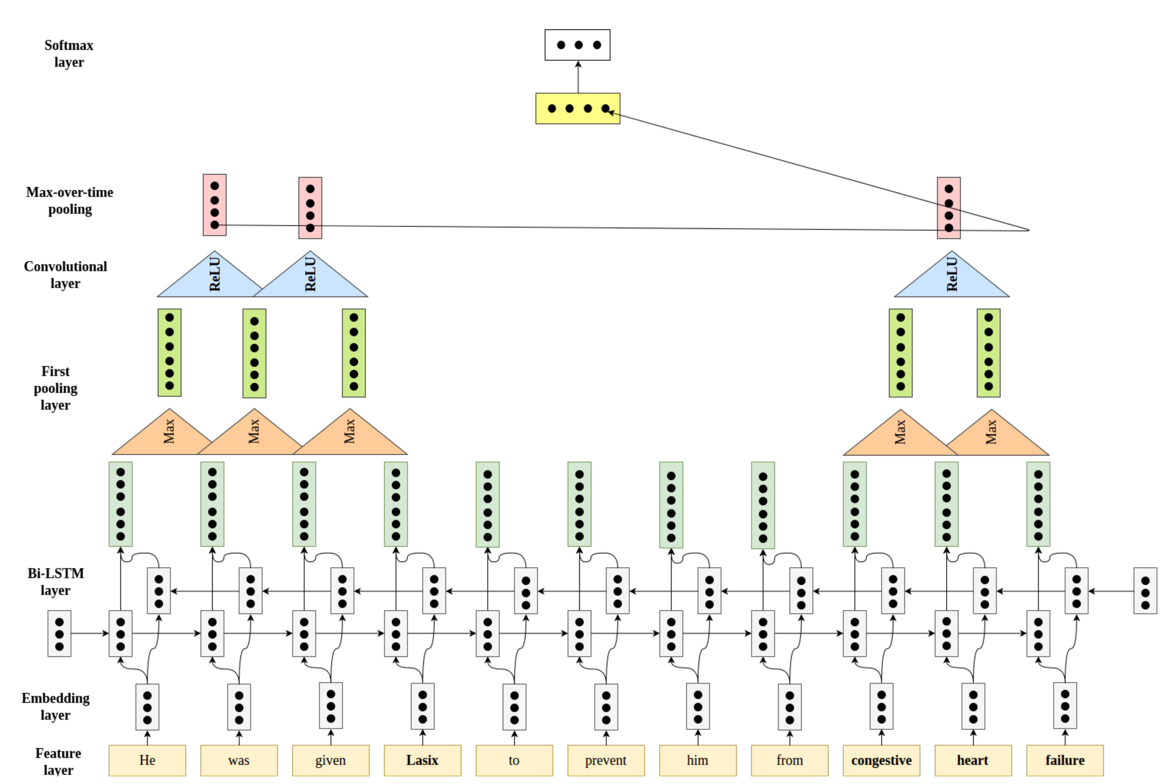

Figure : CRNN-Max architecture.

CRNN-Att

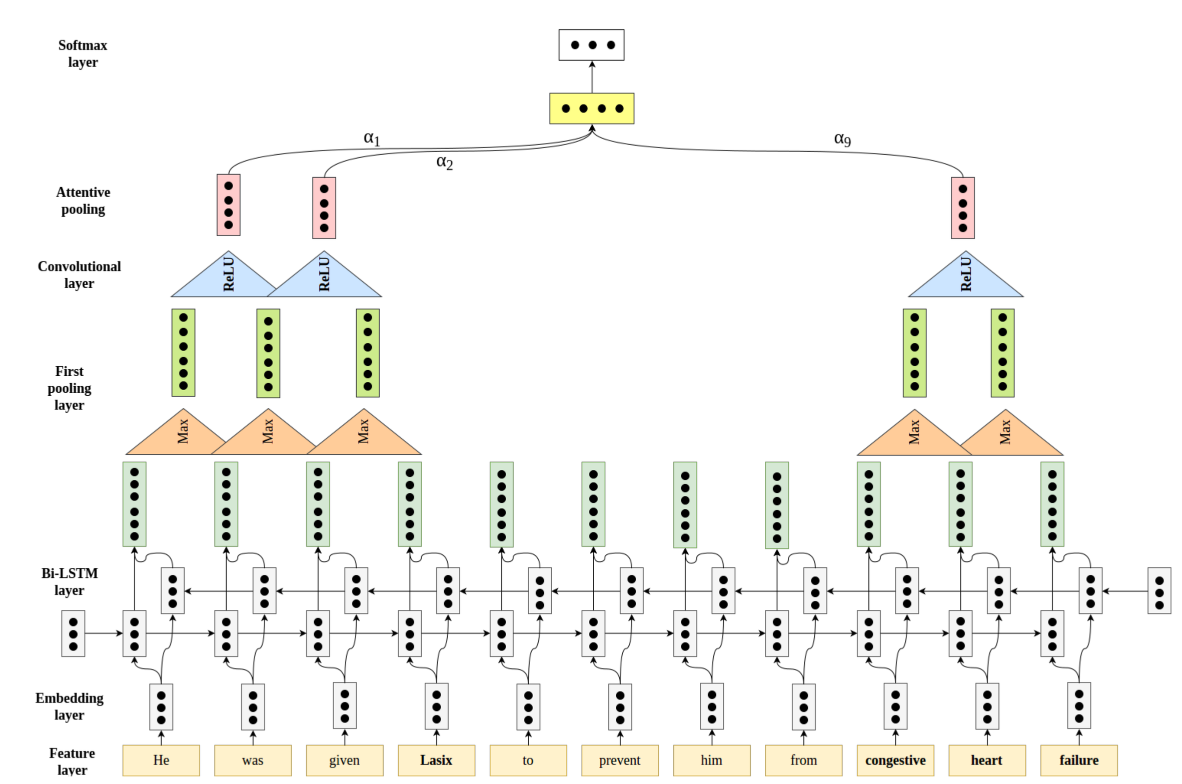

Figure : CRNN-Att architecture.

Important Result

The 2-layer CRNN model performs better because it learns both long and short term dependencies more efficiently. Furthermore, a "recurrent+pooling" layer learns regional embeddings from a sentence.

Experimental Results

The CRNN-Max model outperforms other models on both datasets. A small filter size for first pooling layer and medium size for second is found to perform best.

Model	i2b2-2010			DDI extraction		
	Precision	Recall	F1 score	Precision	Recall	F1 score
SVM	**73.20**	61.17	64.29	65.39	40.13	49.74
CNN-Max	55.73	50.08	49.42	68.15	46.58	54.05
LSTM-Max	57.54	55.40	55.60	**73.98**	59.96	65.41
LSTM-Att	65.23	56.77	60.04	53.43	**64.86**	58.27
RCNN	50.07	45.34	46.47	–	–	–
CRNN-Max	67.91	61.98	**64.38**	72.91	60.88	**65.89**
CRNN-Att	64.62	**62.14**	62.45	69.03	59.04	63.24

Table : Comparison of our proposed models CRNN-Max and CRNN-Att, with baselines, on the i2b2-2010 and DDI extraction datasets.

$f_1 \backslash f_2$	2	3	4	5	6
1	59.97	58.96	59.30	59.18	60.03
2	59.84	56.69	60.89	**62.45**	61.03
3	60.46	61.77	58.85	57.34	59.81

Table : Average F1 scores on varying filter sizes f_1 and f_2 in the CRNN-Att model for i2b2 dataset.

Why CRNN performs well?

1. LSTM layer learns **regional embeddings** efficiently. This result has also been seen in [1].
2. Together, the two layers correctly classify a larger range of sentence lengths (see Figure).
3. Similar distribution is obtained on considering entity separation instead of sentence length.

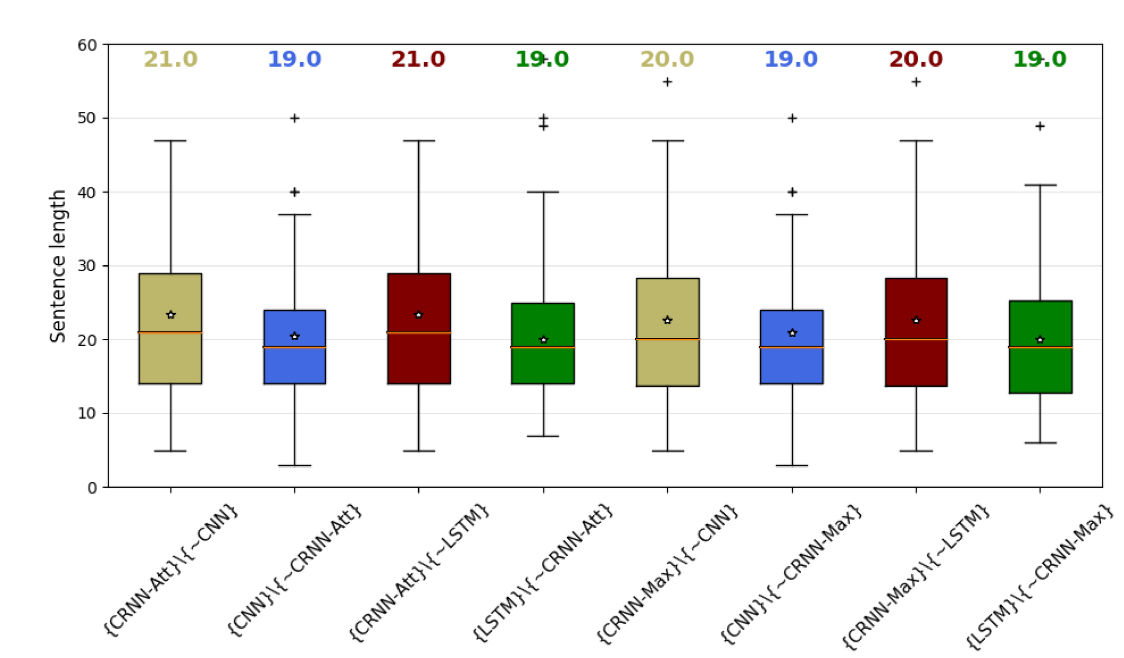

Figure : Distribution of sentence lengths for various sentence sets.

Effect of linguistic features and attention

1. Random initialization of word embeddings performs as well as PubMed initialization, similar to observations made by Johnson et al. [1].
2. An SVM trained on linguistic features has high specificity but low recall. In comparison, our CRNN model **generalizes better**.
3. **Attentive pooling** scheme is able to select important phrases depending upon the classification label.

Conclusion

- Our proposed CRNN-Max model outperforms existing methods on both datasets, since it learns a better sentence representation.
- An attentive pooling strategy effectively learns to weigh important words higher than common words.
- It may be interesting to see whether tree-based or non-continuous convolutions work well for learning dependencies.

References

[1] Rie Johnson and Tong Zhang.
Supervised and semi-supervised text categorization using lstm for region embeddings.
arXiv preprint arXiv:1602.02373, 2016.

[2] Sunil Kumar Sahu and Ashish Anand.
Drug-drug interaction extraction from biomedical text using long short term memory network.
arXiv preprint arXiv:1701.08303, 2017.

Contact Information

- Web: http://www.rdesh26.wixsite.com/home
- Email: r.desh@iitg.ac.in