Title

Anonymous EMNLP submission

Abstract

We show that syntactic information is important when performing natural language inference (NLI). We describe new modeling techniques for NLI that improve upon previous baselines by adding minimal amounts of syntactic information.

1 Introduction

Natural language inference (NLI) is the task of characterizing entailment and contradiction relationships between texts. In general, most NLI tasks are formulated as characterizing the relationship between a pair of sequences—a premise and a hypothesis. An NLI model should predict whether the hyothesis is entailed by the premise, contradicts the premise, or is neutral to the premise.

We seek to improve models which perform NLI by adding syntactic features into existing models. In particular, we investigate the well-known decomposable attention (DA) neural network (Parikh et al., 2016) which obtained state of the art results on the SNLI (Bowman et al., 2015) dataset at its time of publication while using drastically fewer parameters than previous NLI models.

We extend the DA model and evaluate new models on SNLI, MultiNLI (Williams et al., 2017), the SNLI and MultiNLI hard splits (Gururangan et al., 2018), and smaller, domain specific datasets like Scitail (Khot et al., 2018).

2 Previous Work

Our project will build on previous work done by Pang et al. (2018).

In previous work, we created the syntail-v1 model whose architecture is shown in figure 1. We have obtained mixed results as shown in table 1 when naively incorporating syntactic information into the DA model.

Without ELMo, syntail-v1 improves an impressive 8.4% in test accuracy over the baseline decomposable attention model on the SciTail dataset. However, once ELMo is added to the model, almost the exact opposite occurs. Adding syntactic information decreases test accuracy by 6.7% with ELMo.

Model	Embedding	Train Acc.	Test Acc.
DA	GLoVe 6b 300d	89.4%	70.1%
Syntail-v1	**GLoVe 6b 300d**	**92.1%**	**78.5%**
DA	ELMo	98.4%	78.0%
Syntail-v1	**ELMo**	**88.8%**	**71.3%**

Table 1: Syntail-v1 and DA model accuracies on the SciTail dataset.

3 Proposed Methods

We plan to use state-of-the-art constituency (Stern et al., 2017) and dependency (Dozat and Manning, 2016) parsers to investigate if incorporating syntactic features improves the performance of the DA model. Additionally, we hope to understand how syntactic features interact with contextual word embeddings like ELMo (Peters et al., 2018) and BERT (Devlin et al., 2018) in a task like NLI. It is unclear from previous work whether contextual word embeddings benefit from additional syntactic information or if something like ELMo can already adequately capture this information.

We will first investigate different ways of incorporating syntactic information. Pang et al. (2018) use a constituency parser to obtain an encoding of the premise and hypothesis that is in the final fully-connected layer of the DA model. We believe that incorporating syntactic features earlier in the pipeline, especially before the premise and hypothesis are attended over one another, should improve performance.

We will then investigate different model architectures under a multi-task training objective where the jointly learns to both parse and predict entailment relationships.

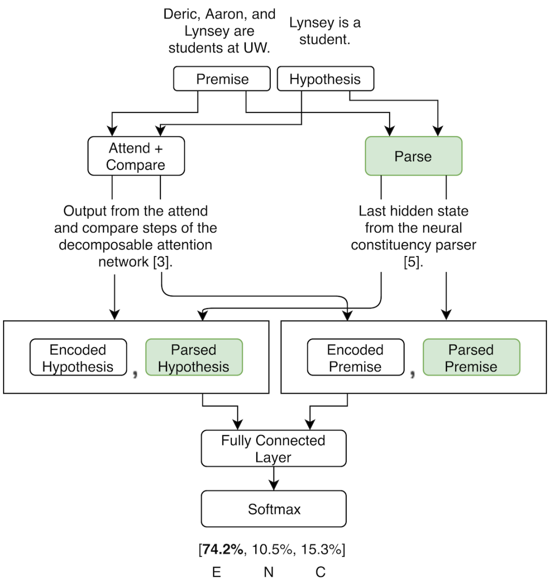

Figure 1: A simple method of incorporating syntax into the decomposable attention model.

4 Expected Outcomes

We will compare our models to the DA model as well as syntail-v1. We hope to evaluate our models on many diverse datasets as stated in section 3.

We expect that one of our modeling ideas will produce results that improve on the DA and syntail-v1 models. In the best case, we will show that syntax is crucial to building an accurate NLI system.

5 Challenges

The first challenge involved in our project will be to modify syntail-v1 so that syntactic information is incorporated much earlier. It will also take significant time to investigate the best parsers to use for our model. After, we will write a new model under a multi-task training setup, which we anticipate will be a significant implementation challenge.

We expect to require significant compute power and time to adequately fine-tune and evaluate our models. Contextual word embeddings are especially costly to train, and the mixing ratio in a multi-task setup is well-known to be difficult to tune.

6 Credits

References

Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding.

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing.

Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel R Bowman, and Noah A Smith. 2018. Annotation artifacts in natural language inference data.

Tushar Khot, Ashish Sabharwal, and Peter Clark. 2018. SciTail: A textual entailment dataset from science question answering. In *Proceedings of AAAI*.

Deric Pang, Lynsey Liu, and Aaron Johnston. 2018. Syntail: Syntactically-informed natural language inference. CSE 599 G1 Deep Learning.

Ankur P Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. 2016. A decomposable attention model for natural language inference.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv:1802.05365 [cs]*.

Mitchell Stern, Jacob Andreas, and Dan Klein. 2017. A minimal Span-Based neural constituency parser.

Adina Williams, Nikita Nangia, and Samuel R Bowman. 2017. A Broad-Coverage challenge corpus for sentence understanding through inference.