Natural Language Inference using External Knowledge

Pavan Kapanipathi
{kapanipa}@us.ibm.com

Abstract

Natural language inference (NLI) – sometimes referred to as textual entailment – is a fundamental task in natural language processing. Most approaches for solving this problem are driven by the natural language text provided for training. However, external knowledge sources such as ConceptNet, DBpedia can add value by enhancing the semantics of text for this task. In this talk, I will focus on techniques that can leverage external knowledge for Natural Language Inference. Particularly, the talk will cover an initial framework that can integrate both text based and external knowledge based models, with emphasis on different ways of effectively exploit external knowledge. We evaluate our approach on multiple textual entailment datasets and show that the use of external knowledge helps the model to be robust and improves prediction accuracy

1 Introduction

Given two natural language sentences, a premise P and a hypothesis H, the textual entailment task – also known as natural language inference (NLI) – consists of determining whether the premise entails, contradicts, or is neutral with respect to the given hypothesis [MacCartney and Manning, 2009]. In practice, this means that textual entailment is characterized as either a three-class (ENTAILS/NEUTRAL/CONTRADICTS) or a two-class (ENTAILS/NEUTRAL) classification problem [Khot *et al.*, 2018b; Bowman *et al.*, 2015].

Performance on the textual entailment task can be an indicator of whether a system, and the models it uses, are able to reason over text. This has tremendous value for modeling the complexities of human-level natural language understanding, and in aiding systems tuned for downstream tasks such as question answering [Harabagiu and Hickl, 2006].

Most existing textual entailment models focus only on the text of the two sentences to improve classification accuracy [Parikh *et al.*, 2016; Zhang *et al.*, 2018; Liu *et al.*, 2019]. A recent and promising line of work has turned towards extracting and harnessing more contextually relevant semantic information from knowledge graphs (KGs) for each textual entailment pair [Wang *et al.*, 2019; Kapanipathi *et al.*, 2020].

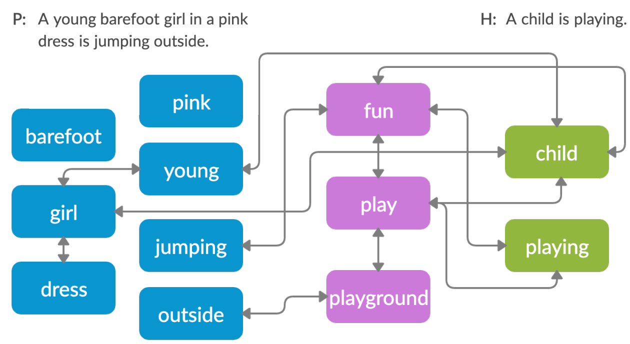

Figure 1: A premise and hypothesis pair along with a relevant subgraph from ConceptNet. Blue concepts occur in the premise, green in the hypothesis, and purple connect them.

These approaches map terms in the premise and hypothesis text to concepts in a KG, such as Wordnet [Miller, 1995] or ConceptNet [Speer *et al.*, 2017] and use information of these mapped concepts for the textual entailment task. Figure 1 shows an example of such mapping, where select terms from the premise and hypothesis are mapped to concepts from a knowledge graph (blue and green nodes, respectively). This talk will focus on these two approaches that can be categorized into a generic framework where text-based models are augmented with external knowledge sources for better performance on NLI task.

These two approaches aim to address the following challenges in using external knowledge bases for NLI: (a) determining the relevant external knowledge source to use; (b) extracting relevant information from large and noisy KGs; (c) effectively leveraging both the semantic and structural information from KGs.

2 Framework and Approaches

In this section, we describe the central contribution of this paper – the KG-augmented Entailment System (KES). As shown in Figure 2, KES consists of two main components. The first component is a standard text encoder that creates a fixed-size representation of the premise and hypothesis texts. The second component selects contextual subgraphs for the premise and the hypothesis from a given KG, and encodes them using multiple different techniques as detailed in below subsections. The final layers of the two components are used as input to a standard feedforward layer for classification. We opted for a combined graph and text approach because the noise and incompleteness of KGs renders a purely graph-based approach insufficient as a standalone solution. However, we show that the KG-augmented model provides valuable context and additional knowledge that may be missing in text-only representations.

2.1 ConSeqNet

Challenges: (1) Many external knowledge sources are available and choosing one that is appropriate for a given NLI dataset is non-trivial; (2) a general framework for augmenting text-based models with external knowledge is needed, as existing NLI approaches that use external knowledge are tightly tuned to one specific KG.

Contributions: The ConSeqNet framework enables the use of various kinds of external knowledge bases to retrieve knowledge relevant to a given NLI instance, by retrieving information related to the premise and hypothesis. We describe our novel architecture and demonstrate its use with a specific external knowledge source – ConceptNet – and evaluate its performance on two other sources, WordNet and DBpedia. We compare the performance of three distinct approaches to augmenting the knowledge used to train for and to predict entailment relationships between given pairs of premises and hypotheses:graph-only, text-only, and text-and-graph. Using both qualitative and quantitative results, we demonstrate that introducing graph-based features boosts performance on the NLI problem, but only when text features are present as well. Our system has a competitive performance (accuracy) of 85.2 (Table 1).

2.2 KG augmented Entailment System (KES)

Challenges: (1) ConSeqNet and existing KG-based models do not possess the ability to select and harness semantic and structural information from the KG. For example, in Figure 1, the ability for models to encode information from paths between blue and green nodes via purple nodes provides better context facilitating the system to more correctly judge entailment. (2) They are not easily integrated with existing NLI models that exploit only the text of the premise and hypothesis. (3) They are not flexible with respect to the type of KG that is used.

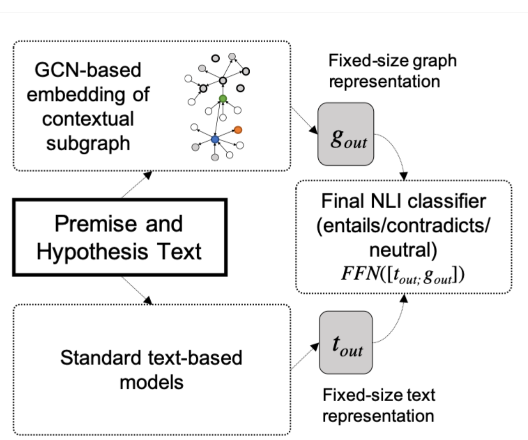

Figure 2: Primary components of KES: standard text-based model, GCN-based graph embedder, and final feedforward classifier.

Model	Dev	Test
Decomp-Attn [Parikh *et al.*, 2016]	75.4	72.3
DGEM* [Khot *et al.*, 2018a]	79.6	77.3
DeIsTe [Yin *et al.*, 2018]	82.4	82.1
BiLSTM-Maxout [Mihaylov *et al.*, 2018]	-	84.0
match-LSTM [Wang and Jiang, 2015]	88.2	84.1
Our implementation		
match-LSTM (GRU)	88.5	84.2
match-LSTM+WordNet* [**?**]	88.8	84.3
match-LSTM+Gmatch-LSTM* ()	**89.6**	**85.2**

Table 1: Performance of entailment models on SciTail in comparison to our best model that uses match-LSTM as the text and the graph model with *Concepts Only* graph and CN-PPMI embeddings. * indicates the use of external knowledge in the approach.

Contributions: We present an approach to the NLI problem that can augment any existing text-based entailment model with external knowledge. Our approach has two major innovations. First, we introduce a neighbor-based expansion strategy in combination with subgraph filtering using Personalized PageRank (PPR) [Jeh and Widom, 2003]. This approach reduces noise and selects contextually relevant subgraphs for premise and hypothesis texts from larger external knowledge source. Second, we encode subgraphs using Graph Convolutional Networks (GCNs) [Kipf and Welling, 2017], which are initialized with knowledge graph embeddings to capture structural and semantic information. This general approach to graph encoding allows us to use any external knowledge source that can be represented as a graph such as WordNet, ConceptNet, or DBpedia [Lehmann *et al.*, 2015]. We show that the additional knowledge can improve textual entailment performance by using four standard benchmarks: SciTail, SNLI, MultiNLI, and BreakingNLI. In particular, our experiments on the BreakingNLI dataset [Glockner *et al.*, 2018], where we see an absolute improvement of 5-20% over four text-based models, shows that our technique is robust and resilient (Table 2).

References

[Bowman *et al.*, 2015] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In *Proc. of EMNLP*. Association for Computational Linguistics, 2015.

[Chen *et al.*, 2018] Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Diana Inkpen, and Si Wei. Neural natural language inference models enhanced with external knowledge. In *Proc. of ACL, Volume 1*, pages 2406–2417, 2018.

Models	Scitail		MultiNLI		SNLI		BreakingNLI	
	Text	**KES**	**Text**	**KES**	**Text**	**KES**	**Text**	**KES**
match-LSTM	82.54	82.22 (0.6)	71.32	71.67 (0.8)	83.60	83.94 (0.6)	65.11	**78.72**
BERT+match-LSTM	89.13	**90.68** (0.2)	77.96	76.73 (0.6)	85.78	85.97 (0.6)	59.42	**77.59**
HBMP	81.37	**83.49** (0.2)	69.27	68.42 (0.6)	84.61	83.84 (0.2)	60.31	**63.60**
DecompAttn	76.57	72.43 (0.8)	64.89	**71.93** (0.6)	79.28	**85.56** (0.6)	51.3*	**59.83**
KIM [Chen *et al.*, 2018]	-	**NE**	-	76.4*	-	88.6*	-	83.1*
ConSeqNet [Wang *et al.*, 2019]	84.2*	85.2*	71.32	70.9	83.60	83.34	65.11	61.12

Table 2: Entailment accuracy results of KES with different text models and text-only versions (Text). Bold values indicate where KES improves performance. PPR θ-values are shown in parentheses. *Reported values from related work.

[Glockner *et al.*, 2018] Max Glockner, Vered Shwartz, and Yoav Goldberg. Breaking nli systems with sentences that require simple lexical inferences. In *Proc. of ACL, Volume 2*, pages 650–655, 2018.

[Harabagiu and Hickl, 2006] Sanda Harabagiu and Andrew Hickl. Methods for using textual entailment in open-domain question answering. In *Proc. of CICLing and ACL*, pages 905–912. Association for Computational Linguistics, 2006.

[Jeh and Widom, 2003] Glen Jeh and Jennifer Widom. Scaling personalized web search. In *Proc. of WWW*, pages 271–279. Acm, 2003.

[Kapanipathi *et al.*, 2020] Pavan Kapanipathi, Veronika Thost, Siva Sankalp Patel, Spencer Whitehead, Ibrahim Abdelaziz, Avinash Balakrishnan, Maria Chang, Kshitij Fadnis, Chulaka Gunasekara, Bassem Makni, et al. Infusing knowledge into the textual entailment task using graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8074–8081, 2020.

[Khot *et al.*, 2018a] T. Khot, A. Sabharwal, and P. Clark. SciTail: A textual entailment dataset from science question answering. In *32nd*, 2018.

[Khot *et al.*, 2018b] Tushar Khot, Ashish Sabharwal, and Peter Clark. SciTail: A textual entailment dataset from science question answering. In *Proc. of AAAI*, 2018.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proc. of ICLR*, 2017.

[Lehmann *et al.*, 2015] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195, 2015.

[Liu *et al.*, 2019] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. In *ACL (1)*, pages 4487–4496. Association for Computational Linguistics, 2019.

[MacCartney and Manning, 2009] Bill MacCartney and Christopher D Manning. *Natural language inference*. Stanford University Stanford, 2009.

[Mihaylov *et al.*, 2018] T. Mihaylov, P. Clark, T. Khot, and A. Sabharwal. Can a suit of armor conduct electricity? A new dataset for open book question answerin. In *2018*, 2018.

[Miller, 1995] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[Parikh *et al.*, 2016] Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. In *Proc. of EMNLP*, pages 2249–2255, 2016.

[Speer *et al.*, 2017] Robert Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proc. of AAAI*, 2017.

[Wang and Jiang, 2015] Shuohang Wang and Jing Jiang. Learning natural language inference with lstm. *arXiv preprint arXiv:1512.08849*, 2015.

[Wang *et al.*, 2019] Xiaoyan Wang, Pavan Kapanipathi, Ryan Musa, Mo Yu, Kartik Talamadupula, Ibrahim Abdelaziz, Maria Chang, Achille Fokoue, Bassem Makni, Nicholas Mattei, and Michael Witbrock. Improving Natural Language Inference Using External Knowledge in the Science Questions Domain. *Proc. of AAAI*, 2019.

[Yin *et al.*, 2018] Wenpeng Yin, Dan Roth, and Hinrich Schütze. End-task oriented textual entailment via deep explorations of inter-sentence interactions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 540–545, 2018.

[Zhang *et al.*, 2018] Zhuosheng Zhang, Yuwei Wu, Zuchao Li, Shexia He, Hai Zhao, Xi Zhou, and Xiang Zhou. I know what you want: Semantic learning for text comprehension. *arXiv preprint arXiv:1809.02794*, 2018.