Multilingual Deep Neural Networks for Paraphrase Aquisition

Nishant Kambhatla, Akash A J

October 30, 2016

1 Introduction

Paraphrases are alternative ways to convey the same information. A method for the automatic acquisition of paraphrases has both practical and linguistic interest. Taking a leap from the conventional statistical approaches for paraphrasing, the recent times have seen quite a few Neuaral Computational approaches to tackle the problem. Interestingly, most of such works have addressed Paraphrase Identification - a classification task - or Paraphrase Generation - an abstractive task. The aim of this project, however, is to structure a framework for Paraphrase Aquisition - an extractive task - in a multilingual setting.

Consider the following example from (Barzilay & McKeown, 2001). It contains two pairs of paraphrases: ("burst into tears", "cried") and ("comfort","console").

Emma burst into tears and he tried to comfort her, sayingthings to make her smile.
Emma cried, and he tried to console her, adorning hiswords with puns.

Datasets:
Paraphrase: PPDB, PPDB 2.0, MSRP, MSCOCO
Parallel Corpora: WMT16, EuroParl

2 Methodologies of Interest

1. **Stacked Residual LSTM Networks** (K. He, Zhang, Ren, & Sun, 2015):
   ResNet is deep learning network with added residue for the purpose of learning. In ResNet, the explicit addition of the residue x to the function being learned allows for deeper network training without overfitting the data. The addition of residual connection does not add any learnable parameters. Therefore, this does not increase the complexity of the model unlike bi-directional models which double the number of LSTM units.

2. **Dynamic Pooling and Recursive Auto-Encoders** (Socher, Huang, Pennin, Manning, & Ng, 2011):
   Recursive autoencoder is learned to represent words/phrases of a sentence, from

a training corpus where sytactic parse trees are given for each sentence. Reconstruction error of encoded vectors is minimized to learn the parameters of the autoencoder. Vector representation of variable sized sentence pairs gives a similarity score (diffrent size for different sentence pairs), which is then converted to a fixed length embedding using a dynamic pooling algo.

3. **Multi-Perspective Sentence Similarity Modeling with CNN**(H. He, Gimpel, & Lin, n.d.):
   A model for comparing sentences that uses a multiplicity of perspectives. This model has 2 key components -

   (a) **A sentence model** for converting a sentence into a representation for similarity measurement; a CNN architecture with multiple types of convolution and pooling is udsed in order to capture different granularities of information in the inputs.

   (b) **A similarity measurement layer** using multiple similarity measurements, which compare local regions of the sentence representations from the sentence model.

4. **Other motivating articles:**
   (Callison-Burch, Koehn, & Osborne, 2006) (Marton, Callison-Burch, & Resnik, 2009) (Yin & Schütze, 2016) (Cheng & Kartsaklis, 2015) (Bahdanau, Cho, & Bengio, 2014) (Bahdanau et al., 2014) (Luong, Pham, & Manning, 2015)

References

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Barzilay, R., & McKeown, K. R. (2001). Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th annual meeting on association for computational linguistics* (pp. 50–57).

Callison-Burch, C., Koehn, P., & Osborne, M. (2006). Improved statistical machine translation using paraphrases. In *Proceedings of the main conference on human language technology conference of the north american chapter of the association of computational linguistics* (pp. 17–24).

Cheng, J., & Kartsaklis, D. (2015). Syntax-aware multi-sense word embeddings for deep compositional models of meaning. *arXiv preprint arXiv:1508.02354*.

He, H., Gimpel, K., & Lin, J. (n.d.). Multi-perspective sentence similarity modeling with convolutional neural networks..

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

Marton, Y., Callison-Burch, C., & Resnik, P. (2009). Improved statistical machine translation using monolingually-derived paraphrases. In *Proceedings of the 2009 conference on empirical methods in natural language processing: Volume 1-volume 1* (pp. 381–390).

Socher, R., Huang, E. H., Pennin, J., Manning, C. D., & Ng, A. Y. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Advances in neural information processing systems* (pp. 801–809).

Yin, W., & Schütze, H. (2016). Discriminative phrase embedding for paraphrase identification. *arXiv preprint arXiv:1604.00503*.