Neural Machine Translation for Paraphrase Extraction

Nishant Kambhatla, Akash Abdu Jyothi

Abstract: Paraphrases are alternate ways to convey the same information. We propose a method to automatically extract paraphrases from monolingual corpus, by using an NMT system trained to translate from our source language to a foreign language. Similar to [1, 2], we use pivoting method to extract paraphrases of the source language using translated "parallel" corpus. We find identical words/phrases in the target side of NMT and map them back to the source side to extract paraphrases. We use paraphrase scoring function similar to [1, 2], with an additional confidence score from NMT alignment model [3]. We use nematus [4] as a base to perform translations.

[1] Chris Callison-Burch, Colin Bannard, and Josh Schroeder. 2005. Scaling phrase-based statistical machine translation to larger corpora and longer phrases. In Proceedings of ACL.

[2] 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.

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

[4] Rico Sennrich; Barry Haddow; Alexandra Birch (2016). Edinburgh Neural Machine Translation Systems for WMT 16. Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers. Berlin, Germany, pp. 368-373