

Chaoran Lin

CS 35L

TA: Mushi Zhou

1 December 2016

Deep Learning in Machine Translation

Deep learning is an area in machine learning that aims to help computers achieve a high level of artificial intelligence by mimicking the process of human learning and knowledge acquisition [1]. Through connected layers of computational units in an artificial neural network, combined with a large set of input data, deep learning systems present a new, more advanced way of tackling problems in computer science and artificial intelligence (AI). In recent years, it has shown its advantage in image recognition and game-playing, most notably through the AlphaGo AI's victory over Go champion Lee Sedol in March 2016 [2].

However, applying deep learning to improve machine translation remains a challenge. To solve this challenge, Google has recently introduced a new algorithm based entirely on deep learning to improve the quality of machine translation as part of its new Google Neural Machine Translation system (GNMT) [3]. The difference is that past approaches used a Phrase-Based Machine Translation technique (PBMT), which breaks input sentences into words and phrases and translated them independently. To improve on this, the Neural Machine Translation system's neural network considers the entire input sentence as a unit for translation [4]. On a high-level overview, the translation will be more considerate and dependent on the overall semantic context of the input sentence in that the system constantly refers back to the original sentence during the translation, which is similar to the process and habits of human translation [5].

In translating a sentence, the GNMT system uses three tools: an encoder, a decoder, and a hidden layer in between which has an attention mechanism. First, it uses an encoder to encode the words as vectors representing the meaning of all the words read so far. After reading the entire sentence, it then uses a decoder that generates the translated sentence in the output language one word at a time. To do so, it pays different strengths of attention to a weighted distribution over the encoded vectors in the input language that are most relevant in generating a translation for the word. It then creates a decoded vector that encompasses the meaning of the vectors that have the most relevance, which it uses to generate the word in the output language. In the end, it not only uses the information encoded in the input language, but also considers the importance of each word in each portion of the sentence, creating a more coherent translation. Figure 1 provides a simple visualization of this process.

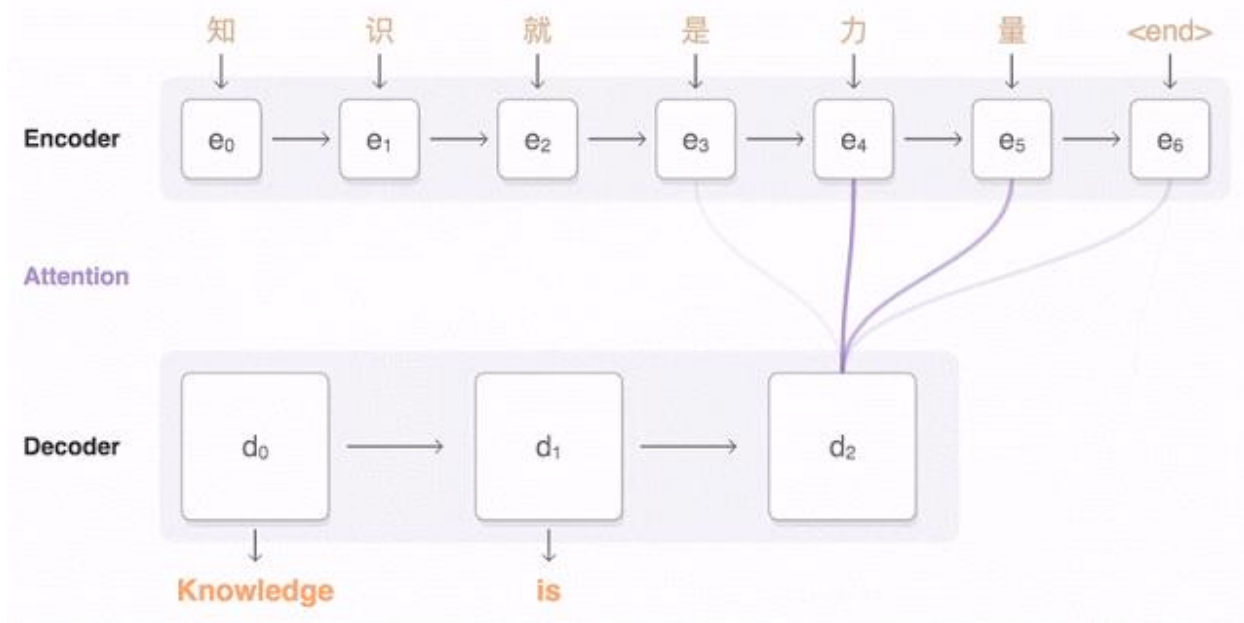


Fig. 1: GNMT's encoder and decoder translating the Chinese phrase "Knowledge is power" to English. The blue link's transparency signifies the amount of attention the decoder is paying to an encoded word. Source: [4]

Behind this whole machine translation model is the use of Google's machine learning toolkit TensorFlow and Tensor Processing Units (TPUs), "which provide sufficient sufficient

computational power to deploy these powerful GNMT models” [4]. Similar hardware was also used in supporting the AlphaGo AI’s legendary victory previously mentioned [3].

To examine the quality of Neural Machine Translation, Google researchers performed a test where they used the system translate sentences from Wikipedia and news articles between major language pairs (e.g. English-Spanish, English-Chinese) and compared it to translation results based on its previous Phrase-Based Machine Translation approach with the help of bilingual human raters. The outcome was that the Neural Machine Translation system reduced translation errors by 55%-85%, using the ratings of the bilingual humans as a metric [4].

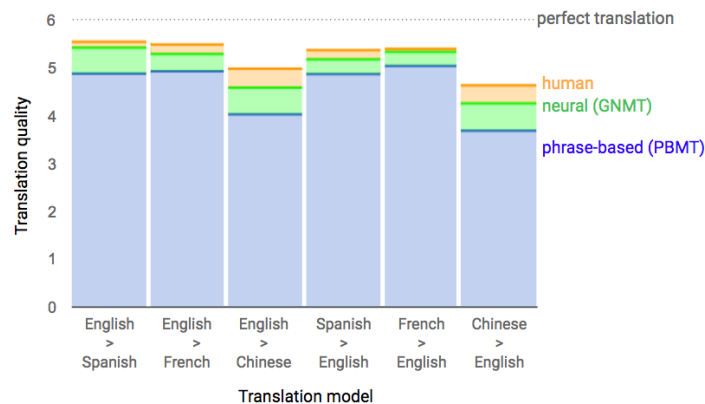


Fig. 2: Ratings of quality of translations between phrase-based machine translation (PBMT), neural machine translation (GNMT), and human translation. Source: [4]

However, there are still some problems and limitations with the Neural Machine Translation system. For example, it can still drop words and mistranslate proper names or rare terms [4]. It was also noted that the test translations were limited by their sample of rather well-crafted and simple sentences [3].

Despite these drawbacks, I believe that the introduction of deep learning to Google’s machine translation system still serves as a milestone in the progression of machine learning. In fact, Neural Machine Translation is already being used in Google’s very own Translate mobile and web apps, bringing its impact to a large base of users [4]. This means that the neural network in

the system can improve continuously as a huge amount of users feed it input data daily, due to the properties of deep learning [1].

On a broader scale, I believe that it can also create meaningful impact outside of translation services. For example, we can anticipate the use of this new technique in improving the quality of computer-assisted language learning software and tools such as Rosetta Stone and Duolingo. Most importantly, these research advancements can help further bridge the gap between the languages—and people—of the world, providing a more seamless translation and connection between users today.

Works Cited

- [1]: Jones, Nicola. "Computer Science: The Learning Machines." Nature.com. Macmillan Publishers, 08 Jan. 2014. Web. 01 Dec. 2016. <<http://www.nature.com/news/computer-science-the-learning-machines-1.14481>>
- [2]: Chouard, Tanguy. "The Go Files: AI Computer Wraps up 4-1 Victory against Human Champion." Nature.com. Macmillan Publishers, 15 Mar. 2016. Web. 30 Nov. 2016. <<http://www.nature.com/news/the-go-files-ai-computer-wraps-up-4-1-victory-against-human-champion-1.19575>>
- [3]: Castelvechi, Davide. "Deep learning boosts Google Translate tool." Nature.com. Macmillan Publishers, 27 Sep. 2016. <<http://www.nature.com/news/deep-learning-boosts-google-translate-tool-1.20696>>
- [4]: Le, Quoc V., and Mike Schuster. "A Neural Network for Machine Translation, at Production Scale." Google Research Blog. Google, Inc., 27 Sept. 2016. Web. 30 Nov. 2016. <<https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>>
- [5]: Hsu, Jeremy. "Google Translate Gets a Deep-Learning Upgrade." Google Translate Gets a Deep-Learning Upgrade. IEEE Spectrum, 03 Oct. 2016. Web. 1 Dec. 2016. <<http://spectrum.ieee.org/tech-talk/computing/software/google-translate-gets-a-deep-learning-upgrade>>