Arabic machine translation using Bidirectional LSTM Encoder-Decoder

BENSALAH Nouhaila*, AYAD Habib*, ADIB Abdellah* and IBN EL FAROUK Abdelhamid+

*Team Networks, Telecoms & Multimedia
LIM@II-FSTM, B.P. 146
Mohammedia 20650, Morocco
+Teaching, Languages and Cultures Laboratory Mohammedia,

bensalah.3.nouhaila@gmail.com, ayad.habib@gmail.com, adib@fstm.ac.ma, farouklettres@gmail.com

Abstract

Due to the language structure, applying the same machine translation approach may not work for Arabic language as for European languages. So, there is a great need to develop a model to solve this issue. Machine Translation (MT) using neural networks has recently become a viable alternative approach to the most widely-used statistical MT . Although a lot of research has been done on MT for Arabic language, to the best of our knowledge, no work has been used a Bidirectional Recurrent Neural Network (BiRNN) encoder/decoder for this task. In this poster, we aim to fulfill this goal by developping a model based mainly on Bidirectional Long Short-Term Memory (BiLSTM) with to map the input sequence to a vector, and we use then another Long Short-Term Memory (LSTM) to decode the target sequence from the obtained vector. Our work offers encouraging results in terms of correlation with human judgment.

Introduction

Automatic machine translation is considered to be the major problem in natural language processing. It has proved to be both the most attractive and the least accessible task. Since the introduction of MT , many approches have been applied, from traditional rule-based methods to the more recent statistical methods.

MT has been an active research topic since 1950s [1]. Originally, MT systems were developed using both dictionaries and rules to generate correct word order. In the 1990s, statistical methods became dominant [2] due to the availability of large corpora, comutational speed, and software for performing basic translation process such as alignement, recordering, filtering, etc.

The particular problem of MT has a long history as well. In 1982, a paper by Nagao [3] applies a rule-based machine translation between English and Japanese to transfer grammatical concepts between the two languages. Another phrase-based statistical machine translation sytems between English and Arabic have been proposed by [4] with an impressive improvement over other sytems without using any neural network. However, the authors state that the results on statistical machine translation achieve only a baseline level of success.

Recently, neural machine translation has been extremely powerful due to its exellent performance on difficult problems such as speech recognition [5] and visual object recognition [6] for a modest number of steps, and have been achieved close to state-of-the art accuracy in machine translation [2]. However, RNNs suffer from the vanishing and exploding gradient problem [7]. So, if we are trying to translate a paragraph of text, RNNs may leave out important information from the beginning.A common solution is to use either LSTM [8] or the Gated Recurrent Unit (GRU) [9] neural networks wich solve these problems and have proved to perform equally well at capturing long-term dependencies.

In this paper, our aim is therfore to present the first result on the Arabic translation using BiLSTM as encoder to map the input sequence to a vector, and a simple LSTM as a decoder to decode the target sentence from the obtained vector.

The outline of the paper is structured as follows. The next section details the proposed approach. The experiments and obtained results are presented in section . At the end of this paper, a conclusion is presented.

The proposed Approach

Most of the state-of-the art machine translation systems employ RNN's [[10], [11], [12], [13]]. These models often use an encoder-decoder approach to predict translations. In this section, we will explain in detail the architecture of the proposed model presented in the experiments results.

The architecture of LSTM

LSTM [14] was created to solve the problem of short-term memory. They have internal mechanism called gates, that can regulate the flow of information. The general architecture of LSTM is illustrated in Fig 1 .

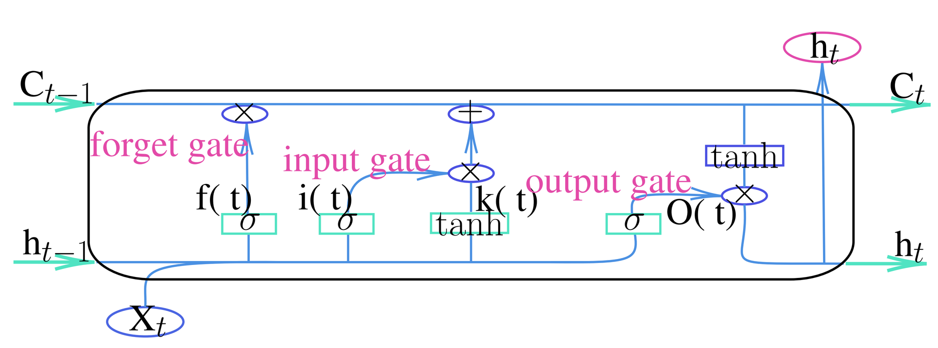

Figure 1: The architecture of LSTM

So, the forget gate is used to keep the important informations in memory from previous steps. The input gate decodes what information is important to add from the current step. Finally, the output gate determines the next hidden state.

The architecture of our model

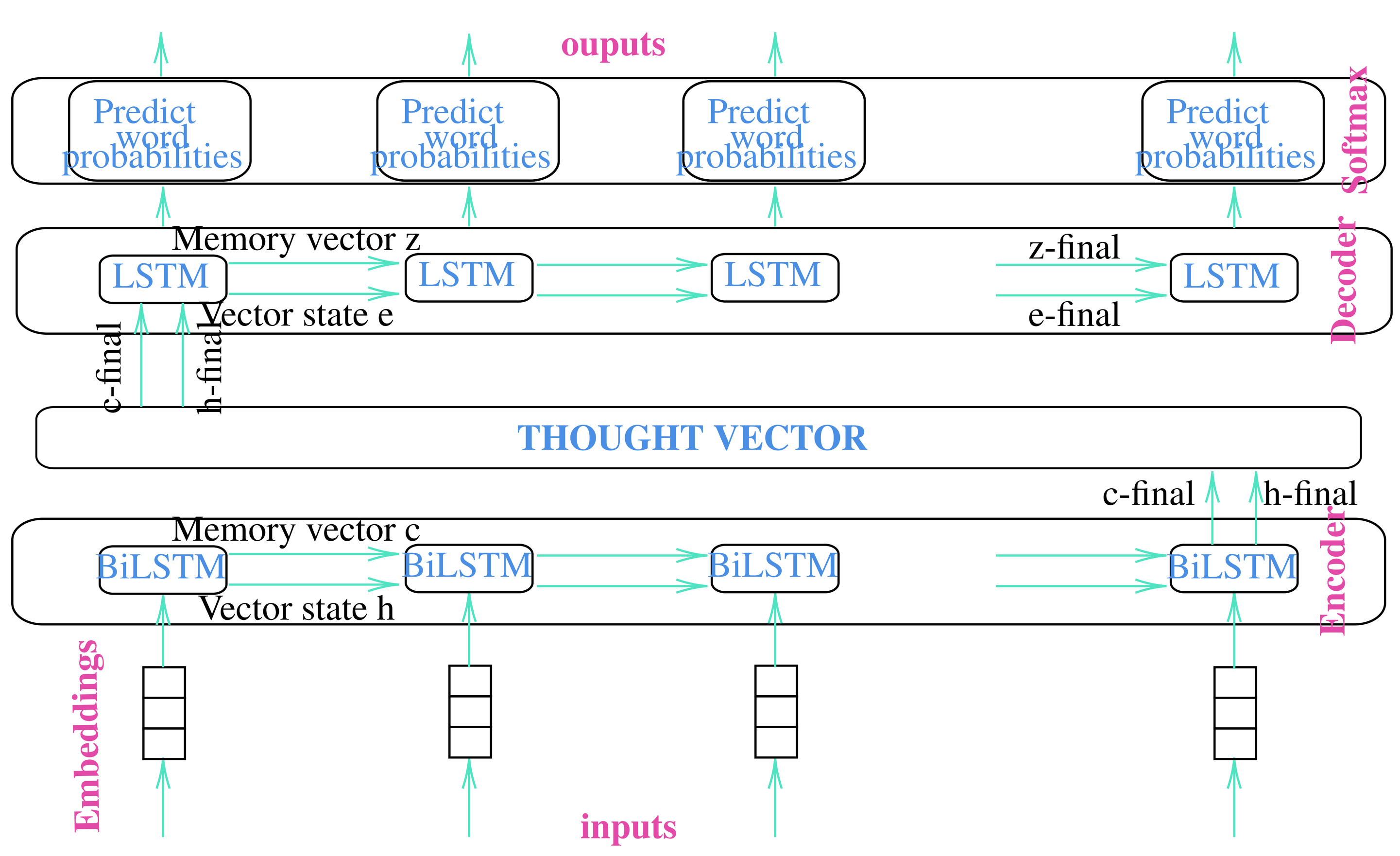

Figure 2: The architecture of our model

[1]https://github.com/susanli2016/NLP-with-Python/tree/master/data

As shown in Fig 2 , we create as first step a vector that represents the English sentence, we embed our input sentence sequence into the BiLSTM encoder word by word until the end of the English sentence sequence. We obtain the hidden and cell (or memory) states and we feed the vectors that represent the meaning of the sentence into the LSTM decoder as its initial state. Finally, the output of the decoder is sent to softmax layer that is compared with the target data.

Results

To build our translation corpus, we have used the English-French parallel corpus from the github website[1] . Since our task is machine translation between the Arabic and English, our system starts by translating the English sentences into Arabic. Then, the best Arabic translation are selected for each English sentence to form our final translation corpus. Finally, all the sentences in both the English and Arabic languages are normalized and tokenized

In our sequence-to-sequence model, an embedding dimension of $\mathbf{R}^{20}$ for inputs and for $\mathbf{R}^{15}$ outputs have been used. The maximal sequence length has been set to 186 words for English and 519 words for Arabic.

A mini-batch size of 64 have been incorporated. The training has been done by means of stochastic Gradient Descent (SGD) with Adam optimization function [15].

Our model implemented using python has been trained using CPU with 4GB of memory.

Table 1: Translation results

Metric	Test Bleu score(%)
Ilya Sutskever et al. [11]	16
Our approach	**18**

(a) English-to-Arabic

Metric	Test Bleu score(%)
Ilya Sutskever et al. [11]	26
Our approach	**27**

(b) Arabic-to-English

The BLEU [16] obtained by our model and the approach proposed by [11] using our corpus are provided in Table 1a and Table 1b for the tasks of translation form English-to-Arabic and Arabic-to-English respectively.

So all these results show that our Bi-seq2seq gives best results, which demonstrates the efficiency of our proposal for the translation task.

Conclusion

In this work, we have presented a BiLSTM encoder and LSTM decoder model for the task of machine translation between English and Arabic texts . Our system addresses the case of machine translation between English and Arabic using a deep learning sequence-to-sequence model, which has not been investigated before, the obtained performances offer encouraging results in terms of correlation with human judgment. This work can be further developed in various directions. One way is to consider the case of translation between other languages besides French. Another interesting future one is to integrate this model into an English-to-Arabic machine transliteration system.

References

[1] John Hutchins. The history of machine translation in a nutshell. page 5.

[2] Eric Greenstein and Daniel Penner. Japanese-to-English Machine Translation Using Recurrent Neural Networks. page 7.

[3] Makoto Nagao. A framework of a mechanical translation between Japanese and English by analogy principle. *[No source information available]*, October 1984.

[4] Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard M. Schwartz, and John Makhoul. Fast and Robust Neural Network Joint Models for Statistical Machine Translation. In *ACL*, 2014.

[5] G. E. Dahl, D. Yu, L. Deng, and A. Acero. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):30–42, January 2012.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, May 2017.

[7] Boris Hanin. Which Neural Net Architectures Give Rise to Exploding and Vanishing Gradients? page 10.

[8] Rahul Dey and Fathi M Salem. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks. page 5.

[9] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555 [cs]*, December 2014. arXiv: 1412.3555.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, June 2017. arXiv: 1706.03762.

[11] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to Sequence Learning with Neural Networks. page 9.

[12] Mitchell Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330, July 2002.

[13] Kyunghyun Cho, Bart van Merrinboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Y Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. June 2014.

[14] Sepp Hochreiter and Jrgen Schmidhuber. Long Short-term Memory. *Neural computation*, 9:1735–80, December 1997.

[15] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, December 2014. arXiv: 1412.6980.

[16] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 311, Philadelphia, Pennsylvania, 2001. Association for Computational Linguistics.