Significant Enhancements in Machine Translation by Various Deep Learning Approaches

Alpana Upadhyay[*]

Department of MCA, Sunshine College, Gujarat Technological University, India

[*]**Corresponding author:** Alpana Upadhyay, Department of MCA, Sunshine College, Gujarat Technological University, India, E-mail: alpana.upadhyay@yahoo.com

Citation: Upadhyay A (2017) Significant Enhancements in Machine Translation by Various Deep Learning Approaches. Am J Compt Sci Inform Technol 5: 2. doi: 10.21767/2349-3917.100008

Abstract

A new and transforming technology for natural language processing and speech processing is deep learning. Deep learning extends various operative ways to train computer systems for learning and it gives significant advances for that. If the right system or architecture is developed with deep learning methods then the systems can automatically learn from data itself without the requirement of designing it overtly. This technique of machine learning changes the perspective of addressing natural language and speech technologies considerably.

Deep learning was first acquainting with Machine Translation in the standard statistical systems. This paper addresses the progress of introduction of deep learning in machine translation. It describes and includes all the topics like integrating deep learning in statistical machine translation, developing end-to-end neural machine translation systems, introducing deep learning in machine translation evaluation. Several research directions are drawn in terms of how deep learning can influence machine translation.

Keywords: Deep learning; Machine learning; Machine translation; Machine translation evaluation; Neural machine translation

Introduction

There are many advances in machine learning techniques now a day enhancing the capability of machine learning. One of the novice machine learning technique is deep learning that has been successfully applied to many extents of machine learning like image processing, speech processing and recognition, natural language processing etc. Deep learning has the prevailing capability to learn from data. Initially deep learning was integrated with standard statistical systems [1]. Deep learning has shown its capability in language modeling and translation [2]. Moreover, deep learning was applied in quality estimation and machine translation evaluation [3].

A new machine learning architype was emerged in terms of neural machine translation (MT) [4]. This new technique of machine learning has produced outstanding results for several language pairs [5]. This new paradigm employs attention based model with encoder decoder style to figure out with an end to end neural machine translation system [6]. This research approach is new and opens up new doors for the research in machine translation with challenges like pacing with huge vocabularies, translation of multimodal, high cost that infers to a new matter for large scale optimization [7].

Scientific community is very much interested in this hovering topic. This research paper is intended to mainspring the readers with insight, understanding, and global vision of limits of deep learning, its challenges and its impacts. This paper focuses on the various topics such as application of deep learning in statistical MT (machine translation), neural MT (neural machine translation), interactive neural machine translation, how deep learning augments machine translation evaluation. In the next this paper describes available alternatives to construct a neural MT system. The rest of the section in this paper deliberates the prominent research aspect of how to apply deep learning for machine translation.

About Neural Machine Translation (Neural MT)

Almost all neural Machine Translation (MT) rely on the approach of encoder-decoder. In this approach input sequence is in source language and it is presented in a low dimensional space from that output sequence in targeted language is produced [4]. To encode the input sequence many options are available to design an encoder. One of the approaches is to utilize simple RNN-recurrent neural network to encode the input sequence [8]. On the other hand to project source side information, compressing a sequence in a fixed sized vector is very depreciatory. To solve this problem a new system was built up that system uses bidirectional RNN-recurrent neural network. In this system source sequence is encoded using annotations by combining two presentations acquired by backward and forward RNN. In this system every annotation vector holds information

from complete source sequence though it concentrates on one exacting word [6].

After this process an attention mechanism implemented by a FFNN-feed forward neural network is used to focus on the precise parts of the input and to establish an association between input and output sequence. Stacked LSTM-Long Short Term Memory is the other method that can work as an alternate to bidirectional RNN [9]. Dealing with the large softmax

Figure 1: Conventional architecture for NMT system (Neural Machine Translation system)

Figure 1 describes the conventional architecture for neural MT system that is used for softmax normalization. The other way to carry out translation is performing it at the sub word level. This method has few advantages like it allows creation of the words which are out of vocabulary. Machine translation at the character level is also presented by few researchers [13,14]. Other than these one more technique called BPE-byte pair encoding is also a widely used method used for variety of language sequences [15].

Classification of Various Deep Learning Approaches

There are various categories where deep learning is used. These categories are as follows.

- Statistical machine translation by deep learning
- NMT-Neural Machine Translation
- INMT-Interactive Neural Machine Translation

Statistical machine translation by deep learning

The very first approach to incorporate deep learning or neural network in Machine Translation was using rescoring n-best lists from statistical machine translation systems, known that statistical machine translation gives up to date outcomes and deep learning aids to find out correct set of weights for statistical data [16,17]. There are improvements suggested for statistical machine translation that use deep learning.

Domain adaptation using joint neural network models: In general MT systems, domain adaptation is still a confront, in view of the fact that parallel data is measured as a sparse resource. Neural translation models like joint models have given away an enhanced adaptation capability due to continuous representation. There are various ways to adapt such type of

normalization on the output that is reliant on the size of target vocabulary, is the main problem with neural machine translation.

Many researchers have researched to attend to this problem. They gave the solution like using structured output layer to manage the output [10], or carry out softmax on a subset of the output [11], or performing self-normalization [12].
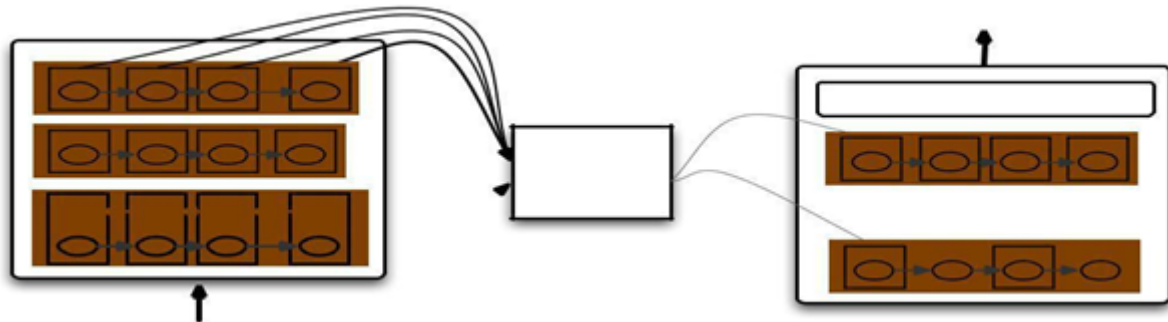
models. Better estimation of weighting and instance selection is an important factor in neural model whereas data selection and mixture modeling is the starting point. Pair wise modeling, which reduces the cross entropy by regularizing the loss function with respect to model selected [17].

Figure 2 shows the conventional architecture of joint neural network model. Joint neural network model which is prepared on plain concatenation can be sub optimal.
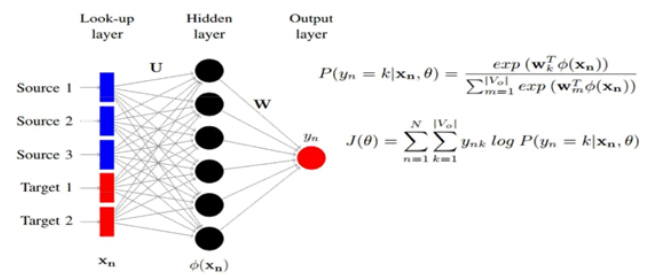

Figure 2: Conventional architecture for NNJM system (Neural Network Joint Model)

The model is trained using weighted concatenation. Flow to out domain model is proscribed using regularizer depending on in domain model as follows.

$$J(\theta_a) = \sum_{n=1}^{N} \sum_{k=1}^{|V_o|} \left[\lambda \, y_{nk} \, log P(y_n = k | \mathbf{x_n}, \theta_a) + (1 - \lambda) \right.$$
$$\left. y_{nk} \, P(y_n = k | \mathbf{x_n}, \theta_i) \, log P(y_n = k | \mathbf{x_n}, \theta_a) \right]$$

Regularizer depends on cross entropy difference of in and out domain model.

$$J(\theta_a) = \sum_{n=1}^{N} \sum_{k=1}^{|V_o|} \Big[\lambda \, y_{nk} \, logP(y_n = k|\mathbf{x_n}, \theta_a) + (1 - \lambda)$$

$$y_{nk} \Big[P(y_n = k|\mathbf{x_n}, \theta_i) - P(y_n = k|\mathbf{x_n}, \theta_a) \Big] \, logP(y_n = k|\mathbf{x_n}, \theta_a) \Big]$$

Source sequence simplification for statistical machine translation: Translating long sentences is again a big challenge in machine translation. Text simplification and hierarchical machine translation decoding with neural rescoring is used to handle this problem.
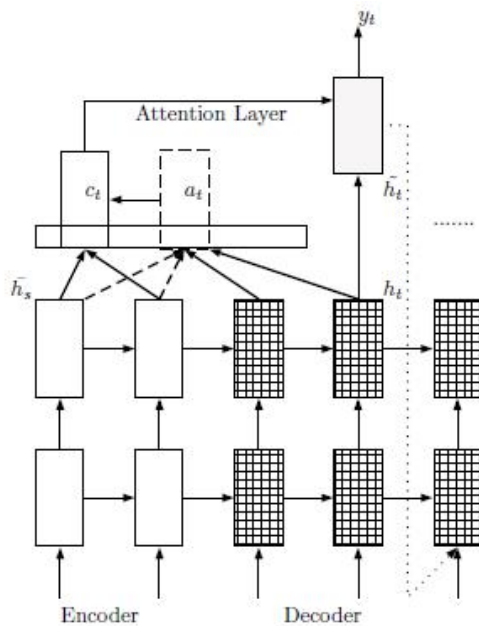
Figure 3: Hierarchical architecture for neural source sequence simplification model

Figure 3 demonstrates the architecture for source sequence simplification model. Context vector ct has been calculated for attention layer and weighted average at has been calculated with the alignment weights. The first step is to simplify the sentence. Every redundant word is removed from the statement. On the other hand at decoder, the assimilation is done through a two-step decoding method. From these two steps, first step translates simplified input and generates n best lists. In the second step, n best lists are used to decode complete input sequence.

NMT-neural machine translation

Encoder-Decoder architecture is come out as a key elucidation for a number of translation requirements [4,18]. Although it gives tremendous results in monolingual data, it does not fit properly for multilingual input. For multilingual inputs it's a challenge. There are various approaches to address the problem of multilingual inputs. Following is the architecture of machine translation for multilingual inputs.
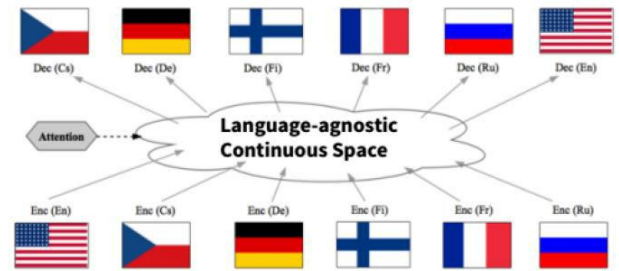
Figure 4: Architecture for multilingual machine translation

Context dependent word representation for NMT: Contextualization and symbolization is used to solve the problem of rare words. Contextualization is carried out by using masquerading few dimensions of targeting word. Word embedding vector supported on the average of non linearly translated to source word embedding. Symbolization is performed by introducing position-dependent special tokens to deal with digits, proper nouns and acronyms (Figure 4).

Multi way multi-lingual NMT: In an extremely multi lingual inputs, multi way multi-lingual neural machine translation method with a shared attention system with across language pair is used. To manage completely multi lingual inputs, multiple encoders and decoders are used and that eventually increases the complexity of the system. One encoder per source language and one decoder per target language is used to encode-decode each source and target language pair and that instance of the multi-lingual environment is treated as mono-lingual. A deep synthesis of all the instances of the system coalesces the scores of neural machine translation.

INMT-interactive neural machine translation

Machine translation has shown remarkable results in last few years, in spite this technology needs some improvements to give good outcomes for some applications and few other domains especially when quality of fully automated system is not sufficient. INMT-Interactive neural machine translation is a competent solution. It delineates association between a machine translation system and human translator. This technology is editing method for translators which interrelate with machine translated outputs. In this method suggestions are made for auto complete the translation by machine translation system. Machine translator either agrees to follow the suggestion or writes new translation. The very first step is that the neural machine translation system is tailored to best fit prefix basis interactive method.

In this method user reads translation hypothesis from left to right and is supposed to correct the translation hypothesis by generating the translation prefix with new hypothesis that is completed by machine translation. Tag vectors are maintained in source input sequence for keeping track of attention history. This attention based system focuses on considering more non translated words. This system keeps track of interaction history by representing source input with both the operations of

reading and writing. Figure 5 demonstrates the architecture of attention based interactive neural machine translation. This interactive system is useful for the decoder to repeatedly keeping track of translated parts and non-translated parts. The biggest advantage of this interactive translation system is it reduces human efforts significantly.
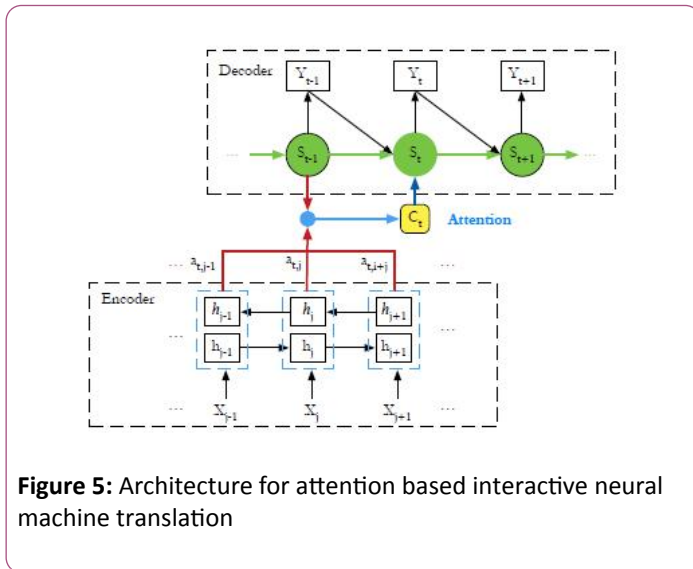

Figure 5: Architecture for attention based interactive neural machine translation

Research Viewpoints

Deep learning is a recent technology and major advancement in machine learning that has been applied very successfully to many areas of machine learning like speech recognition, image processing, natural language processing and many more. Deep learning techniques have very powerful ability to quickly learn complex tasks from data. This technology has astounded both communities namely industry and academy. Deep learning has very strong capacity in terms of rescoring, tuning, phrase extraction, reordering, language modeling and most importantly in machine translation. It can also be used for quality estimation and machine translation evaluation. Neural MT is pretty new paradigm of deep learning that has produced very promising outcomes for various language pairs. This new paradigm can deal with creating a new architecture of end to end NMT system that can easily deal with morphologically rich languages and very large vocabularies. This technology has grasped the attention of scientific community at a large scale and this paradigm has unlocked a huge array of research perceptions.

Presently almost all the neural machine translation systems are dependent on auto encoder and decoder architecture. To manage syntactic differences and lengthy range reordering, it needs to have different encoders or attention based more powerful interactive model. Recent technology has some limitations to handle the complexity associated to manage large output vocabularies, non-canonical texts like social media and morphologically rich languages. To evade these limitations some solutions are under research.

References

1. P Koehn, F J Och, D Marcu (2003) Statistical phrase-based translation, in: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology 1: 48–54.

2. A Vaswani, Y Zhao, V Fossum, D Chiang (2013) Decoding with large-scale neural language models improves translation, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL.

3. R Gupta, C Orasan, J van Genabith (2015) Machine translation evaluation using re-current neural networks, in: Proceedings of the Tenth Workshop on Statistical Machine Translation, Association for Computational Linguistics, Lisbon, Portugal.

4. I Sutskever, O Vinyals, QV Le (2014) Sequence to sequence learning with neural networks, in: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems, Mont real, Quebec, Canada.

5. R Sennrich, B Haddow, A Birch (2016) Edinburgh neural machine translation systems for wmt 16, in: Proceedings of the First Conference on Machine Translation, Association for Computational Linguistics, Berlin, Germany, 371–376.

6. D Bahdanau, K Cho, Y Bengio (2015) Neural machine translation by jointly learningto align and translate, in: International Conference on Learning Representations.

7. MR Costa-jussa, J A R Fonollosa (2016) Character-based neural machine translation, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics Association for Computational Linguistics, Berlin, Germany, 2: 357–361.

8. K Cho, B van Merrienboer, C Gulcehre, D Bahdanau, F Bougares, et al. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, 1724–1734.

9. S Hochreiter, J Schmidhuber (1997) Long short-term memory, Neural computation 8: 1735–1780.

10. Jean S, Cho K, Memisevic R, Bengio Y (2014) On using very large target vocabulary for neural machine translation. Arxiv Preprint.

11. H-S Le, I Oparin, A Messaoudi, A Allauzen, F Yvon (2011) Large vocabulary SOUL neural network language models, in: INTERSPEECH.

12. J Devlin, R Zbib, Z Huang, T Lamar, R Schwartz, et al. (2014) Fast and robust neural network joint models for statistical machine translation, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1370–1380.

13. W Ling, I Trancoso, C Dyer, A W Black (2015) Character-based neural machine translation, CoRR abs/1511.04586.

14. M R Costa-Jussa, J A R Fonollosa (2016) Character-based neural machine translation, CoRR abs/1603.00810.

15. R Sennrich, B Haddow, A Birch (2016) Neural machine translation of rare words with subword units, CoRR abs/1508.07909

16. H Schwenk, M R Costa-Jussa, J A R Fonollosa (2006) Continuous space language models for the IWSLT 2006 task, in: 2006 International Workshop on Spoken Language Translation, IWSLT Keihanna Science City, Kyoto, Japan, 166–173.

17. A Upadhyay, Kumbharana, C K (2017) Analysis of Functional Parameters to Implement Knowledge Management for Sustainable e-Governance in Agriculture Sector of Saurashtra Region of Gujarat State. In ICT Based Innovations 233-244. Springer, Singapore.

18. N Kalchbrenner, P Blunsom (2013) Recurrent continuous translation models, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, Washington, USA 1700–1709.