International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(1), pp.399-403
DOI: http://dx.doi.org/10.21172/1.101.73
e-ISSN:2278-621X

RECENT APPROACHES OF NEURAL MACHINE TRANSLATION: A REPORT

Mr. Satish Kumar[1], Dr. Mohammed Asger[2]

Abstract- Neural Machine Translation (NMT) is a recent advance proposal technique in machine translation. Neural Machine Translation surpassed the results of Statistical Machine Translation. This report discusses some significant works on the recent advances in neural machine translation. The aim of NMT is to build a single neural network that generates maximum translation performance. This report present the overview of neural network, neural language model which include Recurrent Neural Models (RNN) and Neural Machine Translation models which include encoder-decoder approach, attention-based neural machine translation.
Key words and Phrased: Neural Network, Neural Machine Translation, Back Propagation Training, Refinements, encoder-decoder approach, attention-based model Feed forward approach.

1. HISTORY OF NEURAL MACHINE TRANSLATION

Machine translation applied the machine power to do automatic translation from source natural language to target language. Machine translation idea fort explore by Warren Weaver 1946. Earlier, rule based machine translation (RBMT) used in study of linguistic information to generating translation based on grammar and dictionary. Later Statistical Machine Translation models are applied to machine translation which generates translation based on the analysis of parallel corpus. R. Neco and M. Foracods in 1997come up with a technique called “encoder-decoder” to do machine translation. In the year 2003, Yoshua Bengio developed a language model using the neural network techniques. A new end to end encoder-decoder structure was proposed by Nal Kalchbrenner and Phil Blunsom for machine translation that source test into a continuous vector by using Convolution Neural network decode state vector into the target source by using Recurrent Neural Network. Neural Machine translation based on non-linear model differs from the linear model of statistical machine translation and used the state vector for semantic equivalence which connects encoder and decoder.

In 2014 Cho et al and Sutskerver at al [4] developed a sequence to sequence method learning based on recurrent neural network for encoder and decoder and proposed a varaity of recurrent neural network called Long short-term memory (LSTM) [8] for neural machine translation. The LSTM effectively explain the “Long Distance Rendering” problem by transferred the basic challenge of neural machine translation to the problem of “Fixed Length Vector”. While comparing the source sentence into fixed length vector using neural network, the complexity and uncertainties increase during decoding, especially for long sentence. In 2014, Yoshua Bengios [2] proposed the attention mechanism to solve the problem of fixed length vector. Initially attention mechanism used for image classification that neural network to focus on appropriate of input during performing prediction task. When a work is generated by decoder to form target language sentence then just small part of source language sentence are relevant and so, to generate a context vector based on the source sentence, a context based attention mechanism are applied. The word of target language predicted based on context vector instead of fixed length vector. By comparing neural machine translation with statistical machine translation, neural machine not required prior domain knowledge and it train multiple table together. Neural machine translation represent better sentence structure and has less syntax error, word order error morphological error which are commonly occur in statistical machine translation. On the other hand neural machine translation face problem and challenge in translation which need to handle. Neural machine translation training and decoding process is low as compared to statistical machine translation.

2. ARTIFICIAL NEURAL NETWORK

Artificial neural networks are non linear computational models that inspired by human brain. Artificial neural network include voice recognition, machine recognition, robotics image recognition and most recent advancements in the field of artificial intelligence. The term neural represent the basic functional unit of “neuron” nerve cell of human nervous system. A typical neural of human brain composed of four major parts include cell body that process all incoming signals to generate result, dendrite which receive information from other neurons and synapses the part of neuron where one neural connect to other neuron. The amount of signal strength transmitted from one neuron to another depends upon synaptic weight. The signal can be excitatory (high strength) or inhibitory (low strength) in nature. In general an artificial neural network is highly interconnected network of neurons.

[1] Department of Information Technology, Baba Ghulam Shah Badshah University, Rajouri-185131, (J&K), India 2018
[2] Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri-185131, (J&K), India 2018

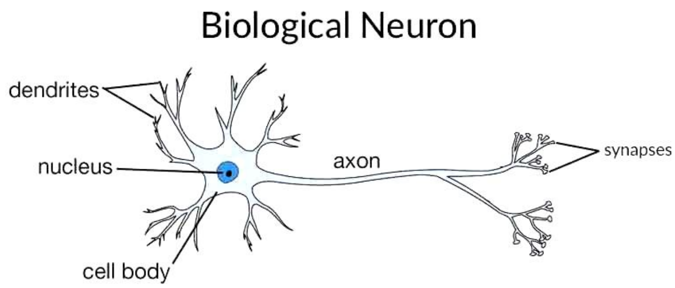

Figure 1: Show elements of biological neuron

How does artificial neural network work? Artificial Neural Network represented as weight diverted graph consist of nodes that represent artificial neurons, directed edges represent synaptic weight. It received input information in the form of pattern and image in vector form. The architecture of typical artificial neural network contains billions or trillion of artificial neurons arrange in series of layers following figure represent the architecture of neural network with different layers.

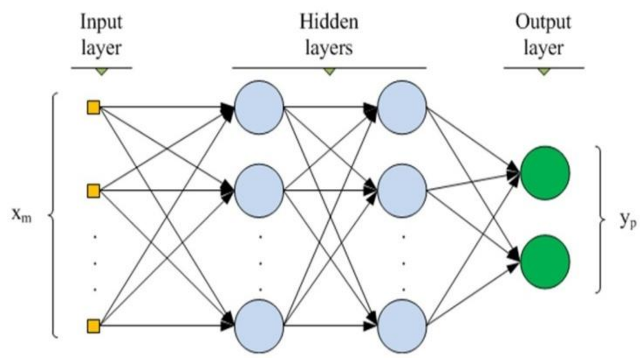

Figure 2: Show neural network architecture

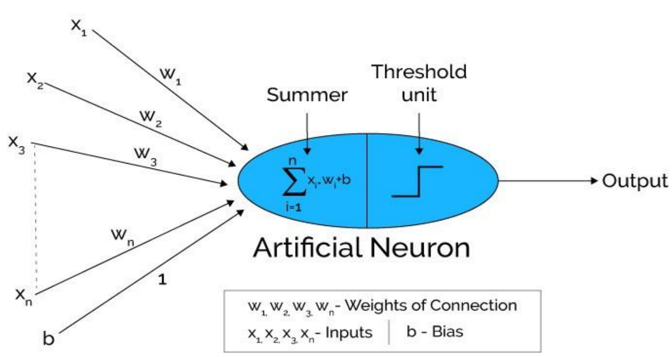

Figure 3: Show single layer perception

The input layer contains artificial neurons which receive input on which network will learn, recognize or process. The output layer contains neurons which respond to information provided by neurons that how it learned by task. The hidden layer placed between input and out layer, process and transform input information so that output layer uses it. There are different types of artificial neural networks which are categorized based on pattern, hidden layer, nature of weight and memory unit. These are:- Feed-forward Neural Network: - has no loop and Recurrent Neural Network has loop because of feedback, both are based on connection pattern.

Single-Layer Neural Network:- having single hidden layer (single percepton)

and Multilayer conatin multiple hidden layers (Multiple Percepton), both are based on hidden layer.

Static Neual Network and Dynamic Neural Network:- These models are based on memory unit. Static neural network doesn't have content memory. In static neural network current input depend on current output. For example feedforward network. Dynamic neural network have content memory. In dynamic neural network current output depend on both current input as well as current output. For example Recurrent Neural Network.

2.1 Neural Language Model

A language model is a technique for learning a function to control salient statistics feature of sequence of word in computational linguistics and make probabilistic predictions of next word. Natural language model are effective technique to model probability distribution having many inputs.

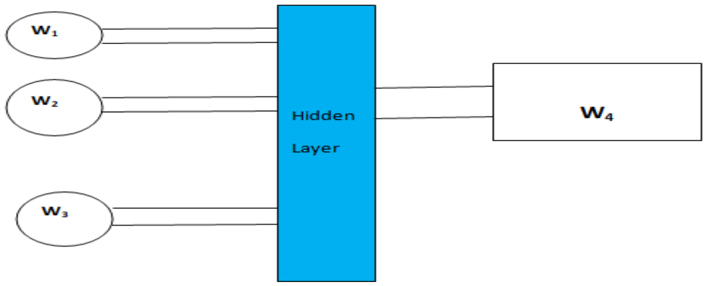

Figure 4: Neural Language Model: - To predict a word wi that is based on its preceding words.

2.2 Recurrent Neural Network based Language Model

In artificial intelligence sequential data prediction examine as a key problem. The statistical language modeling predicts the next word in green content of data. While structuring language models in machine translation we only concerned sequential data prediction. Many such models proposed that are particular for language domain. One of the most widely and general model used for language based on n-gram statistics which assume that language consist of sequence of atomic symbols call words that consist sentences. A question arise is there any advance progress in language modeling over n-gram model. If we models as progress tool to predict sequential data then improvements have achieved such models are like cache models, class based models etc. For machine translation huge amount of data is required to language models.

In this overview we have mention recurrent neural network approach for modeling sequential data. Feedforward neural networks approach of artificial neural network used in statistical language model for fixed length content. This approach show better performance than other model like class based model --- in his research explore that models based on neural network generated significant improvement in machine translation for major task as compare to mixture of other models. Feedforword approach of neural network has limitation that its use fixed length context that require ad hoc before training, which means neural network need some preceding words when predicting next one. So, a model is required that encode temporal language data absolutely for unpredictable length context.

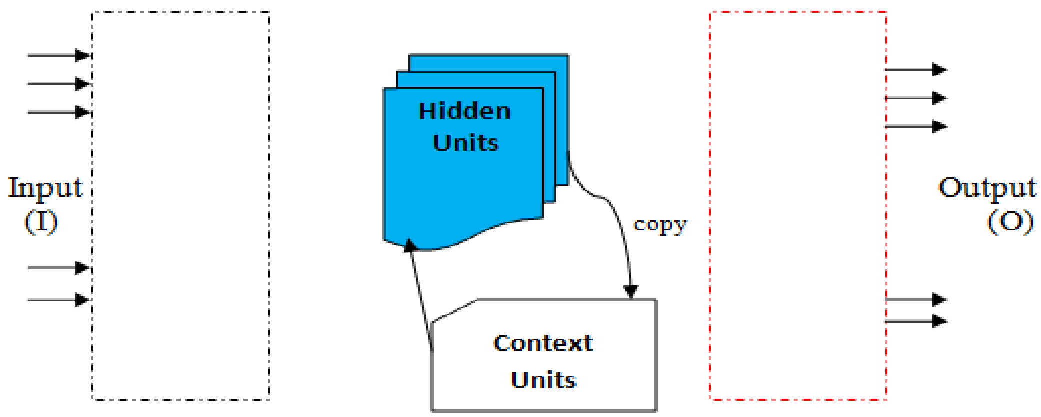

Figure 5: Show Recurrent neural network architecture

Recurrent neural network use unlike feedforword neural network use unlimited size of context. In recurrent neural network, information loop inside the network for random amount of time. But result shows that it is different to learn long term dependences to target stochastic gradient descent. In this work we have analyze a simple recurrent neural network. This recurrent neural network architecture is very simple and easy to implement. As show in figure a recurrent neural network has an input layer (i), hidden layer or context layer (h) and output layer (o). at a time t the input to network is i(t), state of hidden layer is h(t) and output of network layer is o(t). The input to network is obtain by concatenating weight w represent input word and output from hidden neurons h at time t-1. All the layers input, hidden, and output are evaluated as :-

$I(t) = w(t) + h(t-1)$

$hj(t) = f(\sum Ii(t)xji)$

ok(t) = g(∑ hj(t)ykj)

and sigmodel function is give by

$f(z) = \frac{1}{1+e^{x}}$

and softmax function computed as

$$g(zn) = \frac{e^{Zn}}{\sum_k e^{zk}}$$

The only difference between recurrent neural network and feedforward neural network proposed by Bengio [6] is parameters selection to tune before training. In feedforword one can tune words layer size, hidden layer size and context length size. In recurrent neural network one has to be set only size of hidden layer. In order to increased the performance, all the words access less than threshold into special token set. The probability of word evaluated as fellow

$$P\ (w\ (t+1))\ w\ ((t), h\ (t-1)) = \begin{cases} \frac{\text{Orare(t) w(t+1) rare}}{\text{Orare}} \\ O_i\ (t)\ \text{otherwise} \end{cases}$$

For further improvement Schwenk et al. described particular approaches and other possible approach also described that can applied mostly to recurrent neural networks. Thus proposed recurrent neural network model is sophisticated because it provides better relation of language model with machine learning, data compression etc.

2.3 Neural Machine Translation Model

Recently Neural Machine Translation has been using attentional mechanism that applied mostly on part of source sentence during translation. There are much useful architecture for attention-based neural machine translation. Here we mentioned two effective approaches, one is global approach and other is local approach [1]. Global approach attends to all source words whereas local attends subset of source of words at a time. In machine translation neural network has achieved state-of-art performance. Neural machine translation required small scale domain of knowledge. Louong et. al. 2015 model attend all the source words until end mark of sentence reached.

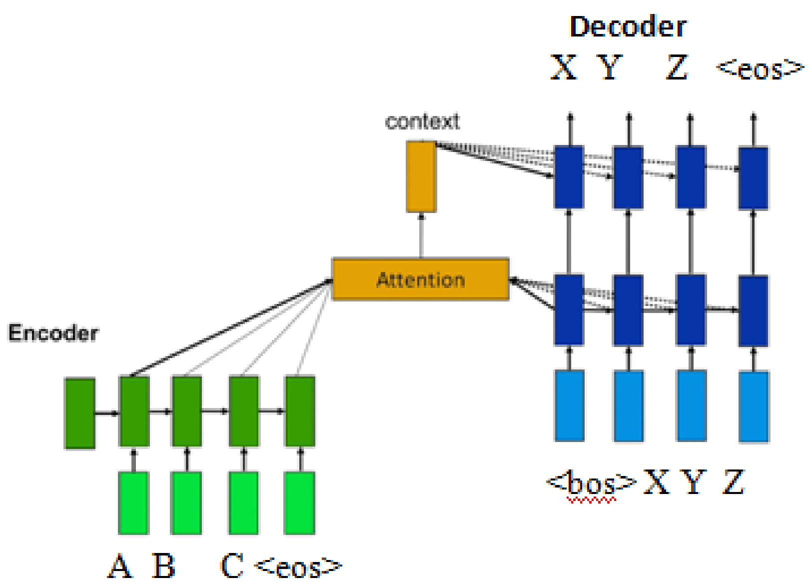

Figure 6: Recurrent Architecture for translation a source sentence A B C into target sequence X Y Z end of sentence and begin of sentence mark with <eos>, <bos> respectively.

In Neural Machine Translation attentional mechanisms is applied to jointly translation and align words [7]. In this paper we explore two novel approaches of attention based model. First, global approach in which all source words attended and second local approach in which only subset of words are applied at time. Both these approached differ by attention i.e. attention is attended all source word N only few source words. Figure illustrate global attentional model and figure illustrate attentional model. In both model given source context vector ct and hidden state ht applied to concatenation layer so that information collected from both state to form an attentional state given as

ht = tanh(wc[ct,ht])

This attentional vector applied to softmax to generate predictivity distribution

P(yt/y<t,x) = softmax(wsht)

Global attention: - In global attention model the variable length alignment vector at is

$$at = (\mathrm{align}(ht^{T}, hs)) = \frac{\exp(\mathrm{score}(ht^{T}, hs))}{\sum \exp(\mathrm{score}(ht^{T}, hs)}$$

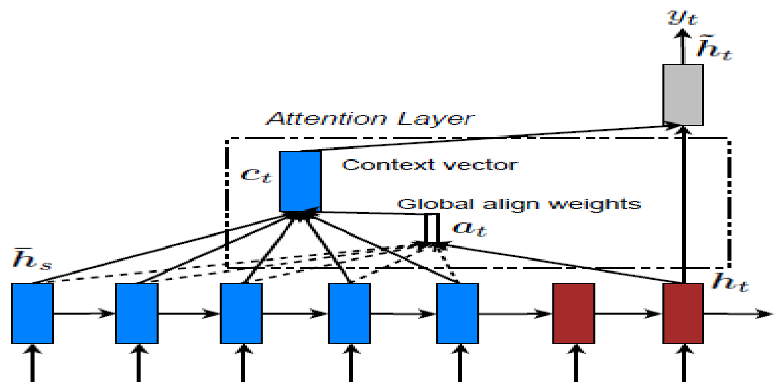

Figure 7: illustrate Global attentional model evaluation steps.

Score (context base function) is one of the following:-

Dot Score(ht, hs) = htT, hs

General score (ht, hs) = htTwa, hs

Concat score (htT, hs) = vatanh(wi[htT, hs])

4. CONCLUSION

Neural Machine Translation has an active research area of machine translation. But still there are many challenges face during translation which can be solved by detail study of natural language. So many new techniques are required to develop a complete automatic neural machine translation.

5. REFERENCES

[1] K. Cho, B. Merrienboer, D. Bahdanau and Y. Bengio. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches.
[2] A Graves, N. Jaitly and A. Mohamed. Hybrid Speech Recognition with Deep Bidirectional LSTM , 2013.
[3] D. Bahdanau, K. Cho and Y. Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. ICLR, 2015.
[4] K. Cho. Introduction to Neural Machine Translation with GPUs .
[5] O. Vinyals, A. Toshev, S. Bengio and D. Erhan. Show and Tell: A Neural Image Caption Generator. 20 Apr 2015.
[6] K. Cho, A. Courville and Y. Bengio. Describing Multimedia Content using Attention-based Encoder--Decoder Networks. IEEE Transactions on Multimedia, 2015.
[7] R.Sennrich, B. Haddow, A. Birch. Neural Machine Translation of Rare Words with Subword Units. ACL, 2016.
[8] M. Schuster and K. Nakajima. Japanese and Korean voice search. ICASSP, 2012.
[9] Y. Wu, M. Schuster, Z. Chen, Q. Le, M. Norouzi and the Google Brain team. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. 8 Oct 2016
[10] A.Graves and N. Jaitly. Towards End-To-End Speech Recognition with Recurrent Neural Networks. ICML, 2014.
[11] J. Gu, G. Neubig, K. Cho and V. Li. Learning to Translate in Real-time with Neural Machine Translation. 10 Jan 2017
[12] Karpathy. The Unreasonable Effectiveness of Recurrent Neural Networks .
[13] Neural Machine Translation (ACL 2016 Tutorial)
[14] Y. Bengio, A. Courville and P.l Vincent. Representation Learning: A Review and New Perspectives. 23 Apr 2014