International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 05, 2019 ISSN: 1475-7192

Machine Translation Using Deep Learning: A Survey

[1]Santosh Behera, [2]Sanjay Behera

Abstract--- *Machine Translation (MT) using Deep Learning is a newly proposed method for machine translation. The term MT is employed within the sense of translation of one language to a different, with no human improvement. It can also be referred to as machine-controlled translation. Unlike the standard statistical machine translation, the neural MT aims at building one neural network that may be collectively tuned to maximize the interpretation performance. This survey reveals the data about Deep Neural Network (*DNN*) and the concept of deep learning in the field of natural language process i.e. MT. It's higher to use* recurrent Neural *Network (RNN) in MT. This article studies numerous techniques wont to train RNN for various languages corpses. RNN structure is extremely sophisticated and to train a large corpus is also a time-consuming task. Hence, strong hardware support (Graphics Processing Unit) is needed. GPU improves the system performance by decreasing training period of time. Integration of deep learning in machine learning increase from re-modelling existing features into a statistical system to the development of a new model. Among the different neural network research work use feed forward neural network (F-CNN), recurrent neural network (R-CNN), and the encoder- decoder schema.*

Index Terms--- *Machine Translation, Neural Network, Deep Learning, Neural Machine Translation*

I. INTRODUCTION

The information society is continuously evolving towards multilingualism: e.g. totally different languages apart from English are gaining additional and a lot of importance within the web, and robust societies,[1] just like the European, are and will continue to be bilingual. Different languages, domains, and language designs are combined as potential sources of data. In such a context, machine translation (MT), which is the task of automatically translating a text from a source language into a target language,[2] is gaining a lot of and more relevance. Both business and academy are powerfully investigating within the field that is progressing at an incredible speed Deep Learning could be a recently used approach for MT. in contrast to the standard machine translation, the neural machine translation is a better choice for a lot of accurate translation, and it also provides better performance[3]. Deep Neural Network (DNN) is wont to improve traditional systems to make them a lot of efficient. Different deep learning techniques and libraries are requiring for developing a higher machine translation system. RNN, LSTMs, etc. are used to train the system which is able to convert the sentence from source language to target language. Adapting the appropriate networks and deep learning methods could be a good selection because it tuned the system towards increasing the accuracy of the translation system as compared to others[4].

Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, santoshkumarbehera@soa.ac.in

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

I.I. Deep Learning

Deep learning (DL) is part of machine learning (ML) techniques based on learning information representations, as opposition task-specific algorithms. Machine Learning will be supervised, partially supervised or unsupervised[5]. Deep learning architectures like deep neural networks, deep belief networks, and recurrent neural networks (RNN) are applied to field including computer vision (CV), speech recognition, natural language processing (NLP), audio recognition, social network filtering, machine translation, and Bioinformatics wherever they created results comparable to and in some cases, superior human experts[6].

Deep learning is a brand-new technique, widely used in completely different machine learning applications. It permits the system to find out like a human and to enhance the efficiency with training[7]. Deep learning strategies have the capability of feature representation by using supervised/unsupervised learning, even there exist higher and a lot of abstract layers. Deep learning presently used in image processing application, huge data analyses, speech recognition, machine translation, etc.[8]

I.II. Deep Neural Network

A deep neural network (DNN) is an Artificial Neural Network (ANN) with multiple hidden layers between the input and output layers[9]. DNNs are usually feed-forward neural networks during which data flows from the input layer to the output layer without looping back. Recurrent neural networks (RNNs), [10]during which data can flow in any direction, are used for applications such as language modeling. Long short-term memory (LSTM) is particularly effective for this use. Convolutional neural networks (CNNs) are used in computer vision. CNN's also have been applied to acoustic modeling for automatic speech recognition (ASR)[11].

I.III. Machine Translation

Machine Translation (MT) could be a sub-field of linguistics that investigates the use of a computer software package to translate text or speech from one natural language to a different language. At its basic level,[10] Machine Translation (MT) performs an easy substitution of words in one, natural language for words in another. An artificial intelligence system is required to translate literary works from any language into native languages. The literary work is fed to the machine translation (MT) system and translation is completed[12]. Such machine translation systems will break the language barriers by creating available work made sources of literature accessible to individuals across the world. Figure 1 shows the method of machine translation within the kind of the pyramid. Deep learning (DL) attracts researchers for using it in machine translation. The most idea behind this is often to develop a system that works as a translator[13]. With the help of history and experiences, a trained deep neural network translates the sentences while not using a massive database of rules.

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

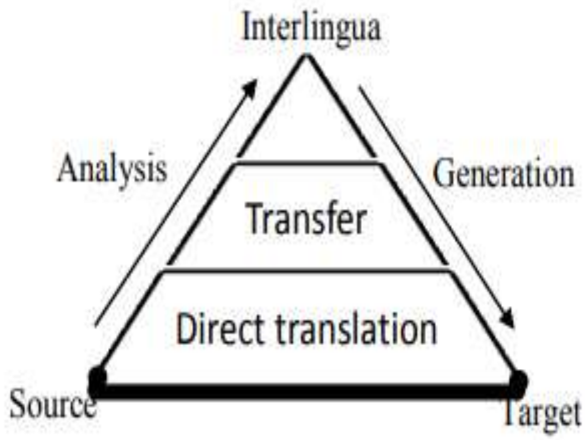

Fig.1: Machine Translation Pyramid

II. ENCODER- DECODER FOR MACHINE TRANSLATION

This type of design has been impressed by auto encoders that try and predict their own input. Encoder-decoder design generalizes the concept of auto encoders allowing for having completely different input and output information[14]. The encoder-decoder design aims at learning an illustration (encoding) of input data and decodes this representation whereas minimizing the quantity of error for recovering the output data. The most purpose of the internal representation could be a dimensionality reduction capable of extracting relevant features from the dataset[15]. A schematic illustration is shown in Figure 2

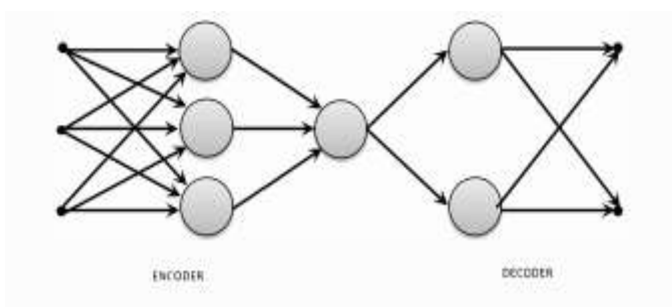

Fig. 2: Encoder – Decoder

Neural machine translation (NMT) models adhere to an Encoder-Decoder design, the role of an encoder is to represent absolute length sentences to a fixed-length real vector that is termed as context vector[16]. This context vector contains all the mandatory features which may be inferred from the source sentence itself. The decoder network takes this vector as input to output target sentence word by word. The perfect decoder is predicted to output sentence that contains the complete context of the source language sentence[9]. Since source and target sentences are sometimes of various lengths initially proposed recurrent Neural Network for each encoder and decoder networks to deal with the problem of vanishing gradient and exploding gradients occurring because of dependencies among word pairs,[17] Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) were proposed rather than Vanilla RNN cell. Fig.1Shows the architectural

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

flow of a basic encoder-decoder based network.

- **Language Modeling**

Language modeling is the task of scoring sequences of words. Approaches with neural networks to language modeling have a protracted history[18]. This segment covers the approach neural networks have enhanced monolingual language modeling in applied mathematics MT systems, whereas bilingual language modeling is according to use a continuous space language model impressed within the classical approaches to enhance the N-gram-based MT system. The neural network is a Multilayer Perceptron,[17] trained as a classifier with the input projection, hidden and output layers. The projection layer of input words represent the distributed encryption of input words and uses a linear activation function. The hidden layer uses a hyperbolic tangent because of the activation function and also the output layer is a Softmax layer. During this case, the continuous space language model is employed in restoring for each a phrase-based and N-gram-based MT system[19].

- **Word Alignment**

Word alignment could be a key task in statistical machine translation systems since it identifies word-by-word relationships given a pair of sentences that are corresponding translations. IBM's models are one of the most standards, probabilistic formulations of this drawback and are using successfully with the GIZA++ implementation[20]. Recently, there have appeared some approaches that use deep learning to perform this task. Yang et al. (2013) use the methodology of DNN used in speech recognition to find out to extract lexical translation data. The model integrates a multi-layer neural network into a Hidden Markov Model (HMM) framework,[21] from wherever they extract context-dependent lexical translation. The model is trained on a bilingual corpus and uses monolingual data to retrain word embedding.

- **Translation Modeling**

Given work in the literature, this paper tend to distinguish studies done on bilingual translation models, on phrase-based models and on syntax-based models[22]. The most difference depends on the actual fact that the bilingual translation models follow a language model structure with bilingual units, whereas the phrase-based models use bilingual units with no context, finally, the syntax-based models incorporate explicitly a representation of syntax by parsing the sources and/or target sentences following a kind of grammar[23].

  - **Bilingual Language Model**

Early approaches in using deep neural networks in bilingual translation models are commonly two-step systems, which implies that an n-best list is proposed in a very traditional method, then the continuous space modeling is used to restore these lists[24]. Schwenk et al. (2007) propose to project bilingual units onto a continuous space as an extension from previous work on monolingual language modeling (Schwenk et al. 2006). Then, this projection allows estimating the translation possibilities during this continuous representation[25]. Bilingual units act as neural network inputs. Again, the authors face the problem of computational complexity that is solved by limiting the vocabulary. Zamora et al. (2010) apply the same neural language model to each the bilingual and therefore, the monolingual language model and additional relevant,[26] the decoder is extended with neural language modeling throughout Viterbi, which supplies higher results

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

than restoring. Lesson et al. (2012) propose the same design, however, authors use two vectors within the input layer coming from the source and target language[27].

III. CONCLUSION

In modern times, machine translation is a very popular research topic in the natural language processing area. Deep learning helps to train a translation system sort of human brain. RNN and RAE give a better lead to text processing as compared to different neural networks. From particularly papers regarding machine translation, I have analyzed that deep learning is much better than the other technique for giving correct results. However, the processing time is high, also it needs more time to train a system. This paper tend to then mention the role of deep learning models in improving different parts of SMT, then this paper tend to shift our discussion on end-to-end neural machine translation (NMT). Our discussion was mostly supported by the fundamental encoder-decoder based NMT, attention-based model. This paper tend to finally list the challenges in Neural Translation models and mentioned future fields of study and open-ended problems.

REFERENCES:

[1] S. Ruder, M. E. Peters, S. Swayamdipta, and T. Wolf, “Transfer Learning in Natural Language Processing,” 2019.
[2] M. Z. Alom et al., “A state-of-the-art survey on deep learning theory and architectures,” Electronics (Switzerland). 2019.
[3] Y. Keneshloo, T. Shi, N. Ramakrishnan, and C. K. Reddy, “Deep Reinforcement Learning for Sequence-to-Sequence Models,” IEEE Trans. Neural Networks Learn. Syst., 2019.
[4] X. He, J. Gao, and L. Deng, “Deep Learning for Natural Language Processing and Related Applications,” Tutor. 2014 IEEE Int. Conf. Acoust. Speech, Signal Process., 2014.
[5] N. Tandon, A. S. Varde, and G. de Melo, “Commonsense Knowledge in Machine Intelligence,” ACM SIGMOD Rec., 2018.
[6] L. Munkhdalai, T. Munkhdalai, O. E. Namsrai, J. Y. Lee, and K. H. Ryu, “An empirical comparison of machine-learning methods on bank client credit assessments,” Sustain., 2019.
[7] M. Indhraom Prabha and G. Umarani Srikanth, “Survey of Sentiment Analysis Using Deep Learning Techniques,” in Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019, 2019.
[8] S. N. Esfahani and S. Latifi, “A Survey of State-of-the-Art GAN-based Approaches to Image Synthesis,” 2019.
[9] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato, “Unsupervised machine translation using monolingual corpora only,” in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
[10] M. Artetxe, G. Labaka, E. Agirre, and K. Cho, “Unsupervised neural machine translation,” in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
[11] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature. 2015.
[12] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Frcitas, “Dueling Network Architectures for Deep Reinforcement Learning,” in 33rd International Conference on Machine Learning, ICML 2016, 2016.
[13] M. Johnson et al., “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,” Trans. Assoc. Comput. Linguist., 2017.
[14] Q. V. Le, “Building high-level features using large scale unsupervised learning,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2013.
[15] J. Dean et al., “Large scale distributed deep networks,” in Advances in Neural Information Processing Systems, 2012.

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

[16] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014.
[17] I. Bello, B. Zoph, V. Vasudevan, and Q. V. Le, "Neural optimizer search with Reinforcement learning," in 34th International Conference on Machine Learning, ICML 2017, 2017.
[18] X. Lin et al., "All-optical machine learning using diffractive deep neural networks," Science (80-.)., 2018.
[19] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in 34th International Conference on Machine Learning, ICML 2017, 2017.
[20] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., 2017.
[21] S. Gu, T. Lillicrap, U. Sutskever, and S. Levine, "Continuous deep q-learning with model-based acceleration," in 33rd International Conference on Machine Learning, ICML 2016, 2016.
[22] R. Collobert, "Deep learning for efficient discriminative parsing," in Journal of Machine Learning Research, 2011.
[23] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers, 2014.
[24] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," Neural Networks. 2019.
[25] Z. Liu, X. Li, P. Luo, C. C. Loy, and X. Tang, "Deep Learning Markov Random Field for Semantic Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., 2018.
[26] N. Carlini, C. Liu, J. Kos, Ú. Erlingsson, and D. Song, "The Secret Sharer: Measuring Unintended Neural Network Memorization & Extracting Secrets," arXiv, 2018.
[27] H. Larochelle and Y. Bengio, "Classification using discriminative restricted boltzmann machines," in Proceedings of the 25th International Conference on Machine Learning, 2008.

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019