International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 2, Issue 6, pp 20-23, August-2018

Removing Language Barrier: A Survey of Machine Transliteration

[1]Er. Hari K.C., [2]Er. Sharan Thapa

[1,2] Department of Electronics and Computer Engineering, Tribhuvan University, Institute of Engineering, Paschimanchal Campus, Nepal.

Abstract **- Nepal is a country with Provinces having varieties of Nepalese spoken languages. The official language is however Nepali, spoken and understand by majority percent people of Nepal. The majority of People don't understand English although English is an universal language. Machine Translation system translating the text from one language script to another language script to enhance the knowledgeable society of Nepalese without any language barrier. Machine Translation is the difficult and challenging task of Natural language Processing. Nepali is the language used by majority of Nepalese and English being the universal language, the necessity of English to Nepali machine translation is significant. This paper is the survey paper that describes the different approaches in the field of Computational Linguistic for Machine translation and their challenges.**

Keywords: Machine Translation, Computational Linguistic, Natural Language Processing, Deep learning, Neural Network.

I. INTRODUCTION

Machine Translation is the field of computational linguistics which translates the text or speech from one natural language to another using machine called computer. It is the semantic analysis of Natural language processing for automatic translation. Source language and target language need to be identify for machine translation task. On a primary level, translation performs simple substitution of words in one language for words in another, but that alone usually cannot produce a good translation of a text. Recognition of whole phrases and their closest counterparts in the target language is needed. Neural techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies [1].

The rise of social networking and World Wide Web is connecting the people from the various parts of the world. Most of the information or data are available in English and many people are unable to access the information due to language barrier. People can be made aware with the information available on internet with machine translation. English to Nepali translation will remove the language barrier for the Nepalese people. Effective machine translation will improve the communication between the people of different countries as it increase the exchange of information. In the past, machine translation was not so much effective due to the use of rule based machine translation. After the increase in computational power of computer, different approaches of deep neural machine translation are providing the good performance in machine translation.

II. RELATED WORK

During 17th century, the research in machine translation started with the idea of using dictionaries to overcome the language barriers [3]. After that, the research continues and today, the neural machine translation has shown the good performance in translation.

In 19th century, IBM developed the statistical model for machine translation. In the beginning of 20th century, phrase based systems are developed. Google also provide machine translation for its user. In Nepal, the research in computational linguistic with machine translation is in very slow pace. No machine translation system has been developed so far. Different machine translation research papers are now available. In paper titled “A new Chinese-English machine translation method based on rule for claims sentence of Chinese patent”, the researcher shows the effective translation of Chinese script to English Script [5]. Similarly, the researcher shows the hindi – English translation in paper “A hybrid approach for Hindi – English machine translation”[6]. The most powerful language Sanskrit had also translated to English in paper “Interlingua based Sanskrit- English machine translation” [7].

III. MACHINE TRANSLATION APPROACHES

Machine translation system requires knowledge of natural language and its usage.

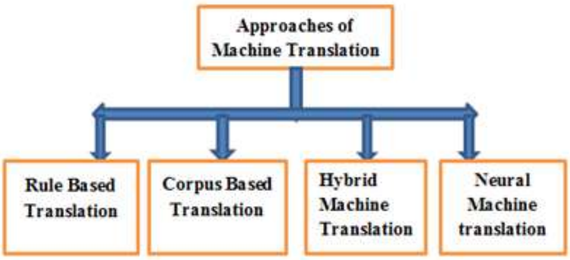
Approaches of Machine Translation
Rule Based Translation
Corpus Based Translation
Hybrid Machine Translation
Neural Machine translation

Figure-1: Approaches of Machine Translation

a) Rule Based Approach of Machine Translation

Rule Based machine translation are traditional based machine translation based on knowledge. It is also known as knowledge based machine translation. By parsing the source text, the intermediate representation is produced. From this intermediate representation, the target script text is produced. The linguistic knowledge about the source and target language is needed. The morphological, syntactic and semantic analysis need to be performed to know information about source and target language.

• Direct Translation

In direct machine translation, word to word translation is performed between the source and target text. Both bilingual and unidirectional translation can be performed in direct machine translation. The challenges in direct translation are:

- Lack of bilingual dictionaries
- Chance of Mistranslation at lexical and word order.

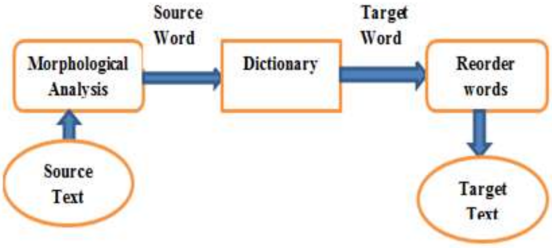

Figure-2: Direct Machine Translation

• Transfer Based Translation

Transfer Based translation is developed to remove the limitation of Direct translation. It consists of 3 stages of analysis, transfer and generation. Database of translation rules are prepared. In first stage, source text syntactic analysis is performed. In second stage, source text syntactic structure is transferred to target text structure. In third stage, target text is generated from syntactic structure of target text. The challenges in Transfer Based translation are:

- Big data have various ambiguities and dialects which make difficult to derive rules.
- Transfer modules need hard and vast rule.

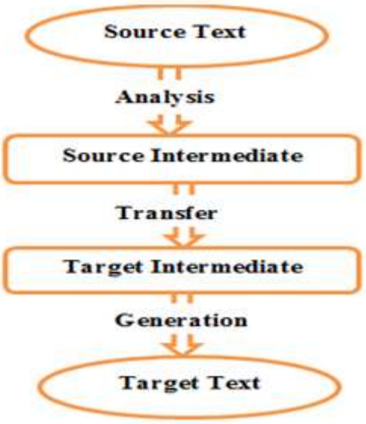

Figure-3: Transfer Based Machine Translation

• Inter- Lingua Translation

Direct machine translations are less efficient approach of traditional approach of machine translation. Inter-lingua is the alternative approach for Direct translation. Source text is translated to intermediate abstract text and then to target text. Since it is language independent, it is also suitable when number of target languages increases. The challenge of this machine translation is to preserve the meaning of a sentence and to build the Interlingua knowledge base.

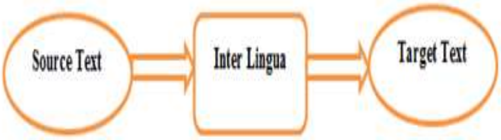

Figure-4: Inter- Lingua Machine Translation

b) Corpus Based Machine Translation

Corpus Based translation are also known as Empirical machine translation. The corpus of source text and target text is prepared and statistical analysis is performed. Large database of parallel corpora need to be prepared for this translation.

• Statistical Machine Translation

Statistical Machine Translation is based on statistical model build by using machine learning algorithms. The statistical model provides statistical information such as correlation between source text and target text. The challenges in statistical machine translation are:

- Difficult to create parallel corpus.
- Different word order text language are not suitable.
- Unexpected Result can be seen.

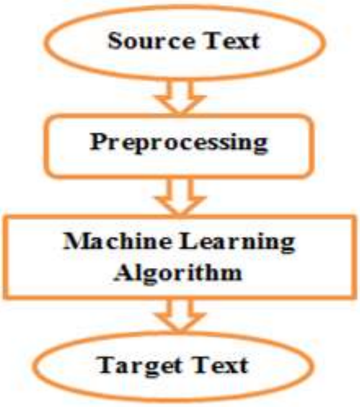

Figure-5: Machine translation using statistical model

• Example Based Machine Translation

Example based machine translation is based on corpus consisting source and target text examples in database. Database is stored in translation memory. It decreases the user effort for re-translating the text. The challenges in Example based machine translation are:

- Difficult to make adaptation and retrieval model
- Difficult for different structure text.

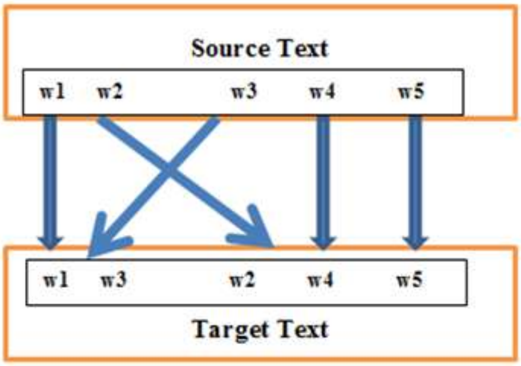

Figure-6: Example based machine translation

c) Hybrid Approach of Machine Translation

Nowadays, hybrid approach is in trend for machine translation. This translation uses multiple translation approaches combined to provide best features. Both the rule based and statistical approach can be combined to give the best performance hybrid approach. Hybrid approach gives best performance compare to other approach. The challenges in Hybrid approach are:

- Building bilingual corpus is very costly and difficult.
- Maximum linguistic resources are needed.

d) Neural Machine Translation

Due to the increase in computation power of computer, neural network deep learning algorithms are used with this machine translation approaches to provide best performance with high accuracy.

• Encoder Decoder Approach

Encoder Decoder approach is an older approach in neural machine translation. In this approach, the feature vector is computed which has a fixed length from the source text. The feature vector corresponding to each vector is fed to the Encoder. The hidden state in encoder captures the relevant information. Weight vectors are learned during training process.

The whole input text vector from encoder is the fed to the decoder then decoder produces the target text using the hidden state and target words.

The key benefits of the approach are the ability to train a single end-to-end model directly on source and target sentences and the ability to handle variable length input and output sequences of text.

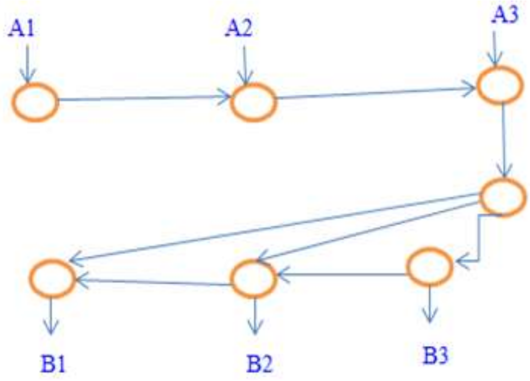

Figure-7: Encoder – Decoder approach of neural machine translation

• Attention Based approach

Attention based approach uses variable length vector in encoder section. It uses the bidirectional recurrent neural network, forward and backward neural network. The context vector is produced from the encoder section. Then, the decoder section, compute conditional probability of a destination word given source words. The weight vector is computed using softmax function. The attention-based approach is a suitable method for handling long sequences and capturing the long-range dependencies in source text [2].

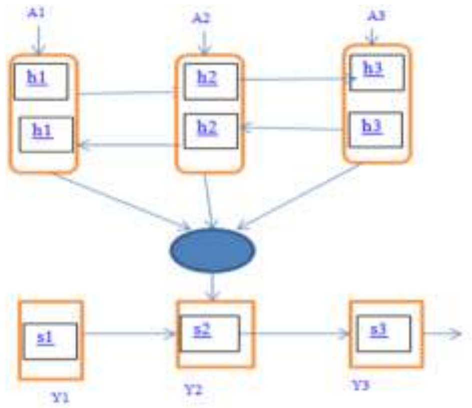

Figure-8: Attention based approach

IV. CONCLUSION

In this paper, different machine translation approaches were reviewed. Various researcher uses machine translation system according to their application. In a multilingual machine translation environment, transfer based approach may be the best option. Neural machine translations achieve the best performance over large lexical database. The research in machine translation has not been initiated in Nepal. The Government and Private research institute should be involved in this area of machine translation. Natural language processing machine translation is a difficult and hard problem for Nepali language since it contains very complex lexical structure.

REFERENCES

[1] X. Zhang, S. Tao, Z. Gong, B. Wu, R. Wang, B. M. Wilamowski, "An improved English to Chinese translation of technical text" in 2015 *IEEE 19th International Conference on Intelligent Engineering System (INES),* 2015.

[2] M. M. Mahasuli, R. Safabakhsh, " English to Persian transliteration using attention based approach in deep learning" in 2017 *Inranian Conference on Electrical Engineering (ICEE),* 2017.

[3] D.V. Sindhu, B.M. Sagar, " Study of Machine Translation approaches for Indian Languages and their languages" in 2016 *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT),* 2016.

[4] http://machinelearningmastery.com/encoder-decoder-recurrent-neural-network-models-neural-machine-translation.

[5] W. Xiong, Y. Jin, " A new Chinese – English machine translation method based on rule for claims sentence of Chinese Patent" in 2011 *7th International Conference on Natural Language Processing and Knowledge Engineering,* pp 378-381, 2011.

[6] O. Dhariya, S. Malviya, U. S. Tiwary, " A hybrid approach for Hindi- English machine translation", in 2017 *International Conference on Information Networking (ICOIN),* 2017.

[7] H.S. Sreedeepa, S.M. Indicula, "Interlingua based Sanskrit- English machine translation", in 2017 *International Conference on Circuit, Power and Computing Technologies(ICCPCT),* pp 1-5, 2017.

Citation of this article:

Er. Hari K.C., Er. Sharan Thapa, "Removing Language Barrier: A Survey of Machine Transliteration", *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, Volume 2, Issue 6, pp 20-23, August 2018.

****