MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

# ASSAMESE TRANSLATION SYSTEM

**Shayanika Hazarika, Poonam Suryawanshi, Nikita Shinde**
**Prof. Monika Dangore**

**Department Of Computer Engineering, Dr. D. Y. Patil School Of Engineering, Pune, India**

***Abstract:*** *Automatic Speech Recognition is the method in which speech signals are translated into the related sequence of characters into text (word). The research on speech recognition has been done for many years ago. After a lot of research work we got the idea of RNN (Recurrent Neural Network) algorithm is more effective and helpful algorithm for speech recognition. The Recurrent Neural Network is the type of Neural Network techniques in which we observe the difference of alphabet. The application or purpose of the RNN is to make people comfortable for the use of Hindi, English and Assamese languages. With the help of this research, people can easily understand languages and get its insight. Languages pose different challenges for speech recognition. This paper discusses about the use of RNN and Statistical Machine Translation algorithms that can be used for translation systems.*

***Keywords****: Recurrent Neural Network, Language Model, Translation Model, Statistical Machine Learning.*

## 1. INTRODUCTION

Speech-to-speech translator systems are developed over the decades, the main motive is to bridge the gap between communication for people who speaks their native language. Suchsystem has been broken in such a way that the Recurrent neural network to hide the background processing i.e. the hidden layer, statistical machine translation method Like deep learning for speech translation. The text-to-text as well as speech-to-speech synthesis for the translation of such language is implemented in our system. Here the Direct Speech to speech translation of the Hindi to Assamese is accomplished. This system which maintains the sequence or the flow to translate the sentence from Hindi Language to Assamese

Language. Here we have used the the Assamese dialects for the translation purpose. This system is depend on the RNN where the output of the previous process is given to the next process as a input for further processing. This system does processing by first training the model and then predicting the output of the given test dataset. The speech model can wait for 14 sec time span to record your voice, until and unless you speak the further output cannot be predicted. The voice or the speech which needs to be translated should be clear, with such accurate input the output you get will be accurate.

## 2. MOTIVATION

Human beings are able to communicate with each other and it is the most fundamental part. Worldwide there are around 7000 different languages. As the world is getting connected, language translation helps in bridging different cultures and economy between people from different countries and groups.

The motivation behind taking up this project is that different Indian languages have Google translators as well as Microsoft translators but languages like Assamese , Konkani etc. are resource scarce language they do not have APIs available for translation. Therefore we tried developing an application as well as web based translator for Hindi to Assamese as well as English to Assamese text as well as Speech translator.

## 3. EXISTING SYSTEM

**Statistical Machine Translation :** Machine translation investigates how to automatically translate text or speech across normally human languages. Machine translation can use a method based on dictionary entries, which means that the words will be translated as a dictionary does – word by word, usually without much correlated of meaning between them. In this paper we are going to discuss about dictionary based machine translation. Dictionary based machine translation is basically suitable for the translation of long phrases but not for sentences level. Statistical machine translation probabilistic models P(T|S) estimated from parallel corpora.

**Parallel Corpora:** Parallel corpora are collections of texts available in two languages. Alignment levels: documents, sentences, words

**Alignment:**Parallel corpora typically aligned at sentence level. But in order to extract pairs of phrase translation, we need word alignments.

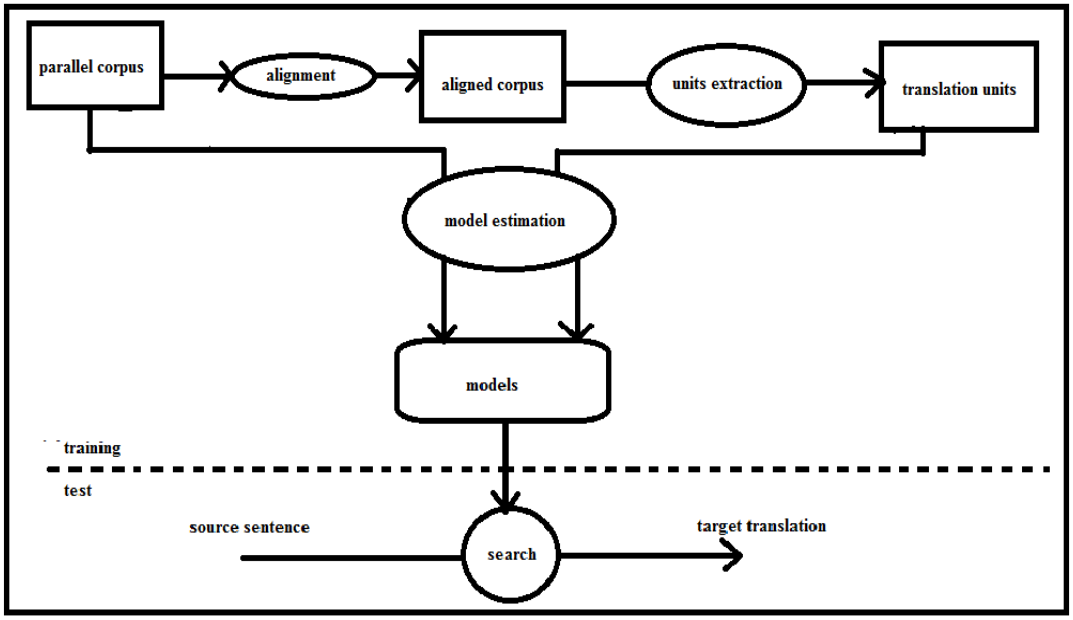


*Fig 1. Statistical Machine Translation*

## 4. PROPOSED SYSTEM

To translate a corpus of English text to Assamese, we need to build a recurrent neural network (RNN).

**Recurrent Neural Network:** In the Neural Networks, the Out of the First process is given as as Input to next process is known to be Recurrent Neural Network. In the standard processing of neural network, the all processes which are executing are independent of each other. In some scenario, the next processing requires the output for the previous process for prediction, so the previous word needed so the system has to remember the previous processing output. For such issues the RNN is discovered. The RNN solves this type of issues Which has the middle layer which is Hidden Layer. Here the Hidden layer needs information of a sequence. The RNN has a memory, it stores all the information for all the previous calculation. The same parameter is used in all processing's or the Middle layer which is hidden to produce output. The RNN specially reduces the tension for the parameters unlike other Neural network.

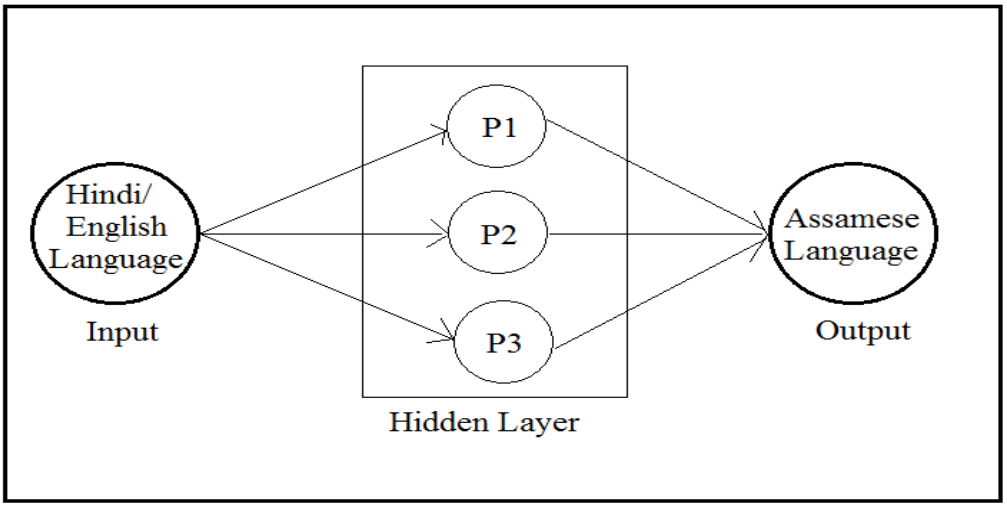


*Fig 2. Recurrent Neural Network*

The Main Advantage of the RNN , It Remember all the information for every step to step prediction . This feature is unique for RNN. It is a Long Short Term Memory.

**Hardware Requirements:**

- System: Pentium IV 2.4 GHz.
- Hard Disk :1 TB.
- Ram: 4GB (min).

**Software Requirements:**

- Operating system : Windows XP/7.
- Coding Language: JAVA, Python.
- IDE : Spyder, AndroidStudio.
- Front End: JSP, CSS etc.

## *5.* ADVANTAGES

**Business**: international trade, investment, contracts, finance

**Commerce**: travel, purchase of foreign goods and services, customer support

**Media**: accessing information via search, sharing information via social networks, localization of content and advertising

**Education**: sharing of ideas, collaboration, translation of research papers

**Government**: foreign relations, negotiation

## 6. CONCLUSION

The travellers can use this model to communicate with the local people. This will reduce the burden of learning new language and will make communication easier. This model translates English as well as Hindi text and speech to Assamese text as well as speech. This model is built using RNN(Recurrent Neural Network) with LSTM(Long Short Term Memory).

## 7. FUTURE SCOPE

This System is used by the people who are travelling to Assam for different purposes, and who are travelling from Assam to all over the India. In future The same type of application will be developed for different resource scarce languages. In Hindi language there is Ambiguity issue, example: Main Hu Aam Aadmi (मैं हूं आम आदमी'). The meaning of "Aam" is Common and the other meaning of "Aam" is mango, so these types of Ambiguity can be solved in future development of such models.

## ACKNOWLEGEMENT

It gives us great pleasure in presenting the paper on "Assamese Translation System". We would like to take this opportunity to thank our guide Prof. Monika Dangore and Dr. Pankaj Agarkar(HOD Computer, DYPSOE) for giving us all the help and guidance we need with indispensable support, suggestions and motivation during the course of the Project. We are really grateful to him. His valuable suggestions were insightful.

## REFERENCES

*[1]. Mani Bansal, Goonjan Jain, "Improvement of English-Hindi Machine Translation using ConceptNet", 2017.*

*[2]. SameerBansal, HermanKamper, Adam Lopez, Sharon Goldwater, "Towards Speech-to-text Without Speech Recognition", April 2017.*

*[3]. N. D. Londhe, IEEE Senior Member, M. K. Ahirwal, P. Londha, "Machine Learning Paradigms for Speech Recognition of an Indian Dialect", International Conference on Communication and Signal Processing, April 6-8, 2016.*

*[4]. ShahramKhandiviand Herman Ney, Senior Member,IEEE "Integration of Speech Recognition and Machine Translation in Computer-Assisted Translation " IEEE TRANSACTIONS ON AUDIO,SPEECH, AND LANGUAGE PROCESSING, VOL.16,NO.8, NOVEMBER 2008*

*[5]. Bowen Zhou ,Senior Member, IEEE " Statistical machine translation for speech: A Perspective on Structure, Learning and decoding " Proceeding of the IEEE , Special Issue: Speech Information Processing*

*[6]. Alex Graves, Abdel- rahman Mohamed and Geoffrey Hinton "speech recognition with deep recurrent neural networks"*

*[7]. WojciechZaremba, IlyaSutskever, Oriolvinyals "recurrent neural network regularization"Under review as a conference paper at ICLR 2015*

[8]. *Dr.R.L.K.venkateswardu, Dr.R. VasanthaKumari, G. VaniJayaSri "Speech recognition by using recurrent neural networks"International Journal of Scientific and Engineering Research Volume 2, Issue 6, June-2011 ISSN 2229-5518*