REGIONAL LANGUAGE TRANSLATOR USING NEURAL MACHINE TRANSLATION.

[1]Vaishali Kharat , Utkarsha Chaudhari , [3]Kirti Kesarwani
Information Technology
Usha Mittal Institute Of Technology
Mumbai,India
kharatvaishali77@gmail.com , utkarshachaudhari24@gmail.com,
Kirtikesarvani99@gmail.com
[4]Prof.Anita Morey
Assistant Professor
Usha Mittal Institute Of Technology, Mumbai,India
annu_km77@rediffmail.com

Abstract: Machine translation is a small component of computer languages focused on the task of automatically translating source text into one language into text in another language.Neural machine Translation (NMT) is a machine translation that uses a neural network to predict the sequence of words, usually a whole sentence model with one integrated model.

Keywords: Artificial Intelligence, Machine Translation, Natural Language Processing.

I. INTRODUCTION

Machine Translation (MT) is an important task that aims to translate natural language sentences using computers. The early approach to machine translation relies heavily on hand-crafted translation rules and linguistic knowledge.In machine translation, the input already has a series of symbols in another language, and the computer system must convert this into a series of symbols in a different language.

As natural languages are inherently complex, it is difficult to cover all language irregularities with manual translation rules. With the availability of large-scale parallel corpora, data-driven approaches that learn linguistic information from data have gained increasing attention.As neural machine translation attracts much research interest and grows into an area with many research directions, we believe it is necessary to conduct a comprehensive review of NMT. We hope that by tracing the origins and evolution of NMT, we can stand on the shoulders of past studies, and gain insights into the future of NMT. Neural machine translation is a radical departure from previous machine translation approaches. On the one hand, NMT employs continuous representations instead of discrete symbolic representations in SMT. On the other hand, NMT uses a single large neural network to model the entire translation process, freeing the need for excessive feature engineering. The training of NMT is end-to-end as opposed to separate turned components in SMT. Besides its simplicity, NMT has achieved state-of-the-art performance on various

language pairs.In practice, NMT also becomes the key technology convert a English sentence to different regional or foreign languages and vice versa from text to voice using Neural Machine Translation for that we develop a translation mechanism NMT is use of neural models to learn a statistical model for machine translation that can give an input sentence.the Our aim is to translate given sentence from one language to another and replace statistical models with neural machine translation for getting better results.

A. Objectives of the Study

In today's world technology plays a very important role. The main objective of the natural language processing machine translation is- Being able to smoothly, automatically and accurately translate one spoken or written language to another.Analyze and compare in detail the structural and performance of the NMT and SMT systems.

II. LITERATURE REVIEW

Hindi to English & vice versa Parallel Corpus Generation from Comparable Corpora for Neural Machine Translation, Prof. Amitabha Mukherjee. 600 peer test time is used to create a peak of nearly 1500 matching pairs made after setting a reasonable limit on alignment parameters to obtain high quality similarity to test the quality of the produced 100 samples of those pairs and rated by 6 human testers.

Development of English-Hindi Interactive Machine Translation. Nidhi Sharma, Sambhavi Seth, Meenal Jain, Mehvish Syed (Oct 2019). In the design of an Interactive Machine Translation system which combines the strengths of both human and machine and provides a high-quality machine assisted translation. Through experiments we have verified our claim, as with the help of our system the speed of the human translators improved. It also improved the productivity of the translators. As an extension to this study, we wish to improve the user friendliness of the system by providing several features like providing a human translator an option to add their example translations in the knowledge base, before doing actual translations. We also wish to perform a more thorough evaluation to understand the expectations of the human translators from this system. One of the possible expectations is to reduce the key in effort. We shall perform the usability evaluation of the system to analyze such expectations and incorporate the same in subsequent versions of our IMT system.

ANUBHARATI Indian Institute of Technology, Kanpur (1995). Based on machine aided translation, which is a variation of the example-based approach, it provides a generic model for translation that is suitable for translation.Six Challenges for Neural Machine Translation, Philipp Koehn and Rebecca Knowles. SMT & NMT: Common toolkits for NMT and traditional phrase-based statistical machine translation with common data sets. They carry out experiments on English–Spanish and German– English. As for these language pairs, large training data sets are available.Spoken Language Translation, Sanket Gandhare Preethi Jyothi Pushpak Bhattacharyya. ASR uses mainly two models: acoustic model and language model to recognize speech.

III. METHODOLOGY

DATASET

Data is one of the most vital affairs of any research study; we are providing data which is a text file. We have to add a start and end token to eat the test and each sentence is padded to fit the maximum length criteria. The source and the target sentence are made into a worth pair format then we convert the textual data into numeric form .It is done by transforming each word to one hot encoding vector.then create the tokenizer and then apply tokenizer on source and target intent. We have to split the data set into frames and test it in a model. we have to use 80 percent data for training and 20 percent for testing.

ENCODING-DECODING

Encoder: The MNT model consists of an encoder and decoder. The system first reads the sentence using the encoder to build a thought vector. A thought vector is a series of numbers that represent some meaning of a sentence .The sentence which outputs the translation of a sentence in the output language.

TRAINING AND TESTING

Foremost downside of the encoder decoder version in sequence to sequence recurrent neural networks is that it could handiest work on short sequences this is the issue of the encoder version to memorize the long sequences and convert it into constant duration furthermore the decoder loses only one piece of records the remaining and sector hidden nation now this is wherein the idea of attention mechanism comes image parent rnn education with teacher forcing mechanism the intuition approximately the attention mechanism is that it can are expecting the subsequent phrase by using concentrating on the few relevant components of the sequencers in preference to looking on the entire sequence.

The intuition about the attention mechanism is that it can predict the next word by concentrating on the few relevant parts of the sequencers rather than looking at the entire sequence. We will be using a teacher forcing technique for faster training of our model.teacher forcing technique is used for deciding the next input to be given to the decoder.teacher forcing technique in which the target word is passed 10 as the next input to the encoder and decoder model with attention is trained using multiple epochs.

IV. VOICE TRANSLATION MODEL

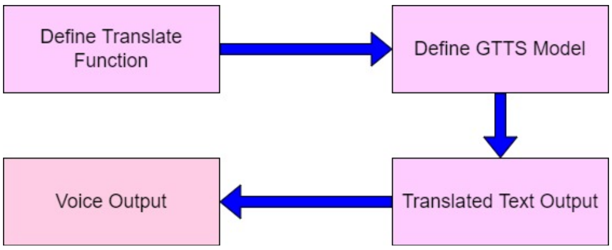

Fig 1: Voice module

This is the basic model of Voice Translation where a real time voice translator can translate text input and give translated voice output generated from it. Machine Translation (MT) components define translation functions in their original language. We are using the GTTS model (google text to speech) Library. It is the process of converting text into vocal audio form. We will import the gtts library from the gtts module which can be used for speech translation. The gtts API supports several languages. then we get translated text and voice output.

V. SYSTEM ARCHITECTURE & ANALYSIS

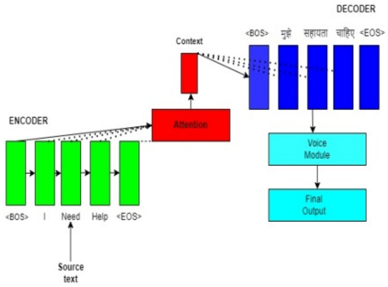

Fig 2: Architecture of system

Artificial Intelligence + Language =Neural Machine Translation: Neural Machine Translation is a very small method when considering the history of machine translation, but it is safe to say that it is now at a very advanced level. It is a very different approach to the language translation challenge that uses deep neural networks and artificial intelligence to train models for different language combinations and content types. As with all data-driven methods on MT, NMT engines learn inter-language translation based on a large number of previous translation examples - so-called training data. However, unlike its predecessors - based on law and MT statistics - Neural MT is able to think like the human brain when producing translation. Instead of translating word for word, or looking at small word order in sequence, it can take all the sentences, and more, in context. The result is a smoother, more accurate, and more efficient translation than ever before. (Block_8) Best Class Technology Encouraged hardware development, Neural MT development is moving at an unprecedented rate. On the basis of Iconic foundations - focused on the world of natural language processing - we have assembled a world-class team of MT experts to help us stay on top of the latest developments, build and improve on these developments, and transform them into productive solutions for our clients. This is reflected in our new technology - Ensemble Architecture - An iconic platform for the development and production of the MT engine. This expandable architecture allows us to build new applications, and connect and play the most effective solutions in the field to deliver the highest quality, regardless of how they are used.

Best-in-Class Technology

Spurred on by advances in hardware, Neural MT development is moving at a pace never seen before. On top of the foundations of Iconic – which are rooted in the world of Natural Language Processing -- we have assembled a world-leading team of MT experts to help us stay on top of the latest advances, to build and improve upon these advances, and to turn them into production-ready solutions for our clients.

This is reflected in our innovative technology – The Ensemble Architecture – Iconic's proprietary platform for MT engine development and production. This extensible architecture allows us to build new applications, and plug-and-play the most effective solutions in the field to deliver best-in-class quality, regardless of the use case.

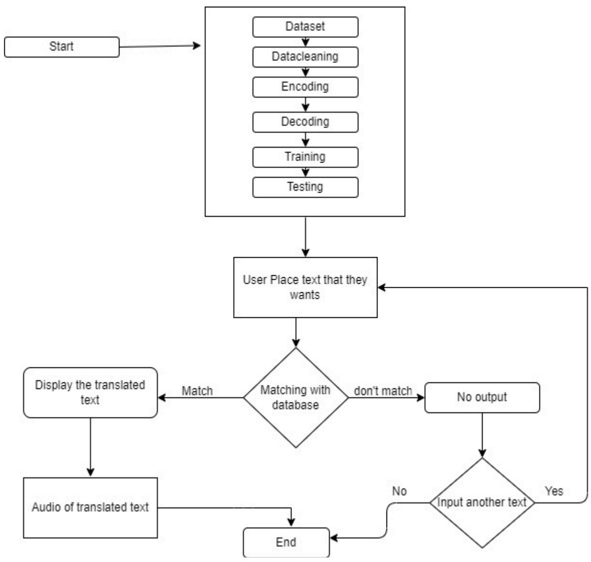

Fig 3 : Flowchart

Above Shows a flowchart of the machine translation using a

Neural network .The system will accept a database as an input.Users will choose the language they want and get output as translated text and audio format as well. We use NMT encoder decoder architecture .The features are extracted from the database and finally RNN training and testing. Then the system will give the correct translation.

VII. RESULTS AND DISCUSSION

Based on the analysis with various machine learning algorithms,we have found Neural machine translation to be the best suited to the needs of this system. We used Recurrent neural network for encoding and decoding and teacher forcing technique to train the model For better performance and quick training.

Fig 4: Translated Text

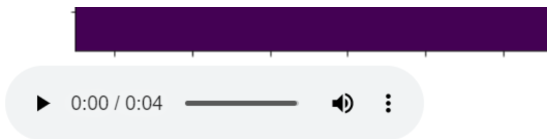

Fig 5 : Translated Audio

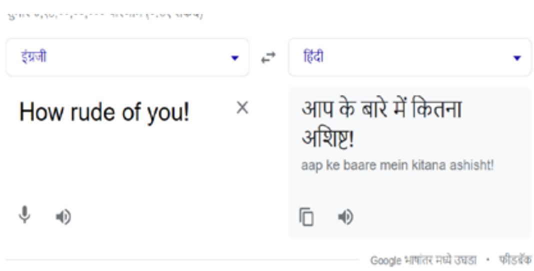

Fig 6 : Google Translation

Here is the comparison (Fig 4 and Fig 6) Between Machine Translation Model and Google Translator. We have compared the same sentence on both translators and the accuracy of google translator is less as compared to the machine translation model.

```
Epoch 100 Batch 0 Loss 0.0054
Epoch 100 Loss 0.0082
Time taken for 1 epoch 7.673804759979248 sec
```

Fig 7 : Accuracy

We trained the model for 100 Epoch's and loss at the 100th Epoch is 0.0082.We can calculate an accuracy as [1-loss].from that an accuracy is computed as 98 percent.

VIII. CONCLUSION AND FUTURE WORK

This report gave detailed information about the project "Natural Language processing Machine translation" .The objective is to develop a model to accurately translate one spoken or written language to another.

In the future, this model can be trained for many different languages at once. We can provide data-sets of different languages. Also ,we will be working on a voice to text module . We can also train the system in such a way that it can generate translation more accurately than before.

REFERENCES

[1] Kalchbrenner and Blunsom, "Recurrent Continuous Translation Models", 2013

[2] Sutskever et al, "Sequence to Sequence Learning with Neural Networks", 2014

[3] Lise volkart,"Statistical vs. Neural Machine Translation", November, 2018

[4] Rafael Prikladnicki,"Speech Recognition for Voice-Based Machine Translation" , January, 2014

[5] Zhixing Tan," Neural machine translation: A review of methods, resources, and tools", May.21,2020

[6]Yang Feng1,"Guiding Teacher Forcing with Seer Forcing
for Neural Machine Translation", August.6, 2021

[7] Singh, M. K, "Mobility and NMT in Sustainable Urban Development-Role of City Developers. Mongolia: Eleventh Regional EST Forum in Asia.", October.2, 2018,.

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova Bert: pre-training of deep bidirectional transformers for language understanding., 2019

[9] Y. Ding, Y. Liu, H. Luan, M. Sun Visualizing and understanding neural machine translation.,2017

[10]Y. Ding, Y. Liu, H. Luan, M. Sun Visualizing and understanding neural machine translation.,2018

[11]S. Edunov, A. Baevski, M. Auli Pre-trained Language Model Representations for Language Generation.,2011

ACKNOWLEDGMENT

We have a great pleasure to express our gratitude to all those who have contributed and motivated us during our project work. For selecting a topic of our choice and studying and understanding it, is a good opportunity to learn and know new technology, understand the way projects are researched. We would like to thank our guide Prof. Anita Morey for his support and guidance, our HOD Dr.Sanjay Shitole, our principal Dr. Shikha Nema, and to all the teachers whose help made us understand our topic well. Thank you all!