English to Bangla Machine Translation with Recurrent Neural Network

|  |  |  |
| --- | --- | --- |
| Shaykh Siddique  Computer Science and Engineering  East West University  Dhaka, Bangladesh  [shaykhsiddiqee@gmail.com](mailto:shaykhsiddiqee@gmail.com) | Rifayet Azam Talukder  Computer Science and Engineering  East West University  Dhaka, Bangladesh  emailaddrs@domain.com | Tahmid Ahmed  Computer Science and Engineering  East West University  Dhaka, Bangladesh  emailaddrs@domain.com |

***Abstract-***

***KeyWords:***

1. INTRODUCTION
2. *Background*

Translation is the communication of meaning from one language to another language. A translation could be word by word translation and another one is translation by sentence. In sentence translation we get more information rather than word by word translation. In this paper our main focus is to translate from English to Bengali by using a machine. Machine translation refers to translation of text or speech. Our task is to translate from English language to Bangla language. With approximately 228 million native speakers and another 37 million as second language speakers. Bengali is the fifth most-spoken native language and the seventh most spoken language by total number of speakers in the world [1].

Natural language processing is used to make the machine intelligent. The way of language processing is enriching day by day. Many studies defined the architecture for natural language processing [2] but few deal with the improvement of English to Bangla language translation. A team works in tense based structure of English to Bangla translation [3]. Another study is on simple sentence structure and comparison of different machine translation systems [4]. Still, now there is a lack of studies with complex sentence structure and recurrent meaning of a sentence.

To make machine intelligent natural language processing is used. Language translation is getting improved but there are a few improvements is getting on English to Bangla translations. There are many ways that translations are solved. Here for Hindi translations, two encoder-decoder neural machine translation architectures are used, which convolutional sequence to sequence model (ConvS2S) and recurrent sequence to sequence model (RNNS2S) [5]. One is for English to Hindi and another is to do the opposite. Here in training data, 1492827 sentences used where 20666365 words for English and 22164816 words for Hindi. The RNNS2S model was trained using Nematus framework and for ConvS2S model was trained using Fairseq-5, an open-source library developed by Facebook for neural machine translation using CNN or RNN networks. Their result showed that ConvS2S performed better on English to Hindi translation which would help to solve our problem. In the Corpus-Based Method using one subject file and one verb file, the translations are solved [4]. Here for each subject, there is a flag corresponded to its verb and the most suitable and meaningful sentences are selected for final translations. The result showed better compared to Google Translator. For another English to Hindi translation feed-forward back-propagation artificial neural network used [6]. For the implementation java used for main programming language to implement the rules and all the modules apart from the neural network model which have been implemented in Matlab. Here, training data is encoded into numeric form by the Encoder which is also implemented in Java. They have used BLEU [7] to calculate the score of the system.

1. *Research Objective*