Paper Title

Machine Translation System Using Deep Learning for English to Urdu

Paper Link

https://www.hindawi.com/journals/cin/2022/7873012/

Summary

The study describes a machine translation system based on deep learning for translating English to Urdu. It employs an LSTM encoder-decoder design with a focus mechanism. A parallel corpus of roughly 30,923 sentences, including religious literature, news, and widely used language, was employed in the investigation. The suggested model received a BLEU score of 45.83 on average, showing its efficacy in machine translation between these languages.

1. Motivation

Machine translation's primary goal is to break down language barriers. The Urdu language, which has previously gotten less attention in machine translation research, was the target of this study. Existing Urdu models focused heavily on statistical methodologies and demonstrated limited performance, as seen by lower BLEU ratings.

1.2 Contribution

The primary contributions of the paper are the building of a news parallel corpus, the development of an English to Urdu machine translation model employing an encoder-decoder with attention, and the evaluation of the model using multiple metrics. This technique addresses the need for more complex machine translation systems for language pairs that are less extensively researched, such as English and Urdu.

1.3 Methodology

Neural machine translation (NMT) was used in the methodology, which is a corpus-based method that maps source and target languages end-to-end. The researchers used an LSTM encoder-decoder architecture with an attention mechanism to effectively parse and convert

normal English text into Urdu. The model was trained and evaluated on a dataset spanning multiple domains.

1.4 Conclusion

The article concludes that neural machine translation, specifically LSTM with the Bahdanau attention mechanism, is an efficient method for translating English to Urdu. The system achieved an average BLEU score of 45.83 after being trained on a corpus of over one million tokens. Future objectives include increasing the corpus to include new domains and incorporating a speech recognition module for speech-to-text translation.

2. Limitations

2.1 The Initial Limitation

The quantity and diversity of the corpus used limit the scope of the study. The model's applicability and accuracy could be improved by expanding the corpus to include other areas.

2.2 Second Restrictions

The current model is limited to text translation. Integrating speech recognition for speech-to-text translation could expand the breadth and utility of the technology.

Synthesis

This study is a big step forward in the field of machine translation, particularly for the English-Urdu language pair. The work highlights the potential of neural machine translation in bridging language barriers in less often studied language pairings by applying deep learning techniques and tackling the specific issues connected with Urdu. The model's successful implementation and high BLEU score demonstrate its usefulness and open the door to additional improvements in multi-domain corpus and speech-to-text translation capabilities.