International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Language Model for Translating the English Language to Garhwali Language

Dr. Pritam Singh Negi

Assistant Professor

Department of Computer Science & Engineering

School of Engineering & Technology

H.N.B. Garhwal University Srinagar (Garhwal), Uttarakhand, INDIA

negipritam@gmail.com

ABSTRACT

The main aim of this paper is to develop a language model for translating the English language to Garhwali, a language spoken in the Garhwal region of Uttarakhand, India. The model will use state-of-the-art natural language processing techniques, including machine learning and neural networks, to enable accurate and efficient translation of English sentences into Garhwali and vice versa. The model will be trained on a large dataset of parallel English-Garhwali text to ensure high accuracy and fluency of the translations. The successful development of this language model will help bridge the language barrier between English speakers and Garhwali speakers, facilitate communication and exchange of ideas, and promote cultural exchange and understanding. Local language conversion system can help to improve the user experience by providing content in the language that the user is most comfortable with. It can help to reduce confusion and frustration, and make the user experience more enjoyable.

Keywords: Statistical Machine Translation, Mathematical model, English-Garhwali, Evaluation.

1. INTRODUCTION

There are two divisions of Uttarakhand State namely Garhwal and Kumaon. In Garhwal division there are seven districts namely Chamoli, Dehradun, Haridwar, Pauri Garhwal, Tehri Garhwal, Rudraprayag and Uttarkashi, whereas six districts namely Almora, Champawat, Bageshwar, Pithoragarh, Udham Singh Nagar and Nainital come under the Kumaon region. There are approximately 325 recognized languages in India, out of which Garhwali language is spoken in the Garhwal division of Uttarakhand. In a study it is found that approximately 2.5 million people speak Garhwali language. The Garhwali language is a central Pahari language that belongs to the Northern Zone of Indo-Aryan and is native to Garhwali. Almost all people who can speak and understand the language can comfortably speak and understand Hindi, the official language of India. Garhwali language is popular among the local villagers and people of cities and towns of Garhwali division, who have been migrated. However, it is also being spoken by people. It is also found that this language has lost the interest by Garhwali people due to large scale migration of population to cities and towns and lack of support from the Government. This problem warrants to develop the methods to conserve the Garhwali Language, so that the kids and the younger generation belonging to this hilly region may be in touch with their language roots. Technological support can play a vital role in the conservation of this language and to generate interest among the local people towards learning the language and explore the meaning of Garhwali words. Translating documents from English to Garhwali is by no means an easy task to accomplish. This is made more apparent if we are not familiar or adept at using this particular language.

Effective and accurate translation of a text from English to Garhwali and vice versa requires a lot of skill, knowledge and understanding. A real English to Garhwali translator should be effective and efficient in translating English to Garhwali text. To be considered a good and reliable translation source, you need to be able to consistently produce accurate and precise translations. This language translation software helps people who want to communicate with people in their local language or communicate with friends and family in their native language. This software helps bridge the language gap by providing an easy way to translate text from one language to another. It also helps people learn new languages by providing a way to practice and understand the language. By providing content in local languages, users can easily understand and interact with the content. This helps increase user engagement and loyalty.

2. OBJECTIVES OF THE PAPER

- To develop a model / translator that translates English test into equivalent Garhwali text.
- To develop a dictionary for providing the meaning to English word in Garhwali and vice versa.
- The work includes to develop a model for a statistical machine translation. SMT uses predictive algorithm to translate test.

3. STEPS INVOLVED FOR DEVELOPING A TRANSLATION SYSTEM

With the help of following steps, we can develop a system for translating the English language to Garhwali

1. Data Collection: Collect a large dataset of parallel English-Garhwali text for training the translation model and develop a database with their parts of speech.
2. Text Preprocessing: Preprocess the text data by removing stop words, punctuation, and special characters, and convert the text into a machine-readable format.
3. Tokenization: Tokenize the preprocessed text data into words and phrases to represent the input and output sequences.
4. Embedding: Map the words and phrases in the input and output sequences into high-dimensional vectors using a pre-trained word embedding model.
5. Model Training: Train a sequence-to-sequence model with an attention mechanism using the preprocessed and embedded text data.
6. Evaluation: Evaluate the performance of the trained model using a validation dataset.
7. Deployment: Deploy the trained model as an application to translate the English and Garhwali languages in real-time.
8. User Interface: Create a user-friendly interface for users to input English or Garhwali text and receive the corresponding translation in the desired language.
9. Continuous Improvement: Continuously improve the translation model by collecting and adding more data in the database, tuning hyperparameters, and optimizing the training process.

4. ALGORITHMFOR PROCESS IN A TRANSLATION SYSTEM

This algorithm is created for translate the English sentence in to the Garhwali Language with the assist of the pattern of comparable kind of group of sentences.

1. Start
2. Input a English sentence of most 128 phrases
3. Save the sentence in to the memory.
4. Break the sentence into its constituent words.
5. Use next module and skip every constituent phrase into it.
   - 5.1. Search for collocation via looking all the viable combos of the phrases.
   - 5.2. If collocation observed then

5.2.1. Take away collocation
5.2.2. Return
5.3 Else
5.3.1. Test consecutive verbs
5.3.2. If observed
5.3.2.1. Substitute a single term in preference to a multiple time period
5.3.2.2. Return
5.3.3 Return
5.4 Return
6. Use Time period seek Module
6.1. Take a look at for presence of every term inside the database.
6.2. If term discovered
6.2.1 Call POS tagging module
6.2.2 Store the term and its associated part of speech
6.2.3 Go back
6.3 Else
6.3.1 Seek the time period against a verb database
6.3.2 If match observed
6.3.2.1 Tag the time period as verb and store the end result
6.3.2.2. Return
6.3.3 Else
6.3.3.1 Tag the term as Noun and shop the result
6.3.3.2 Return
6.4 Use L1 GRAPH module to show the sentence graph.
6.5 Return
7. Use Translation Module
7.1 Calculate the interpretation as according to database and show it.
7.1.1 Display proposed word collection as in step with.
7.1.2 Return.
7.2 Else
7.2.1 Show Translation no longer performed
7.2.2 Ask user to manually input the values in database.
7.2.3 Save the values in database file.
7.2.4 Go back
7.3 Return
8. Return

5. SNAPSHOTSANDDESCRIPTIONOF DIFFERENT MODULEPRESENTINTHEPROPOSED MODEL

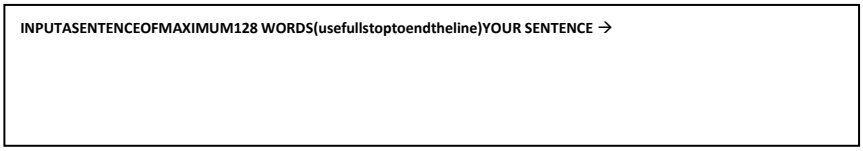

Fig 5.1 Main module screen waiting for user to input an English sentence

In this module user can enter an English sentence in right way. This computes the wide variety of phrases separated by means of areas in the given sentence and calls different sub-modules eventually. It passes the entire sentence to the subsequent sub-module which later tests for collocation and removes if it's found.

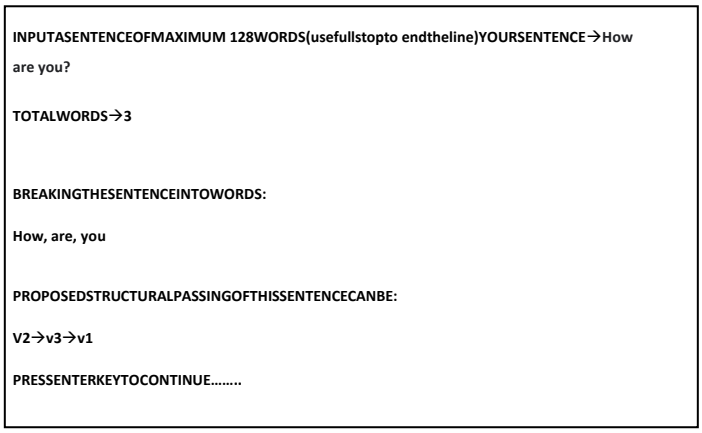

Fig 5.2 Detection of collocation and structural parsing of sentence

It's referred to as by the principal module after a person has entered a sentence finishing with a full forestall. This module accepts the whole sentence as argument and breaks it into consecutive phrases. Then it makes a mixture of each term with each term to check for possible combination of phrases by matching these combos in opposition to a database a collocation database. If any phrases or phrase combos are observed then they're adjusted via concatenating them collectively.

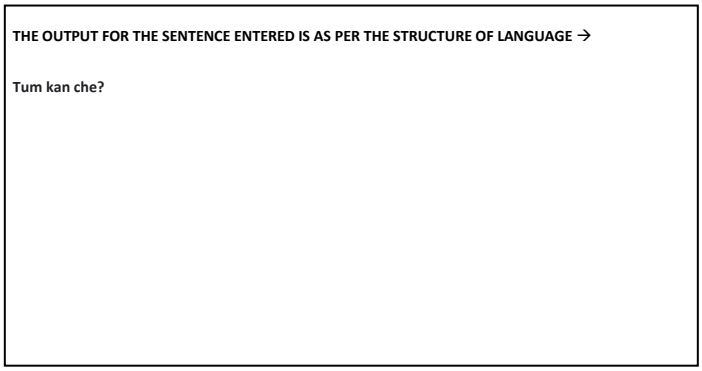

Fig 5.3 Final output windows of the output result

It is known as through the output module via which user can get the output as in keeping with the algorithms. Right here we determined that when we translate an English sentence right into a Garhwali language, then not most effective words are convertedbut the series of phrases is also modified.

The translation will now not take area consistent with the phrase sequence however in truth it'll take in step with the parts of Speech sequence.

CONCLUSION

In the existing system, still we have no such software through which we convert the English language to the Garhwali Language. Here in this paper, we proposed a model through which we inset a text in English language and translates the text in the Garhwali language.

In conclusion, the development of a system for translating the English Language to Garhwali Language and vice versa can greatly facilitate communication and exchange of ideas between English and Garhwali speakers, promoting cultural exchange and understanding. The system employs natural language processing techniques, including machine learning and neural networks, to enable accurate and efficient translation of English sentences into Garhwali and vice versa. The system's success depends on the quality and size of the dataset used for training the translation model, as well as the pre-processing, tokenization, embedding, and training stages of the algorithm. Overall, the development of such a system would contribute to bridging the language barrier between English and Garhwali speakers, promoting inclusivity and cultural exchange.

REFERENCES

[1] Divyapushpalakshmi M, Ramalakshmi R. (2021), An efficient sentimental analysis using hybrid deep learning and optimization technique for Twitter using parts of speech (POS) tagging, *Int J Speech Technol.;*24(2):329–39.

[2] L. Sharma (2018). Entrepreneurial intentions and perceived barriers to entrepreneurship among youth in Uttarakhand state of India: A cross-cultural investigation across genders, *Int. J. Gend. Entrep.*

[3] K. Cho, M. Artetxe, G. Labaka& E. Agirre (2018). Unsupervised Neural Machine Translation. *Published as a conference paper at ICLR.*

[4] L. Jain and P. Agrawal (2017), English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool, *International Journal of Advanced Research in Computer Science*, Volume 8, No. 1, Jan-Feb 2017

[5] Lv C, Liu H, Dong Y, Chen Y. (2016). Corpus based part-of-speech tagging, *Int J Speech Technol.;*19(3):647–54.

[6] K. Hans and R. S. Milton (2016). Improving the performance of neural machine translation involving morphologically rich languages. *arXiv preprint arXiv:1612.02482*

[7] Singh J, Joshi N, Mathur I. (2013). Development of Marathi part of speech tagger using statistical approach, *In: Proc. 2013 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI2013*, pp. 1554–1559.

[8] P. J. Antony (2013), Machine translation approaches and survey for Indian languages, *Int. J. Comput. Linguist. Chinese Language Processing*, 47–78.

[9] U. Kheradia and A. Kondwilkar (2012), Speech To Speech Language Translator, *International Journal of Scientific and Research Publications,* Volume 2, Issue 12, ISSN 2250-3153.

[10] Pathak, Sanket. Ahmad, Rashid. and Ojha, Atul. (2012). A Language Engineering Approach to Enhance the Accuracy of Machine Translation Systems.*ShodhPrerak - Lucknow*, ISSN: 2231 – 413X, pp.205-214.

[11] M. M. S. Rauthan, Pritam Singh Negi and H. S. Dhami (2012) , Pattern Recognition System for Translating the English Sentence into Hindi, *International Journal of Computer Applications*, (0975 – 8887), Volume 39– No.2.

[12] Pritam Singh Negi, M. M. S. Rauthan and H. S. Dhami (2011), Text Summarization for Information Retrieval using Pattern Recognition Techniques, *International Journal of Computer Applications* (0975 – 8887), Volume 21– No.10.

[13] Mannem, P., & Dara, A. (2011). Partial parsing from bitext projections. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 1597-1606). Association for Computational Linguistics.*

[14] Kevin Knight. (1999). Decoding complexity in word-replacement translation models, *Computational Linguistics*.