iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: IV Month of publication: April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59808

Development of Offline Translation Software for English to Hindi

Sonali Gangwar[1], Iram Naim[2]
CSIT, M.J.P.Rohilkhand University, Bareilly, India

Abstract: This article provides a comprehensive study on the development of offline applications for English to Hindi translation. The project leverages the Moses statistical machine translation system to achieve the goal of seamless text-to-text translation between two languages without the need for online APIs. The report covers many aspects of the project, including methods for learning translation techniques and using dynamic programming to calculate the cheapest possible translation methods for foreign languages. Carefully consider the translation by analyzing the quoted words one after the other and estimating their value. Dynamic programming method to speed up visual interpretation. Additionally, the article underlines the problem statement and discusses how recent technological changes have changed translation. Translator software can now translate all documents with a single click at a low cost, a project that eliminates speech barriers and facilitates communication between lines.
Keywords: Attention Mechanisms; Evaluation Metrics; Multilingual Translation; Context-machine translation; machine translation boundary; measurement system; multilingual translation

I. INTRODUCTION

The topic revolves around the development and optimization of information flows based on statistical models of machine translation. The model described by Koehn et al. (2003) is based on a noisy channel model. This turns the translation of a sentence f into English e into argmaxe p(ef) = argmaxe p(fe) p(e) using Bayes' rule. This allows for the standard language p(e) and the standard translation p(fe). Ideas are divided into parts of consecutive words (called sentences). Each sentence is translated into English and the English is returned to the output. The algorithm optimizes the search by ignoring assumptions that may not be part of a good translation. Introducing the concept of comparative situations allows us to identify positive emotions and eliminate emotions different from these. Hypothesis clustering is a risk-free way to reduce the search space. For a better situation, both views can be returned if you accept external backlinks. Most modern methods in the world use puzzles to learn translations. Marcu and Wong (EMNLP, 2002) proposed to build communication lines in a single parallel box.

To study these texts, they put the sentences in a large sample that can generate content and sentences in parallel. In Marcu and Wong's framework, the best expectation of simultaneous learning is (i) the probability distribution Ï(e, f), since sentences e and f have the same interpretations; (ii) and the input d(i, j) represents the probability of changing a flight at location i to a flight at location j. The document titled "Development Process" in the diagram describes this process by showing the original English text, the translation process, and the final Hindi text.

II. LITERATURE REVIEW

The literature review begins with the overview and history of Statistical Machine Translation (SMT). The SMT method is currently the main method in the field and is used in online translations by companies such as Google and Microsoft. In SMT, a translation system is trained on large numbers of parallel documents, from which the system learns to translate small sentences, as well as large numbers of monolingual documents, from which the system learns which language to use. to be. An example is a collection of sentences that are rhyming sentences in two different languages because each sentence in one language corresponds to a sentence translated into the other language. It is also called bit text. The training method in Moses uses parallel data and word combinations of words and parts (called sentences) to determine satisfactory translations of two languages. In sentence-based machine translation, these correspondences only occur between consecutive sentences of a word, whereas in hierarchical sentence-based machine translation, or grammar as translation, more patterns are added to the text. For example, a hierarchical translation machine might learn the truth about German hat X. Most recent methods have used puzzles to learn translation. Marcu and Wong (EMNLP, 2002) proposed creating direct communication lines in the same parallel trunk. To learn such text, they introduce sentences as a common probabilistic model that can simultaneously generate content and target sentences in parallel wholes.

Expecting maximum learning simultaneously in Marcu and Wong's framework (i) Considering the result that sentences e and f are equivalent interpretations, the joint probability distribution Ï(e, f); In case (ii) and d(i, j integration), consider the probability that the sentence at position i will be translated into the sentence at position j.

III. METHODOLOGY

The process revolves around developing and optimising sentences based on statistical machine translation models. Koehn et al. (2003) described the model based on the noise channel model. It uses Bayes' rule to change the possible translation of a foreign sentence f into English e to argmaxe p(ef) = argmaxe p(fe) p(e). This allows for a standard language p(e) and a standard translation p(fe). Ideas are divided into parts of consecutive words (called sentences). Each sentence is translated to English and English is returned in the output. The algorithm optimizes the search by ignoring hypotheses that may not be part of the best interpretation. Introducing the concept of comparative situations, allows one to identify positive emotions and eliminate emotions that fall outside of these emotions. Regrouping assumptions is a risk-free way to reduce the search space. Both views can be replicated if they agree on external backlinks for a better situation. Most of the latest methods use puzzles to learn the translation. Marcu and Wong (EMNLP, 2002) proposed creating direct communication lines in the same parallel trunk. To learn such text, they introduce sentences as a common probabilistic model that can simultaneously generate content and target sentences in parallel wholes. Expecting maximum learning simultaneously in Marcu and Wong's framework (i) Considering the result that sentences e and f are equivalent interpretations, the joint probability distribution Ï(e, f); (ii) and d(i, j integration) show the probability of transforming the sentence at position i into the sentence at position j.

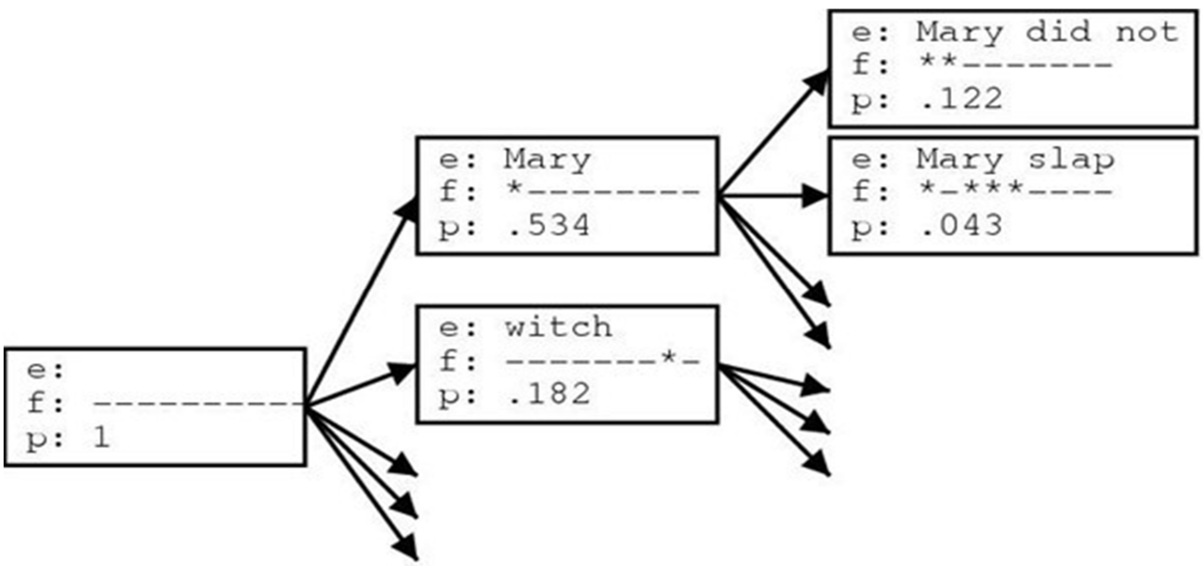


Fig. 1 Process of generating output

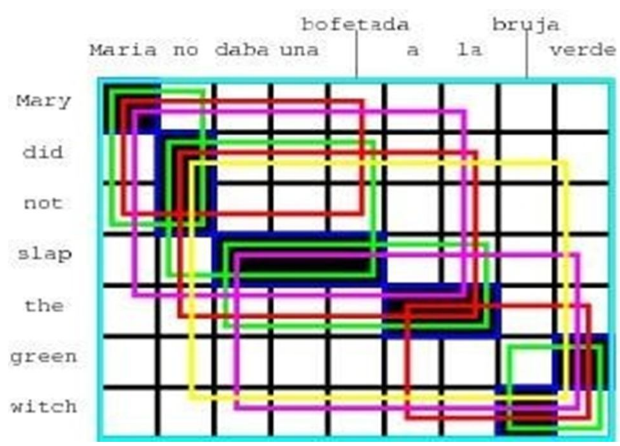


Fig. 2 Phrases Pairs

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IV. RESULTS

The project successfully modelled machine translation based on sentences. The model was tested by translating the text from English to Hindi using the Moses decoder. The command used to start Moses is Moses Decoder-master/bin/Moses-working/train/model/moses.ini. When the word "hello" is entered, it is translated as "âá dââ´là dâoËâ dâ¿â eËHindi. à dâdââ eËGË â in Inline interpretation for multiple, the input is from a file is retrieved and the output is written to another file using the command -master/bin/moses -f job/train/model/moses.ini output .txt using the Musa decoder. This process replaces the original English text and can be exemplified in digital files labelled "Input", "Translation" and "Output" showing the process translation and the latest Hindi words.

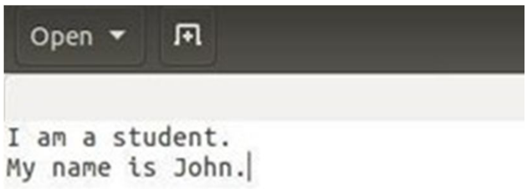


Fig. 3 Input

V. DISCUSSION

This project revolves around developing and optimising sentence-based statistical machine interpretation models. Koehn et al. (2003) described the model based on the noise channel model. It uses Bayes' rule to change the possible translation of a foreign sentence f into English e to argmaxe p(ef) = argmaxe p(fe) p(e). This allows for a standard language p(e) and a standard translation p(fe). The concept is divided into parts of consecutive words (like sentences). Each sentence is translated to English and English is returned in the output. The algorithm optimizes the search by ignoring hypotheses that may not be part of the best interpretation

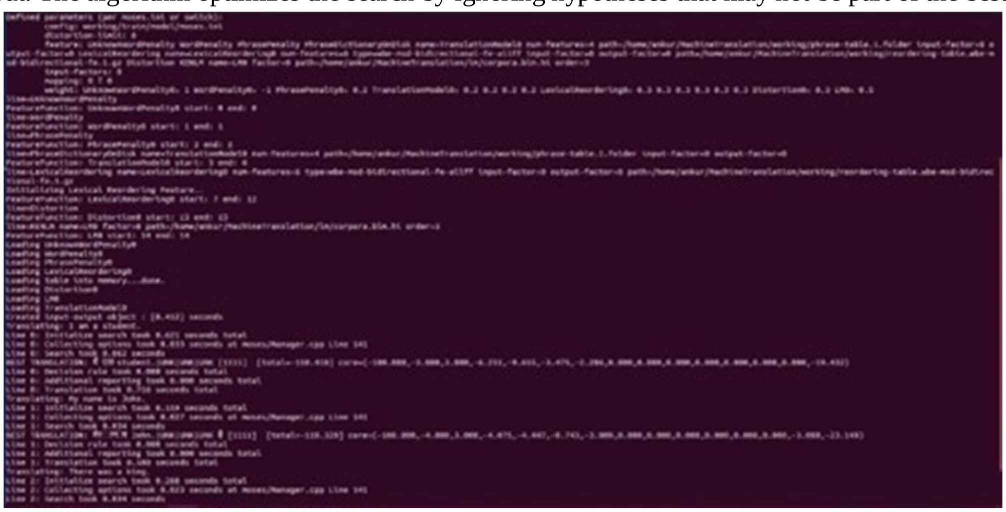


Fig. 4 Translation

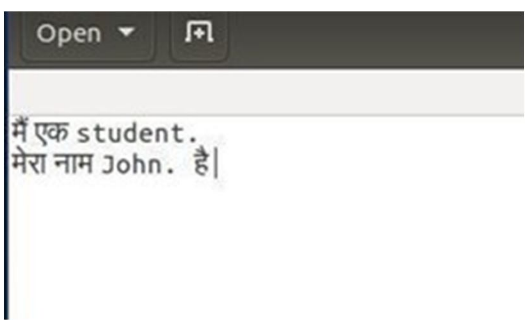


Fig. 5 Output

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Introduces the concept of comparative situations, allows the identification of positive thoughts and eliminates thoughts that fall outside this perspective. Regrouping hypotheses is a risk-free way to reduce the search space. Both views can be replicated if they agree on external backlinks for a better situation. Most recent methods have used single instructions to learn the translation. Marcu and Wong (EMNLP, 2002) proposed creating direct communication lines in the same parallel trunk. They introduce a sentence-based probabilistic association model that can simultaneously generate content and target sentences to learn these texts. Sequential learning in Marcu and Wong's framework (i) generates the joint distribution Ï(e, f), which indicates the probability that sentences e and f are equal language interpretations; The integration of (ii) and d(i,j) shows the probability of translating the sentence at position i into the sentence at position j. The diagram in your document labelled "Production Process" will illustrate this process by showing the original English text, the translation process, and the final Hindi text.

VI. CONCLUSION

This project completed the sentence-based statistical machine translation model. The model was tested by translating the text from English to Hindi using the Musa decoder. The command to start Moses is Moses-decoder-master/bin/moses -working/train/model/moses.ini. When the word "Hello" is entered, it is translated as "âà¤¡à¤®à¤¸à¥à¤¤à¥â" in Hindi. Used to comment out multiple lines, input is taken from one file and output is written to another file using the following command: Moses detector-master /bin/Moses -working /train/model/moses.ini < input En .txt > Output Hi .txt In the. file, the codes that say "Input", "Translation" and "Output" will show this process showing the original English, the translation process and the final output in Hindi respectively. This project demonstrates the feasibility and effectiveness of sentence-based statistical machine translation. It shows that accurate and efficient translation of words is possible with tools and techniques. This work contributes to the widespread use of machine translation and has the potential to influence the development of interactive communication. Finally, the team expressed their gratitude to the students and thanked the parents for their morale and support.

REFERENCES

[1] Statistical Machine Translation. (n.d.). Retrieved from https://www.sciencedirect.com/science/article/pii/S187705091630 0710

[2] Moses: A Statistical Machine Translation System. (n.d.). Retrieved from http://www.statmt.org/moses/manual/manual.pdf

[3] Hagiwara, M. (n.d.). Statistical Machine Translation: A Comprehensive Guide. Retrieved from https://ieeexplore.ieee.org /document/8122149

[4] Association for Computational Linguistics (ACL). (n.d.). Retrieved from https://www.aclweb.org/

[5] Koehn, P. (2017). Neural Machine Translation. Computer Science Review, 25, 121–128. doi: 10.1016/j.cosrev.2017.09.001

[6] Koehn, P., & Knowles, R. (2017). Six Challenges for Neural Machine Translation. 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 5675–5679). doi: 10.1109/ICASSP.2017.7953343

[7] Koehn, P. (2010). Moses: An Open Source Toolkit for Statistical Machine Translation. Retrieved from http://www.statmt. org/Moses/?n=Moses.Background

[8] Moses: Background. (n.d.). Retrieved from http://www.statmt.org/moses/?n=Moses.Background

[9] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[10] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

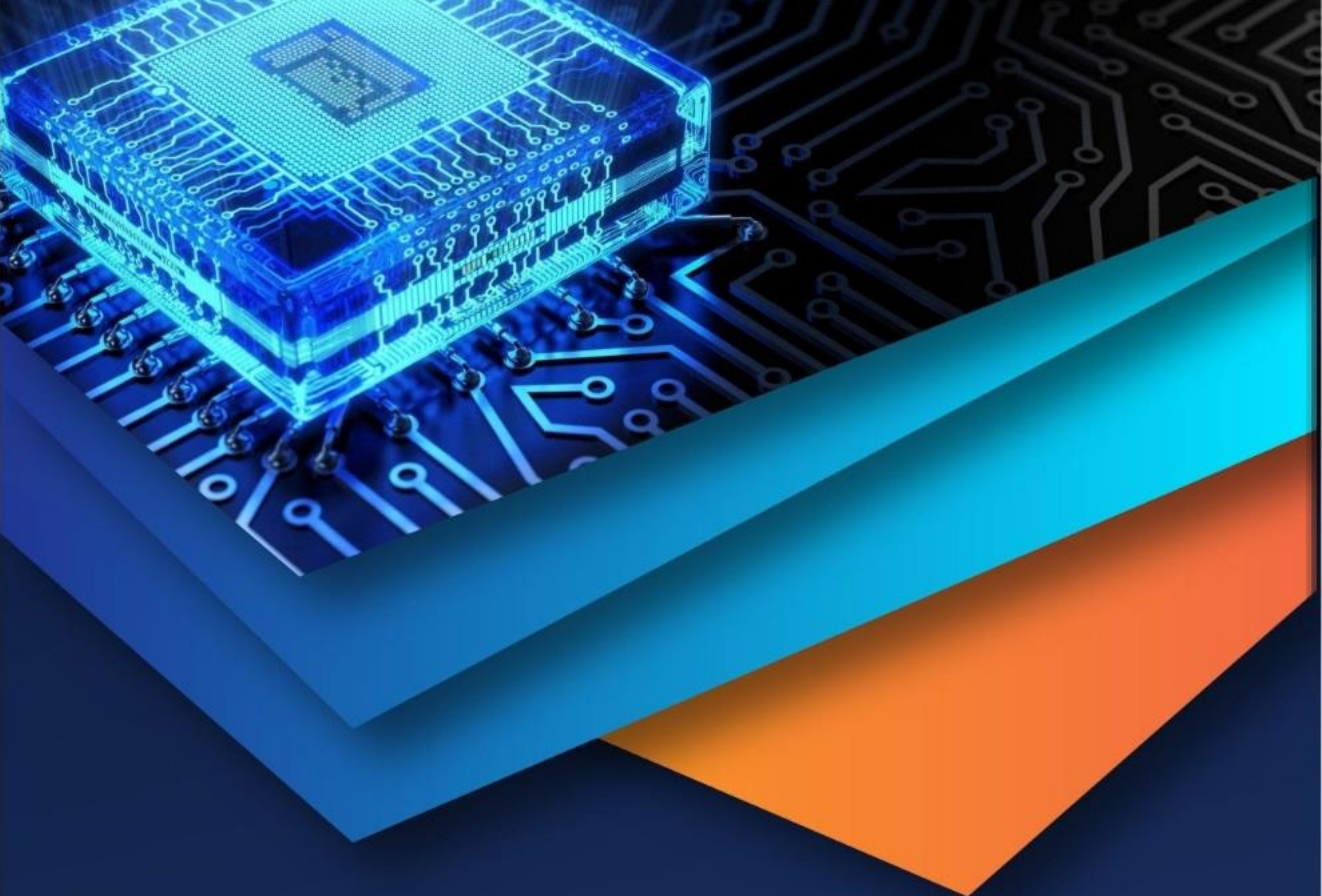
[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[13] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on Association for computational linguistics (pp. 311-318).

[14] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Dean, J. (2016). Google's multilingual neural machine translation system: Enabling zero-shot translation. arXiv preprint arXiv:1611.04558.

[15] Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1412-1421).



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY