

San-Eng: Sanskrit to English Translator using Machine Learning

Shetty Ramakrishna Mohan, Rohan S Bhat, Ranjith V Shetty, Aniruddha

Department of Computer Science and Engineering,
Canara Engineering College, Benjanapadavu, India

Abstract: *In the past decade, machine learning has made great strides in improving automatic translation. In general, machine learning algorithms have been able to achieve better translation quality by learning from large amounts of data. One approach to machine translation is to use a statistical machine translation (SMT) system. SMT systems learn to translate by statistical methods, using a large parallel corpus of text in multiple languages. The most successful SMT systems are based on the translation model known as the phrase-based translation model. In recent years, a new approach to machine translation has emerged, known as neural machine translation (NMT). NMT systems use artificial neural networks to learn to translate. NMT systems have shown to be very successful in translating between languages that are closely related, such as English and Hindi or Reverse is also true. In recent years, a new approach to machine translation has emerged, known as neural machine translation (NMT). NMT systems use artificial neural networks to learn to translate. NMT systems have shown to be very successful in translating between languages that are closely related, such as English and Hindi or Reverse is also true.*

Keywords: Machine Learning, Statistical machine translation, Neural machine translation, Encoder-Decoder, Optical Character Recognition

REFERENCES

- [1]. S. Kondo, "Machine Translation with Pre-specified Target-side Words Using a Semi-autoregressive Model," no. 2019, pp. 68–73, 2021.
- [2]. R. K. Chakrawarti, J. Bansal, and P. Bansal, "Machine translation model for effective translation of Hindi poetries into English," J. Exp. Theor. Artif. Intell, vol. 34, no. 1, pp. 95–109, 2022, doi: 10.1080/0952813X.2020.1836033.
- [3]. A. Pathak and P. Pakray, "Neural machine translation for Indian languages," J. Intell. Syst., vol. 28, no. 3, pp. 465–477, 2019, doi: 10.1515/jisys-2018-0065.
- [4]. B. Premjith, M. A. Kumar, and K. P. Soman, "Neural machine translation system for English to Indian language translation using MTIL parallel corpus," J. Intell. Syst., vol. 28, no. 3, pp. 387–398, 2019, doi: 10.1515/jisys-2019-2510.
- [5]. Sitender and S. Bawa, "Sanskrit to universal networking language EnConverter system based on deep learning and context-free grammar," Multimed. Syst., 2020, doi: 10.1007/s00530-020-00692-3.
- [6]. Rivera-Trigueros, "Machine translation systems and quality assessment: a systematic review," Lang. Resour. Eval., vol. 56, no. 2, pp. 593–619, 2022, doi: 10.1007/s10579-021-09537-5
- [7]. O. Hellwig, S. Sellmer, and S. Nehrlich, "Obtaining more expressive corpus distributions for standardized ancient languages," CEUR Workshop Proc., vol. 2989, pp. 92–107, 2021. A.
- [8]. O. Hellwig, S. Scarlata, E. Ackermann, and P. Widmer, "The treebank of Vedic Sanskrit," Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc., pp. 5137– 5146, 2020.
- [9]. O. Hellwig, "Dating and Stratifying a Historical Corpus with a Bayesian Mixture Model," Proc. LT4HALA 2020 - 1st Work. Lang. Technol. Hist. Anc. Lang., no. May, pp. 1–9, 2020, [Online]. Available: <https://aclanthology.org/2020.lt4hala-1.1>
- [10]. O. Hellwig and S. Sellmer, "Detecting Diachronic Syntactic Developments in Presence of Bias Terms," Proc. Second Work. Lang. Technol. Hist. Anc. Lang., no. June, pp. 10–19, 2022, [Online]. Available: <https://aclanthology.org/2022.lt4hala1.2>
- [11]. P. Dhar, A. Bisazza, and G. Van Noord, "Optimal Word Segmentation for," no. 2012, pp. 181–190, 2021.

- [12]. M. Singh, R. Kumar, and I. Chana, "Machine Translation Systems for Indian Languages: Review of Modelling Techniques, Challenges, Open Issues and Future Research Directions," Arch. Comput. Methods Eng., vol. 28, no. 4, pp. 2165–2193, 2021, doi: 10.1007/s11831-020-09449-7.
- [13]. O. Hellwig, "Dating Sanskrit texts using linguistic features and neural networks," Indogermanische Forschungen, vol. 124, no. 1, pp. 1–46, 2019, doi: 10.1515/if2019-0001.
- [14]. R. Haque, M. Hasanuzzaman, and A. Way, "Investigating terminology translation in statistical and neural machine translation: A case study on English-to-Hindi and Hindi-to-English," Int. Conf. Recent Adv. Nat. Lang. Process. RANLP, vol. 2019- Septe, no. 2017, pp. 437–446, 2019, doi: 10.26615/978-954-452-056-4_052
- [15]. S. R. Laskar, A. F. U. R. Khilji, D. Kaushik, P. Pakray, and S. Bandyopadhyay, "Improved English to Hindi Multimodal Neural Machine Translation," WAT 2021 - 8th Work. Asian Transl. Proc. Work., pp. 155–160, 2021, doi: 10.18653/v1/2021.wat-1.17.
- [16]. O. Hellwig and S. Nehrlich, "Sanskrit word segmentation using character-level recurrent and convolutional neural networks," Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018, pp. 2754–2763, 2018, doi: 10.18653/v1/d18-1295.
- [17]. W. Lu, "Word sense disambiguation based on dependency constraint knowledge," Cluster Comput., vol. 22, pp. 7549–7557, 2019, doi: 10.1007/s10586-018-1899-3.
- [18]. A. R. Pal, D. Saha, N. S. Dash, and A. Pal, "Word Sense Disambiguation in Bangla Language Using Supervised Methodology with Necessary Modifications," J. Inst. Eng. Ser. B, vol. 99, no. 5, pp. 519–526, 2018, doi: 10.1007/s40031-018-0337-5.
- [19]. S. R. Laskar, A. F. U. R. Khilji, D. Kaushik, P. Pakray, and S. Bandyopadhyay, "Multimodal Neural Machine Translation for English to Hind," WAT 2021 - 8th Work. Asian Transl. Proc. Work., pp. 109–113 2021, Available: <https://aclanthology.org/2020.wat-1.11>
- [20]. W. Lu, "The Vedic corpus as a graph. An updated version of Bloomfield's Vedic Concordance," Cluster Comput., vol. 22, pp. 754–775, 2019, doi: 10.1009/s10556-018-1899-3