A Hybrid Translation Model for Pidgin English to English Language Translation

Data, Engineering and Applications

- Saviour Oluwatomiyin,

- Sanjay Misra,

- John Wejin,

- Akshat Agrawal &

- Jonathan Oluranti

- **471** Accesses

Abstract

The African continent is made up of people with rich diverse cultures and spoken languages. Despite the diversity, one common point of unification, especially among the West African communities is the spoken pidgin-English language. With the development in web technology and the English language

dominancy of web content, this growing population stands disadvantaged in understanding content on the web. To proffer a solution, researchers in machine translation from Pidgin English to the English language have leveraged only unsupervised and supervised Neural Machine Translation-based models. In this paper, we propose a hybrid-strategic model that improves the accuracy of the baseline Neural Machine Translation Model (NMT) in translating pidgin English to the English language. From the JW300 public dataset, we used 22,047 sentence pairs for training our model,1000 for tuning, and 2520 for testing. The Bi-Lingual Evaluation Understudy (BLEU) score was employed as a metric of measurement. From our findings, our hybrid model outperforms the baseline NMT model with a BLEU score of 1.05 on two-level translation. This indicates that the accuracy is dependent on the level and type of hybrid used. Studies that look at in-depth pre-translation strategies for developing translation machine model are green areas for pidgin-English translation.

This is a preview of subscription content, <u>log in via an institution</u> to check access.

Similar content being viewed by others
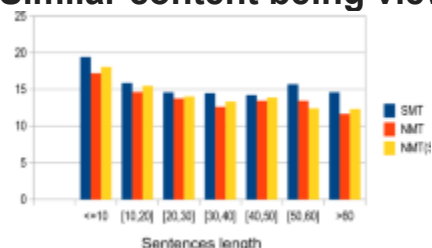
An empirical analysis on statistical and neural machine translation system for English to Mizo language

Article 13 September 2023

Improving translation between English, Assamese bilingual pair with monolingual data, length penalty and model averaging

Article 30 January 2024

English to Hindi Machine Translation Using Sub-classed Model

Chapter © 2022
References

1. Bbc starts pidgin digital service for west Africa audiences (2017). https://www.bbc.com/news/world-africa-40975399

2. Statista. Mobile internet user penetration in Nigeria from 2015 to 2025". Available: Nigeria mobile internet user penetration 2025 | Statista. Accessed 15 March 2021

   Google Scholar

3. Lopez A (2007) A survey of statistical machine translation. https://doi.org/10.21236/ada466330

4. Brown PF et al (1990) A statistical approach to machine translation. Comput Linguist 16(2):79–85

   Google Scholar

5. Marcu D, Wong W (2002) A phrase-based, joint probability model for statistical machine Translation Daniel Marcu. https://doi.org/10.3115/1118693.1118711

6. Koehn P, Och FJ, Marcu D (2003) Statistical phrase-based translation. In: Proceedings of the 2003 human language technology conference of the north american chapter of the association for computational linguistics. pp 127–133. https://www.aclweb.org/anthology/N03-1017. Accessed 29 Apr 2020

7. Artetxe M, Labaka G, Agirre E (2020) Unsupervised statistical machine translation. ArXiv180901272 *Cs*. http://arxiv.org/abs/1809.01272. Accessed 28 Apr 2020

8. Karan S (2015) Methods for leveraging lexical information in SMT. M.S Thesis, Comp Sci, IIIT, Hyderabad, India. https://www.researchgate.net/publication/279181014_Methods_for_Leveraing_Lexical_Information_in_SMT. Accessed 15 March 2021

9. Ravi S, Knight K (2011) Deciphering Foreign Language. In: Proceedings of the 49th Annual Meeting of the Association for Computational

Linguistics: Human Language Technologies, Portland, Oregon, USA, pp 12–21. https://www.aclweb.org/anthology/P11-1002

10.   Sánchez-Martínez F, Forcada ML (2009) Inferring shallow-transfer machine translation rules from small parallel corpora. J Artif Intell Res 34:605–635. https://doi.org/10.1613/jair.2735

Article **MATH** **Google Scholar**

11.   Mikolov T, Le QV, Sutskever I (2020) Exploiting similarities among languages for machine translation. ArXiv13094168 Cs. http://arxiv.org/abs/1309.4168. Accessed: 16 May 2020

12.   Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. ArXiv E-Prints, ArXiv:1409.3215

Google Scholar

13.   Cho K, van Merrienboer B, Gülçehre C, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. *CoRR* abs/1406.1078. http://arxiv.org/abs/1406.1078

14.   Zhou L, Zhang J, Zong C (2018) Look-ahead attention for generation in neural machine translation. In: Huang X, Jiang J, Zhao D, Feng Y, Hong Y (eds) Natural Language Processing and Chinese Computing. NLPCC 2017. Lect Notes Comput Sci 10619. Springer, Cham. https://doi.org/10.1007/978-3-319-73618-1_18

15.   Ogueji K, Ahia O (2019) PidginUNMT: Unsupervised Neural Machine Translation from West African Pidgin to English

Google Scholar

16.   Lample G, Ott M, Conneau A, Denoyer L, Ranzato M (2018) Phrase-based & neural unsupervised machine translation. In: Proceedings of the 2018 conference on empirical methods in natural language processing, Brussels, Belgium. pp 5039–5049. https://doi.org/10.18653/v1/D18-1549

17.   Ahia O, Ogueji K (2020) Towards supervised and unsupervised neural machine translation baselines for nigerian pidgin. ArXiv200312660 Cs. http://arxiv.org/abs/2003.12660. Accessed 12 May 2020

18.   Vaswani A et al. (2017) Attention Is All You Need. ArXiv170603762 Cs. http://arxiv.org/abs/1706.03762. Accessed 24 Apr 2020

19.   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 of the Association for Computational Linguistics, Philadelphia, Pennsylvania, USA. pp 311–318. https://doi.org/10.3115/1073083.1073135

20.   Lavie A, Sagae K, Jayaraman S (2004) The significance of recall in automatic metrics for MT evaluation, vol. 3265, pp. 134–143. https://doi.org/10.1007/978-3-540-30194-3_16

21.   Lavie A, Agarwal A (2020) METEOR: an automatic metric for MT evaluation with high levels of correlation with human judgments. In: Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic. pp 228–231. https://www.aclweb.org/anthology/W07-0734. Accessed 24 July 2020

22.   Agić Z, Vulić I (2019) JW300: A wide-coverage parallel corpus for low-resource languages. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. pp 3204–3210. https://doi.org/10.18653/v1/P19-1310

23.   Adubi SA, Misra S (2016) Syllable-based text compression: a language case study. Arab J Sci Eng (Springer) 41(8):3089–3097

Article **Google Scholar**

24.   Ojumah S, Misra S, Adewumi A (2017). A database for handwritten yoruba characters. In: International Conference on Recent Developments in Science, Engineering and Technology. Springer, Singapore, pp 107–115

Google Scholar

25.   Sharma I, Anand S, Goyal R, Misra S (2017). Representing contexual relations with sanskrit word embeddings. In: International conference on computational science and its applications. Springer, Cham, pp 262–273

Google Scholar

26.   Akman A, Bayindir H, Ozleme S, Akin Z, Misra S (2011) A lossless text compression technique using syllable-based morphology. Int J Inf Technol 8(1):1–9

[Google Scholar](#)

[Download references](#)

Author information

Authors and Affiliations

1. **Center of ICT/ICE Research, Covenant University, Ota, Ogun State, Nigeria**
   Saviour Oluwatomiyin, John Wejin & Jonathan Oluranti
2. **Department of Computer Science and Communication, Østfold University College, Halden, Norway**
   Sanjay Misra
3. **Amity University Haryana, Gurgaon, India**
   Akshat Agrawal

Corresponding author

Correspondence to [Akshat Agrawal](#) .

Editor information

Editors and Affiliations

1. **School of Information Technology, Rajiv Gandhi Technical University, Bhopal, Madhya Pradesh, India**
   Sanjeev Sharma
2. **Department of Creative Technologies and Product Design, National Taipei University of Business, Taiwan, Taiwan**
   Sheng-Lung Peng
3. **Department of Computer Science and Engineering, Rajiv Gandhi Technical University, Bhopal, Madhya Pradesh, India**
   Jitendra Agrawal
4. **Department of Computer Science and Engineering, Oriental Institute of Science and Technology, Bhopal, Madhya Pradesh, India**
   Rajesh K. Shukla

5. **Department of Information Technology, Haiphong University, Haiphong, Vietnam**

Dac-Nhuong Le

Rights and permissions

Reprints and permissions

Copyright information

About this paper

Cite this paper

Oluwatomiyin, S., Misra, S., Wejin, J., Agrawal, A., Oluranti, J. (2022). A Hybrid Translation Model for Pidgin English to English Language Translation. In: Sharma, S., Peng, SL., Agrawal, J., Shukla, R.K., Le, DN. (eds) Data, Engineering and Applications. Lecture Notes in Electrical Engineering, vol 907. Springer, Singapore. https://doi.org/10.1007/978-981-19-4687-5_29

Download citation

- .RIS
- .ENW
- .BIB

Publish with us

Policies and ethics

Access this chapter

Chapter

EUR 29.95
Price includes VAT (Nigeria)

- Available as PDF
- Read on any device
- Instant download
- Own it forever

Buy Chapter

eBook

EUR 117.69

Softcover Book

EUR 149.99

Hardcover Book

EUR 149.99

Tax calculation will be finalised at checkout

Purchases are for personal use only

Institutional subscriptions

- Sections
- References