Investigation of English to Hindi Multimodal Neural Machine Translation using Transliteration-based Phrase Pairs Augmentation

Sahinur Rahman Laskar[1], **Rahul Singh**[1], **Md Faizal Karim**[1]
Riyanka Manna[2], **Partha Pakray**[1], **Sivaji Bandyopadhyay**[1]
[1]Department of Computer Science and Engineering, National Institute of Technology, Silchar, India
[2]Department of Computer Science and Engineering, Adamas University, Kolkata, India
{sahinurlaskar.nits, rahuljan, faizal.karim.metro}@gmail.com
{riyankamanna16, parthapakray, sivaji.cse.ju}@gmail.com

Abstract

Machine translation translates one natural language to another, a well-defined natural language processing task. Neural machine translation (NMT) is a widely accepted machine translation approach, but it requires a sufficient amount of training data, which is a challenging issue for low-resource pair translation. Moreover, the multimodal concept utilizes text and visual features to improve low-resource pair translation. WAT2022 (Workshop on Asian Translation 2022) organizes (hosted by the COLING 2022) English to Hindi multimodal translation task where we have participated as a team named CNLP-NITS-PP in two tracks: 1) text-only and 2) multimodal translation. Herein, we have proposed a transliteration-based phrase pairs augmentation approach, which shows improvement in the multimodal translation task. We have attained the second best results on the challenge test set for English to Hindi multimodal translation with BLEU score of 39.30, and a RIBES score of 0.791468.

1 Introduction

The multimodal NMT (MNMT) concept aims to include different input modalities, such as images in addition to text and attempts to improve low-resource pair translation by merging visual features in addition to textual features (Shah et al., 2016). The attention-based encoder-decoder architecture for NMT handles various issues of long-term dependency and variable-length phrases via sequence-to-sequence learning and attains a state-of-the-art technique of machine translation (MT) (Bahdanau et al., 2015; Luong et al., 2015). Also, NMT shows remarkable performance for low-resource Indian languages (Pathak and Pakray, 2018; Pathak et al., 2018; Laskar et al., 2019a,b, 2020a, 2021c,b). Further, to handle the data scarcity problem, the authors (Sen et al., 2020) augmenting phrase pairs and the source language transliteration-based (Laskar et al., 2022) approach to enhance text-only based for low-resource pair translation. This paper aims to investigate the English to Hindi multimodal translation task in WAT2022 with a proposed transliteration-based phrase pairs augmentation approach (as discussed in 3.2).

The rest of the paper is organized as follows: Section 2 presents the review of related works. The system description is briefly discussed in Section 3. Section 4 reports the results and Section 5 concludes the paper with future scope.

2 Related Works

The literature survey explores existing works on MNMT for English-Hindi language pair (Dutta Chowdhury et al., 2018; Sanayai Meetei et al., 2019; Laskar et al., 2019c, 2021a). We participated in WAT2020 on multimodal translation task for English to Hindi translation and attained the best results with a BLEU score of 33.57 on the challenge test set(Laskar et al., 2020b) using RNN-based MNMT model (Calixto and Liu, 2017; Calixto et al., 2017) and taking advantage of pre-trained word embeddings of the monolingual corpus. Later, we improved the results in WAT2021 (Laskar et al., 2021a) using phrase pairs augmentation. In this work, we have investigated a proposed transliteration-based phrase pairs augmentation approach to enhance the multimodal translational performance of English to Hindi.

3 System Description

The experiments are carried out in four operations, namely, transliteration-based phrase pairs augmentation, data preprocessing, model training, and testing. The OpenNMT-py (Klein et al., 2017) tool is utilized to build multimodal and text-only models independently. The difference between our previous work (Laskar et al., 2020b) and this work is that the current work uses transliteration-based phrase pairs augmentation.

Proceedings of the 9th Workshop on Asian Translation, pages 117–122
October 17, 2022.

3.1 Dataset Description

The dataset namely, Hindi Visual Genome 1.1[1] (Parida and Bojar, 2020) is used in the multimodal translation task of English-to-Hindi, which is provided by WAT2022 organizer (Nakazawa et al., 2022). In this dataset, duplicates (text and image) are present in the train set (Laskar et al., 2020b), which have image ID numbers 2328549, 2391240, and 2385507. Therefore, we have removed those duplicates and thus train set contains 28,927 images and the same number of corresponding English-Hindi parallel sentences. The validation, test (evaluation and challenge) set contains 998, 1,595, and 1,400 images and parallel sentences.

3.2 Transliteration-based Phrase Pairs Augmentation

In this operation, the English-Hindi parallel train set is first used to extract source-target phrase pairs, which are then added to the train set. (Sen et al., 2020) used SMT-based phrase pairs to enrich training data in order to enhance low-resource pair translation. To extract phrase pairs (Laskar et al., 2021a), we have used Giza++ (Och and Ney, 2003) following (Sen et al., 2020). Duplicates and blank lines are eliminated before adding to the parallel train set. Table 1 presents the statistics for the phrase pairs that were extracted. Table 2 presents the data statistics for the train set (before and after augmentation). Afterwards, the sentences from English sources are transliterated into Hindi script using indic-trans [2] (Bhat et al., 2014). The transliteration strategy aims to enable lexical sharing at the sub-word level between source and target sentences that will take place during training. The sample overlaps words (bold marks) of transliterated En and Hi tokens of the training set, are presented in Figure 1.

3.3 Data Preprocessing, System Training, and Testing

The pre-trained CNN-VGG19[3] is used to extract the image/visual features from the image data. Unlike (Laskar et al., 2020b, 2021a), we have considered the co-ordinate or bounded box region information (X, Y, width, height) of the images as the visual features, which are available in the Hindi Visual Genome 1.1 (Parida and Bojar, 2020). Moreover, we have augmented image features of extracted phrase pairs. To select relevant images of the corresponding phrase pairs, we have searched each phrase in the original parallel corpus, if it is found then the corresponding image and its co-ordinate information are considered. But there is a problem if multiple sentences contain the same phrase subset. To handle this issue, a filtering step solution is considered.

Image ID	En	En Transliterated Hi	Hi
17	**Computer** screens turned on	**कंप्यूटर** स्क्रीन्स टर्न्ड ऑन	**कंप्यूटर** स्क्रीन चालू
21	**photo album** open on an adult's lap	**फोटो एल्बम** ओपन ऑन अन अडुल्ट'स लैप	एक वयस्क की गोद में **फोटो एल्बम** खुला
23	there is a group of girls beside the black **car**	तेरे इस आ ग्रुप ऑफ गर्ल्स बेसाइड थे ब्लैक **कार**	काली **कार** के पास लड़कियों का एक समूह है
80	the **cabinet** is wood	थे **कैबिनेट** इस वुड	**कैबिनेट** लकड़ी है

Figure 1: Sample overlap words (bold marks) in the transliterated source (En) and target (Hi) sentence (train set).

- First, for every En-Hi phrase pair extracted from the corpus, we found the matching English segments from the corpus which have the English part of the phrase pair as a sub-string (filter-1).
- If the length of the resulting data-frame i.e., the number of matching English segments for the English part of the phrase is 0, then the phrase is skipped as it is invalid. If the length is 1, since only one English segment matches it, that segment is directly selected.
- On the other hand, if the length is more than 1 i.e. more than 1 English segments have the English phrase as sub-string, the resulting English segments are again filtered (filter-2) to check if the corresponding Hindi phrase of the phrase pairs also has subset in the Hindi segments.
  - If after filter-2, the result is 0, i.e., there are no matching Hindi segments that have the Hindi phrase as sub-string, then from the filter-1 data-frame, i.e. the final segment from matching English segments is randomly selected.

[1] https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-3267

[2] https://github.com/libindic/indic-trans

[3] https://github.com/iacercalixto/MultimodalNMT

Number of Phrase Pairs	Tokens	
	En	Hi
158,131	392,966	410,696

Table 1: Data Statistics of extracted phrase pairs.

Train Set	Number of Parallel Sentence/Segments
Before Augmentation	28,927
After Augmentation	187058

Table 2: Data Statistics of train set (before and after augmentation).

- If the number of matches after Hindi segment matching is 1, then that single segment is selected.
- If the number of Hindi phrase matches is more than 1, then a matching segment is randomly selected with a seed value.

The OpenNMT-py toolkit has been used for text data tokenization, preprocessing, and conducting independent training sessions for text-only and multimodal NMT. We have followed the default settings of (Calixto and Liu, 2017; Calixto et al., 2017) and employed the bidirectional RNN (BRNN) at the encoder and doubly-attentive RNN at decoder during the training process of multimodal NMT. We have used a batch size of 32, a dropout value of 0.3, and an Adam optimizer with 0.002 learning rate during the training process. We have trained on a single GPU with early stopping criteria i.e., the model training is halted if does not converge on the validation set for more than 10 epochs. The obtained optimum trained models of multimodal and text-only NMT were applied to the evaluation and challenge test set. The basic difference in the testing phase is that multimodal NMT uses visual features of image test data. The source English sentences of test data are transliterated and then applied to the trained model to generate the predicted target Hindi sentences.

4 Result and Analysis

The WAT2022 shared task organizer (Nakazawa et al., 2022) published the evaluation result[5] of the multimodal translation task for English to Hindi. We participated with the team name CNLP-NITS-PP in the multimodal and text-only submission tracks of the same task where four teams participated. The automatic evaluation metrics, BLEU

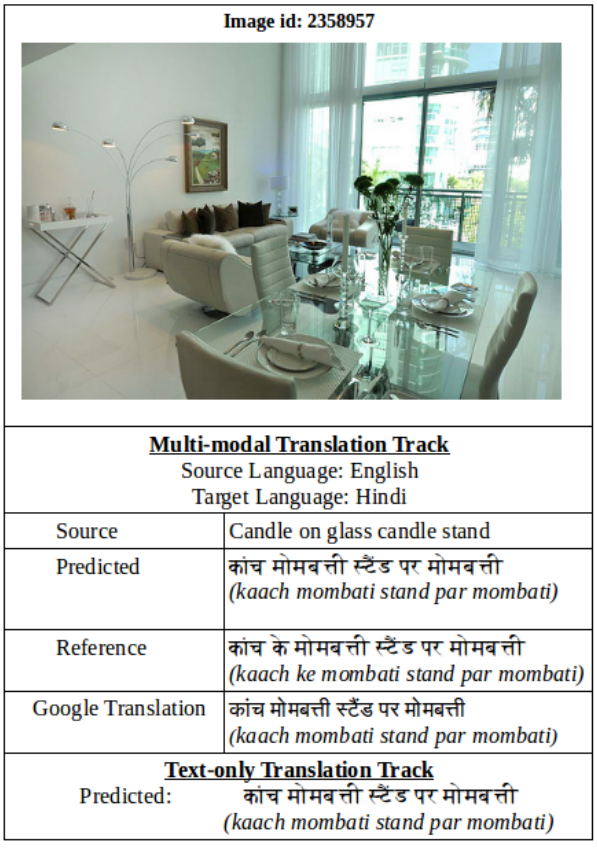

Figure 2: Sample predicted output on challenge test data.

[5] http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

Our System	Test Set	BLEU	RIBES
Text-only NMT	Challenge	37.20	0.770640
	Evaluation	37.00	0.795302
Multi-modal NMT	Challenge	39.30	0.791468
	Evaluation	39.40	0.802635

Table 3: Our system's results (official) on English to Hindi multimodal translation Task.

Other System	Test Set	BLEU	RIBES
Multi-modal NMT (Team: Volta (First Position))	Challenge	51.60	0.859645
Multi-modal NMT (Team: Organizer)	Challenge	20.34	0.644230

Table 4: Other system's results (official)[4] on English to Hindi multimodal translation Task.

Image ID	MNMT Output (Without Transliterated)	MNMT Output (With Transliterated)	Source	Reference
2407547	**अदालत** में दो खिलाड़ी हैं	**कोर्ट** में दो खिलाड़ी हैं	there are two players in the **court**	**क़ोर्ट** में दो खिलाड़ी हैं
2368444	पेड़ों का एक <**unk**>	पुस्तकों का एक **स्टैंड**	A **stand** of trees	पेड़ों का एक **झाड़**
2402752	गहरे लकड़ी की <**unk**> **स्टैंड**	गहरे लकड़ी के **अंडे स्टैंड**	dark wooden wash **stand**	ग़ाढे रंग की लकड़ी का वाश **स्टैंड**

Figure 3: Manual comparison of MNMT output (without transliteration and with transliteration).

(Papineni et al., 2002), RIBES (Isozaki et al., 2010) were used for evaluation of results. Table 3 and 4 reported the official results of our and other systems (Team: Volta (first position) and Organizer). Our team and another team (Volta) achieved the second and first position in multimodal submission for the challenge test set. The quantitative results show that the MNMT outperforms text-only NMT due to the use of visual and textual features. Furthermore, our system's results have improved compared to our previous work on the same task (Laskar et al., 2020b). The BLEU, RIBES scores of the present work show (+5.73, +0.037327), increments on the challenge test set for MNMT, where it is realized that transliteration-based phrase pairs augmentation improves translational performance. The sample examples of outputs, along with Google translation and transliteration of Hindi words, are presented in Figure 2. Moreover, Figure 3 presents a manual comparison of MNMT predicted outputs where we have considered with or without transliteration in the phrase pairs augmentation model.

5 Conclusion and Future Work

In this work, we have proposed the use of transliteration-based phrase pairs augmentation in the WAT2022 multimodal translation task for English to Hindi translation. Our multimodal NMT attained a higher score than that of the text-only NMT model and existing work of (Laskar et al., 2020b). A multilingual-based approach will be investigated to improve the translational performance of low-resource multimodal NMT.

Acknowledgements

We want to thank the Department of Computer Science and Engineering, Center for Natural Language Processing (CNLP), Artificial Intelligence (AI) Lab at the National Institute of Technology, Silchar for providing the requisite support and infrastructure to execute this work.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Irshad Ahmad Bhat, Vandan Mujadia, Aniruddha Tammewar, Riyaz Ahmad Bhat, and Manish Shrivastava. 2014. Iiit-h system submission for fire2014 shared task on transliterated search. In *Proceedings of the Forum for Information Retrieval Evaluation*, FIRE '14, page 48–53, New York, NY, USA. Association for Computing Machinery.

Iacer Calixto and Qun Liu. 2017. Incorporating global visual features into attention-based neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 992–1003, Copenhagen, Denmark. Association for Computational Linguistics.

Iacer Calixto, Qun Liu, and Nick Campbell. 2017. Doubly-attentive decoder for multi-modal neural machine translation. In *Proceedings of the 55th Annual*

Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pages 1913–1924. Association for Computational Linguistics.

Koel Dutta Chowdhury, Mohammed Hasanuzzaman, and Qun Liu. 2018. Multimodal neural machine translation for low-resource language pairs using synthetic data. In *"."*, pages 33–42.

Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 944–952, Cambridge, MA. Association for Computational Linguistics.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. Association for Computational Linguistics.

Sahinur Rahman Laskar, Abinash Dutta, Partha Pakray, and Sivaji Bandyopadhyay. 2019a. Neural machine translation: English to hindi. In *2019 IEEE Conference on Information and Communication Technology*, pages 1–6.

Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Darsh Kaushik, Partha Pakray, and Sivaji Bandyopadhyay. 2021a. Improved English to Hindi multimodal neural machine translation. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 155–160, Online. Association for Computational Linguistics.

Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Partha Pakray, and Sivaji Bandyopadhyay. 2020a. EnAsCorp1.0: English-Assamese corpus. In *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, pages 62–68, Suzhou, China. Association for Computational Linguistics.

Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Partha Pakray, and Sivaji Bandyopadhyay. 2020b. Multimodal neural machine translation for English to Hindi. In *Proceedings of the 7th Workshop on Asian Translation*, pages 109–113, Suzhou, China. Association for Computational Linguistics.

Sahinur Rahman Laskar, Partha Pakray, and Sivaji Bandyopadhyay. 2019b. Neural machine translation: Hindi-Nepali. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 202–207, Florence, Italy. Association for Computational Linguistics.

Sahinur Rahman Laskar, Partha Pakray, and Sivaji Bandyopadhyay. 2021b. Neural machine translation: Assamese–bengali. In *Modeling, Simulation and Optimization: Proceedings of CoMSO 2020*, pages 571–579. Springer Singapore.

Sahinur Rahman Laskar, Partha Pakray, and Sivaji Bandyopadhyay. 2021c. Neural machine translation for low resource assamese–english. In *Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India*, volume 170, page 35. Springer.

Sahinur Rahman Laskar, Bishwaraj Paul, Partha Pakray, and Sivaji Bandyopadhyay. 2022. Improving english-assamese neural machine translation using transliteration-based approach. In *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications, FICTA 2022*. In press.

Sahinur Rahman Laskar, Rohit Pratap Singh, Partha Pakray, and Sivaji Bandyopadhyay. 2019c. English to Hindi multi-modal neural machine translation and Hindi image captioning. In *Proceedings of the 6th Workshop on Asian Translation*, pages 62–67, Hong Kong, China. Association for Computational Linguistics.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

Toshiaki Nakazawa, Hideya Mino, Isao Goto, Raj Dabre, Shohei Higashiyama, Shantipriya Parida, Anoop Kunchukuttan, Makoto Morishita, Ondřej Bojar, Chenhui Chu, Akiko Eriguchi, Kaori Abe, Yusuke Oda, and Sadao Kurohashi. 2022. Overview of the 9th workshop on Asian translation. In *Proceedings of the 9th Workshop on Asian Translation (WAT2022)*, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.

Shantipriya Parida and Ondřej Bojar. 2020. Hindi visual genome 1.1. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Amarnath Pathak and Partha Pakray. 2018. Neural machine translation for indian languages. *Journal of Intelligent Systems*, pages 1–13.

Amarnath Pathak, Partha Pakray, and Jereemi Bentham. 2018. English–mizo machine translation using neural and statistical approaches. *Neural Computing and Applications*, 30:1–17.

Loitongbam Sanayai Meetei, Thoudam Doren Singh, and Sivaji Bandyopadhyay. 2019. WAT2019: English-Hindi translation on Hindi visual genome dataset. In *Proceedings of the 6th Workshop on Asian Translation*, pages 181–188, Hong Kong, China. Association for Computational Linguistics.

Sukanta Sen, Mohammed Hasanuzzaman, Asif Ekbal, Pushpak Bhattacharyya, and Andy Way. 2020. Neural machine translation of low-resource languages using smt phrase pair injection. *Natural Language Engineering*, page 1–22.

Kashif Shah, Josiah Wang, and Lucia Specia. 2016. SHEF-multimodal: Grounding machine translation on images. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 660–665, Berlin, Germany. Association for Computational Linguistics.