Neural Machine Translation in Low-Resource Setting: a Case Study in English-Marathi Pair

Aakash Banerjee   abanerjee@cse.iitb.ac.in
Aditya Jain   adityajainiitb@cse.iitb.ac.in
Shivam Mhaskar   shivammhaskar@cse.iitb.ac.in
Sourabh Deoghare   sourabhdeoghare@cse.iitb.ac.in
Aman Sehgal   aman.sehgal@iiml.org
Pushpak Bhattacharyya   pb@cse.iitb.ac.in
Department of Computer Science and Engineering, IIT Bombay, India.

Abstract

In this paper, we explore different techniques of overcoming the challenges of low-resource in Neural Machine Translation (NMT), specifically focusing on the case of English-Marathi NMT. NMT systems require a large amount of parallel corpora to obtain good quality translations. We try to mitigate the low-resource problem by augmenting parallel corpora or by using transfer learning. Techniques such as Phrase Table Injection (PTI), back-translation and mixing of language corpora are used for enhancing the parallel data; whereas pivoting and multilingual embeddings are used to leverage transfer learning. For pivoting, Hindi comes in as assisting language for English-Marathi translation. Compared to baseline transformer model, a significant improvement trend in BLEU score is observed across various techniques. We have done extensive manual, automatic and qualitative evaluation of our systems. Since the trend in Machine Translation (MT) today is post-editing and measuring of Human Effort Reduction (HER), we have given our preliminary observations on Translation Edit Rate (TER) vs. BLEU score study, where TER is regarded as a measure of HER.

1 Introduction

The aim of this work is to improve the quality of Machine Translation (MT) for the English-Marathi language pair for which less amount of parallel corpora is available. One of the major requirements for good performance of the Neural Machine Translation (NMT) models is the availability of a large parallel corpora. As a result, there is a need to come up with additional resources by augmenting parallel corpora or by using knowledge from other tasks using transfer learning.

Kunchukuttan and Bhattacharyya (2020) have shown that the lexical and orthographic similarities among languages can be utilized to improve translation quality between Indic languages when limited parallel corpora is available. English and Marathi does not have common ancestry and hence are not related languages, whereas Hindi and Marathi are related languages. Also, among the various English to Indic language

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 35

corpora, English-Hindi corpus is comparatively larger. In our pivot based transfer learning, combined corpus, and multilingual experiments we try to utilize this high resource English-Hindi language pair in various ways to assist in English-Marathi translation. In our Phrase Table Injection (PTI) experiment, we explore how the phrases generated during Statistical Machine Translation (SMT) model training can be further utilized in NMT. We also explore how the monolingual corpus of the target language can be leveraged to create additional pseudo-parallel sentences using back-translation. We also try to understand the correlation between the BLEU and Translation Edit Rate (TER) scores by fitting a linear regression line on the TER vs BLEU graph, where TER is regarded as a measure of Human Effort Reduction (HER).

2 Related Work

Transformer model (Vaswani et al., 2017) was introduced in 2017 and gave significant improvements in the quality of translation as compared to the previous Recurrent Neural Network (RNN) based approaches (Bahdanau et al., 2014; Cho et al., 2014; Sutskever et al., 2014). Self-Attention and absence of recurrent layers enabled models to train faster and get better performance. However, this did not help improve the translation quality in the low-resource setting.

Various methods have been proposed over the years to deal with the low-resource NMT problem. Some methods which use monolingual data involve integrating a separately trained language model (Gülçehre et al., 2015) into the decoder, using an autoencoding objective (Luong et al., 2015) or augmenting pseudo-parallel data using back-translation (Sennrich et al., 2016). Sen et al. (2018) introduced a method for combining SMT and NMT by taking phrases from SMT training and augmenting them to NMT. Zoph et al. (2016) introduced a transfer learning approach where a parent model trained on a high resource language pair is used to initialize some parameters of the child model, which is then trained on a low-resource language pair. Kim et al. (2019) also uses a transfer learning approach with the help of a pivot language to learn parameters initially which are then transferred. Multi-lingual NMT (Zoph and Knight, 2016; Firat et al., 2016; Johnson et al., 2017) is another approach which uses knowledge transfer among various languages to improve the performance of all the language pairs involved.

3 Our Approaches

In this section, we discuss the details of the various techniques that we have explored to deal with the problem of low-resource English-Marathi language pair.

3.1 Phrase Table Injection (PTI)

Sen et al. (2018) and Dewangan et al. (2021) used this technique, shown in Figure 1, to combine both SMT and NMT. We know that the phrase table, generated during training of a SMT model, plays a key role in the SMT translation process. It contains a probabilistic mapping of phrases from the source language to the target language. The phrases present in the phrase table are combined with the available parallel corpora; thereby increasing the data available to train the NMT model. This also helps the model to learn translation of short correct phrases along with long sentences.

3.2 Expansion of data using Back-Translation

Back-translation (Sennrich et al., 2016) is a technique that uses monolingual data of the target language to improve the translation performance of low-resource language

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track
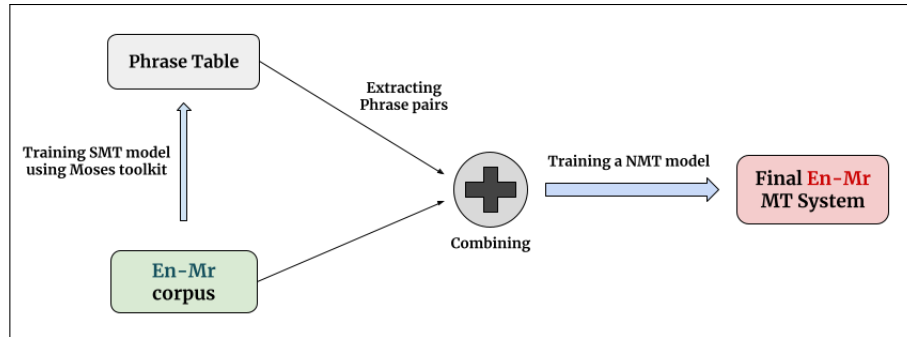
Page 36

Figure 1: Phrase Table Injection

pairs. The amount of available monolingual data in the target language typically far exceeds the amount of parallel data. In SMT, this monolingual data can be used to train a language model, which accounts for fluent translations in SMT. This ability of leveraging the monolingual data for training can be incorporated in NMT by the process of back-translation.

Initially, the available parallel corpora is used to train a Marathi-English NMT model. This model is then used to translate the Marathi monolingual data to create a pseudo-parallel corpus, which in turn is combined with the available parallel corpora to train the NMT model. The ratio of parallel corpora to pseudo-parallel corpora is tuned depending on various factors like quality of target to source model, languages involved in translation, to name a few.

3.3 Combined Corpus

In this technique we exploit the knowledge from similar languages on the target side. As shown in Figure 2, we first train a NMT model using combined corpora from English-Marathi and English-Hindi (EnglishEnglish-HindiMarathi) language pairs. This model is then fine-tuned with the English-Marathi parallel corpora only, using the same vocabulary as that used while training. The intuition is that a model which at the start of training knows how to translate mixed languages is better than a model initialized with random weights.
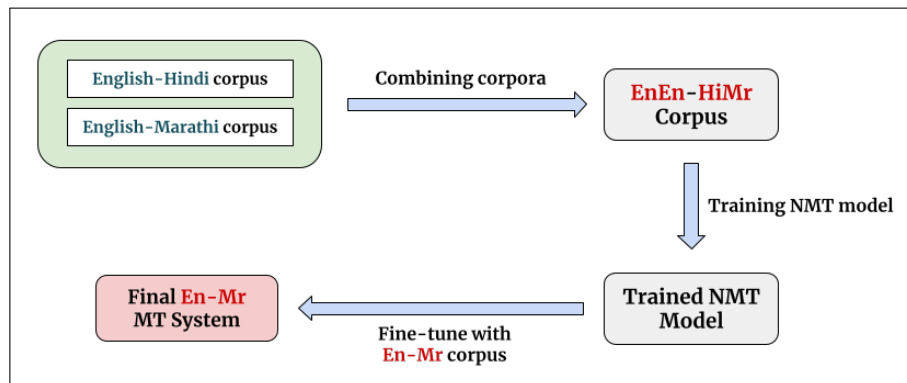

Figure 2: Combined Corpus

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 37

This technique will be more effective if the languages at the target side are similar as this will potentially lead to a partial overlap in the target side vocabulary. Here Hindi and Marathi are the target languages which are similar as both belong to the same language family (Indo-Aryan) and have an overlap in their alphabet set.

3.4 Transfer Learning Approach

The transfer learning approach we used utilizes a pivot language. For the task of English to Marathi translation we use Hindi as a pivot language which assists this task. We chose Hindi as the pivot language because Hindi and Marathi are linguistically close languages. Also English-Hindi parallel corpus is larger as compared to other English to Indic language pairs. We use two pivot based transfer learning techniques proposed by Kim et al. (2019), both of which are discussed below.
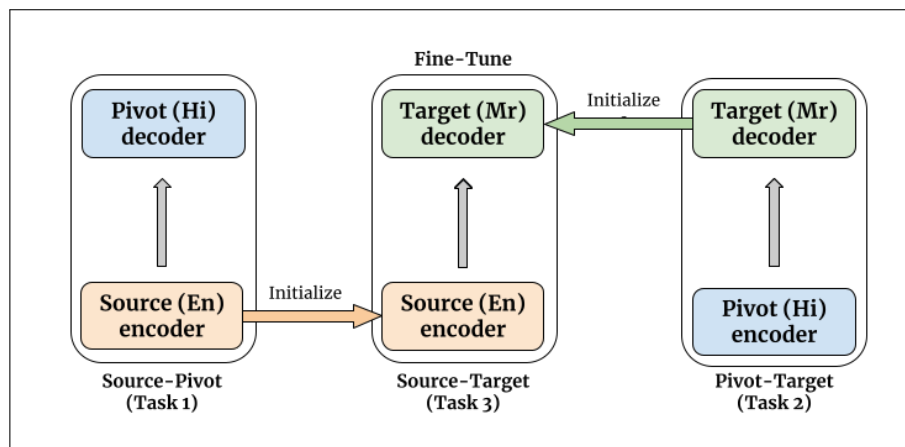

Figure 3: Direct Pivoting (En:English, Hi:Hindi, Mr:Marathi)

3.4.1 Direct Pivoting

In this technique we train two separate NMT models, a source-pivot model and a pivot-target model. As demonstrated in Figure 3, we first separately train an English-Hindi (source-pivot) model (task 1) and a Hindi-Marathi (pivot-target) model (task 2) on their respective parallel corpus. We then use the encoder of the English-Hindi (source-pivot) model and the decoder of Hindi-Marathi (pivot-target) model to initialize the encoder and decoder of the English-Marathi (source-target) model, respectively. Finally, we fine-tune this English-Marathi (source-target) model using the available English-Marathi parallel corpus.

The problem with this approach is that the final English-Marathi (source-target) model is built by combining the encoder trained to produce outputs for the pivot decoder instead of the target decoder; and the decoder trained on the outputs of the pivot encoder instead of the source encoder.

3.4.2 Step-wise Pivoting

As shown in Figure 4, here we first train an English-Hindi (source-pivot) model. Then we use the encoder of the English-Hindi (source-pivot) model to initialize the encoder of the Hindi-Marathi (pivot-target) model. After this, we train the Hindi-Marathi (pivot-target) model on the Hindi-Marathi corpus by freezing the encoder parameters. Then

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track
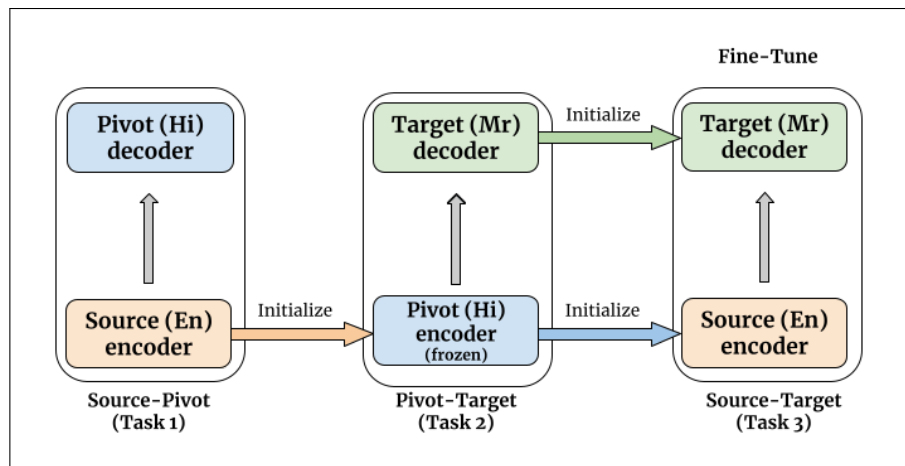
Page 38

Figure 4: Step-wise Pivoting (En:English, Hi:Hindi, Mr:Marathi)

the encoder and decoder from this Hindi-Marathi (pivot-target) model are used to initialize the encoder and decoder of the English-Marathi (source-target) model. Finally, we fine-tune the English-Marathi (source-target) model using the available English-Marathi corpus.

3.5 Multi-Lingual MT System

The various types of multilingual models are one-to-many, many-to-one and many-to-many. Among these, we use the one-to-many multilingual model with source language as English and target languages as Hindi and Marathi. One of the ways to achieve this is by making use of a single encoder for the source language and two separate decoders for the target languages. The disadvantage with this method is that, as there are multiple decoders, the size of the model increases. Another way to achieve this is to use a single encoder and a single shared decoder. An advantage of this method is that the representations learnt by English-Hindi task can further be utilized by the English-Marathi task. Also, Hindi and Marathi being similar languages, the overlap between their vocabulary is large resulting in a smaller shared vocabulary.

4 Experiments

In this section, we discuss the details of the various experiments that we have carried out to improve the quality of the English-Marathi translation.

4.1 Dataset Preparation

The NMT models were trained using a corpus formed by combining the Indian Languages Corpora Initiative (ILCI) Phase 1 corpus (Jha, 2010), Bible corpus (Christodouloupoulos and Steedman, 2015; Jha, 2010), CVIT-Press Information Bureau (CVIT-PIB) corpus (Philip et al., 2021), IIT Bombay English-Hindi corpus (Kunchukuttan et al., 2017) and PMIndia (PMI) corpus (Haddow and Kirefu, 2020). The English-Marathi corpus, English-Hindi corpus and Hindi-Marathi corpus consisted of 0.25M, 2M and 0.24M parallel sentences, respectively. Barring the ILCI corpus, the remaining Hindi-Marathi data was synthetically generated by translating the English sentences

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 39

	Number of Sentences		
	English-Marathi	**English-Hindi**	**Hindi-Marathi**
ILCI	46,277	46,277	46,277
Bible corpus	60,876	62,073	58,375
IITB corpus	—	1,603,080	—
CVIT-PIB	114,220	266,545	108,220
PMIndia	28,974	50,349	28,973
Total Corpus Size	250,347	2,028,324	241,845

Table 1: Corpora statistics of the three language pairs: English-Marathi, English-Hindi and Hindi-Marathi

of the English-Marathi corpus to Hindi using the Google Cloud Translation API[1]. The detailed corpus statistics are mentioned in Table 1.

For reporting the results, the test set introduced in WAT 2021[2] MultiIndicMT: An Indic Language Multilingual Task[3] and the test set from ILCI corpus are used. The test set from WAT 2021 contains 2,390 sentences and is a part of the PMIndia corpus. The PMIndia corpus from WAT 2021 task is used for training to avoid any overlap between the train and test sets. The test set from ILCI corpus consists of 2000 sentences.

The English sentences are tokenized and lowercased using Moses[4] (Koehn et al., 2007) toolkit. The Hindi and Marathi sentences are lowercased and normalized using Indic NLP Library (Kunchukuttan, 2020). Byte Pair Encoding (BPE) (Sennrich et al., 2015) is used as a segmentation technique; as breaking up words into subwords has become standard now and is especially helpful for morphologically rich languages like Marathi and Hindi.

4.2 Training Setup

The Transformer architecture was used to train the NMT models. The PyTorch version of OpenNMT (Klein et al., 2017) was used to carry out the PTI, combined corpus and back-translation experiments. For the pivot language based transfer learning and multilingual NMT experiments, the fairseq (Ott et al., 2019) library was used. The SMT model for PTI was trained using the Moses toolkit.

For the baseline model, a vanilla transformer model was trained using the default parameters[5] for 200K training steps. In the experiment with PTI, Moses toolkit was used to train the model to get phrases from the phrase table. The grow-diag-final-and method was used for symmetrization and msd-bidirectional-fe method was used for lexicalized reordering. While making batches for training, the parallel data and parallel phrases were selected in the ratio 4:1, as giving less weightage to phrases enhances the performance. For back-translation experiment, the amount of pseudo-parallel sentences used is same as that of the available corpus. Both the corpus were combined and a model was trained with the default parameters. In the combined corpus experiment, the model was trained for 200k training steps and then was further fine-tuned for 100k training steps.

---

[1]https://cloud.google.com/translate

[2]http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2021/index.html

[3]http://lotus.kuee.kyoto-u.ac.jp/WAT/indic-multilingual/

[4]https://github.com/moses-smt/mosesdecoder

[5]https://opennmt.net/OpenNMT-py/FAQ.html

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 40

For the pivot language based transfer learning experiments, a transformer model from the fairseq library was used. The optimizer used was adam with betas (0.9, 0.98). The initial learning rate used was 5e-4 with the inverse square root learning rate scheduler and 4000 warm-up updates. The dropout probability value used was 0.3 and the criterion used was label smoothed cross entropy with label smoothing of 0.1. All the models were trained for 400 epochs. Same training setup was used for the multilingual NMT experiments as well. A one (English) to many (Hindi, Marathi) multilingual model was used. As the multilingual model we used had a shared decoder, the source sentence was prepended with a target language token, both at the training and the inference time, to specify the target language during translation.

5 Results and Analysis

We use the BLEU evaluation metric (Papineni et al., 2002) to report our results. Sacrebleu (Post, 2018) python library was used to calculate the BLEU scores. We detokenize the translated sentences before calculating the BLEU scores. The results of all our experiments are summarized in Table 2.

Model	ILCI	WAT 2021
Baseline	16.03	16.26
Phrase Table Injection (PTI)	15.81	17.15
Combined Corpus (CC)	17.69	18.02
Backtranslation (BT)	15.90	15.78
BT + PTI	15.83	16.34
CC + BT	17.51	17.45
CC + PTI	17.75	17.97
CC + PTI + BT	17.47	17.43
Direct Pivoting	18.32	16.68
Step-wise Pivoting	17.94	16.74
Multi-Lingual	18.83	17.09

Table 2: BLEU scores of English-Marathi language pair using various techniques (CC: Combined Corpus, BT: Backtranslation, PTI: Phrase Table Injection).

In PTI experiment we observe an increase in BLEU score on WAT 2021 test set while the BLEU score on ILCI test set decreases. For combined corpus there is improvement of more than 1.5 BLEU score on both the test sets, indicating that English-Hindi corpus helped during the training. We observe that using back-translation, the BLEU score decreases. This can be attributed to the fact that the Marathi-English model used for back-translating the Marathi monolingual corpora was not of good quality. This Marathi-English model was trained using 0.25M parallel sentences which affects the quality of back-translated data. We also tried out experiments by combining the above mentioned methods, among which, the combination of phrase table injection and combined corpus methods give the best results.

The results of the direct pivoting technique show an improvement of 2.29 BLEU score over the baseline model on the ILCI test set and of 0.42 on the WAT 2021 test set. The results of step-wise pivoting show an improvement of 1.91 BLEU score over the baseline on the ILCI test set and of 0.74 on the WAT 2021 test set. The reason

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 41

for this BLEU score increase is that, the encoder and decoder used to initialize the English-Marathi model before training have already learned some representations. This is because the encoder and decoder are initialized from the encoder and decoder of the trained English-Hindi (source-pivot) and Hindi-Marathi (pivot-target) models, respectively. We observe that this initialization of encoder and decoder performs better than a random initialization.

The results of the multilingual model on English-Marathi translation task show a BLEU score increase of 2.8 on the ILCI test set and 0.83 on the WAT 2021 test set over the baseline model. In a multilingual model, we use a shared decoder for both the target languages, due to which, the representations learnt by the model for the task of English-Hindi translation helps in the English-Marathi task as well. This leads to a better performance of the multilingual model over the baseline model. For direct pivoting, step-wise pivoting and multilingual model we observe that the BLEU score increase on ILCI test set is more than that on the WAT 2021 test set. Our conjecture is that as the size of the ILCI corpus used in training is larger than that of the PMIndia corpus(from which WAT 2021 test set is derived), the BLEU score increase for ILCI test set is more.

6 Extensive Evaluation

In this section, we discuss the analysis that we have carried out to compare our models with baseline; and also understand the correlation of BLEU score with TER, where TER is regarded as a measure of HER.

6.1 Qualitative Analysis

In this sub-section, we present the analysis of few sentences to demonstrate how our model performs better than the baseline. In each of the below given examples, **En-Source** represents source English sentences, **Mr-xx** represents translated Marathi sentence using "**xx**" model, **Mr-xx-Transliterate** represents translated Marathi sentence transliterated in English, and **Mr-xx-Gloss** represents word-to-word English translations of the translated Marathi sentence.

- **Example 1: Translation of named entity**

  En-Source: The toy train from Kalka to Shimla is considered as the most beautiful rail line in India.

  Mr-Baseline: कालका ते सिमलापर्यंत धावणारी रेल्वे भारतात सर्वात सुंदर रेल्वे लाईन मानली जाते.

  Mr-Baseline-Transliterate: kalka te shimlaa dhavanari railway bharatat cervat sunder railway laiin maanli jate.

  Mr-Baseline-Gloss: Kalka to up-to-Shimla running railway in-India most beautiful railway line considered is.

  Mr-DirectPivoting: कालकापासून सिमला पर्यंतची टॉय ट्रेन भारतातील सर्वात सुंदर रेल्वे लाइन मानली जाते .

  Mr-DirectPivoting-Transliterate: kalkaapasun shimla paryantachi toy train bharatatil sarvat sundar railway laiin maanli jate.

  Mr-DirectPivoting-Gloss: From-Kalka Shimla up-to toy train in-India most beautiful railway line considered is.

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 42

The English source sentence contains a named entity "***toy train***". The baseline translated "***toy train***" incorrectly as "रेल्वे" (which means "***railway***"), whereas our model was able to translate "***toy train***" correctly as "टॉय ट्रेन" (which means**"toy train"**).

- **Example 2: Translation of long sentences**

  En-Source: The Prime Minister expressed happiness that on this occasion, the devotional hymn Vaishnav Jan To, which was so dear to Bapu, had been rendered by artistes in about 150 countries across the world.

  Mr-Baseline: यावेळी पंतप्रधानांनी आनंद व्यक्त केला की, पूज्य बापूंच्या देशांचे निष्ठावंत भजन असणाऱ्या वैष्णव जन यांना जगातील सुमारे 150 देशांमध्ये पारितोषिके देण्यात आली होती.

  Mr-Baseline-Transliterate: yawelii pantpradhanani anand vyakt kela kii, pujy bapunchyaa deshanche nithavant bhajan assnaryaa vaishnav jan yaannaa jagaatil sumaare 150 deshaanmadhye paaritoshik deanyaat aali hoti .

  Mr-Baseline-Gloss: This-time Prime-Minister happiness expressed did that, reverend Bapu's of-countries loyal hymn is vaishnav jan to in-world around 150 in-countries awards given came was.

  Mr-PTI: सुमारे 150 देशांमध्ये बापूंना प्रिय असलेले वैष्णव जन तो भक्तीगीत जगभरातल्या 150 देशातल्या कलाकारांनी सादर केल्याबद्दल पंतप्रधानांनी आनंद व्यक्त केला.

  Mr-PTI-Translitarate: sumaare 150 deshaanmadhye baapunnaa priya aslele vaishnav jan to bhaktigeet jagbharatlyaa 150 deshatlyaa kalakaranii saadar kelyaabaddadal pantprdhaanaanii aanand vyakt kelaa .

  Mr-PTI-Gloss: Around 150 in-countries to-Bapu loved vaishnav jan to devotional-song around-world 150 in-countries artists performed for-doing Prime-Minister happiness express did.

  The baseline model was not able to completely translate the long source English sentence adequately. The model was able to translate the entire long source English sentence adequately.

- **Example 3: Translation of low readability sentences**

  En-Source: The Union Cabinet chaired by the Prime Minister Shri Narendra Modi has given its ex post-facto approval to the MoU between India and Singapore on cooperation in the field of urban planning and development.

  Mr-CombinedCorpus: पंतप्रधान नरेंद्र मोदी यांच्या अध्यक्षतेखाली केंद्रीय मंत्रिमंडळाने भारत आणि सिंगापूर दरम्यान शहर नियोजन आणि विकास क्षेत्रातील सहकार्याबाबतच्या सामंजस्य कराराला कार्योत्तर मंजुरी दिली.

  Mr-CombinedCorpus-Transliterate: pantprdhaan narendra modi yaanchyaa adhyakshatekhali kendriya mantrimandalaane bharat aani singapore darmyaan shahar niyojan aani vikas kshetraatiil sahakaaryaabaabatchyaa saamanjsy karaaraalaa kaaryottar manjurii dilii .

  Mr-CombinedCorpus-Gloss: Prime-Minister Narendra Modi his chaired-under central cabinet India and Singapore during city planning and development in-field regarding-cooperation Memorandum-of understanding after-work approval given.

  The above example shows the performance of our model on sentences with low readability i.e. sentences with high Automated Readability Index (ARI) (Senter and Smith, 1967). Our model was able to translate the low readability sentence adequately and fluently.

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 43

6.2 Understanding the Correlation between BLEU and TER
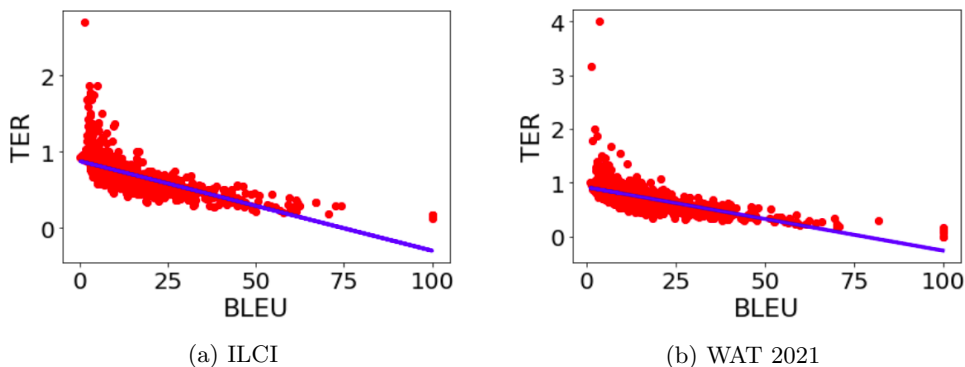

(a) ILCI   (b) WAT 2021

Figure 5: TER vs BLEU for ILCI and WAT 2021 test sets

The trend in machine translation these days is to perform post-editing on the output of the MT system. When post-editing is performed by humans on a large amount of sentences, it is very important to measure the reduction in human effort by the MT system. This can be achieved by calculating the HER which can be an important MT evaluation metric. In this paper, we use TER (Snover et al., 2006) as a measure of HER. TER measures the amount of editing that is required by a human to convert a system output to a reference translation.

In order to understand the correlation between BLEU and TER scores, we plot sentencewise TER vs BLEU score graphs for the ILCI and WAT 2021 test sets. Figure 5 shows that as the BLEU score increases the TER decreases. A linear regression line was fitted on TER vs BLEU graph for both the ILCI and WAT 2021 test sets.

$$y = -0.0117x + 0.8805 \tag{1}$$

$$y = -0.0117x + 0.9124 \tag{2}$$

Equation 1 represents the linear regression line on the ILCI test set having slope of -0.0117 and the y-intercept as 0.8805. Equation 2 represents the linear regression line on the WAT 2021 test set having a slope of -0.0117 and the y-intercept as 0.9124. We observe that the slope of the line is negative for both the equations indicating that BLEU and TER are negatively correlated. This is expected as BLEU is a measure of how *good* the sentence got translated, whereas TER is a measure of how *bad* the sentence got translated. This supports the use of TER as a metric of HER.

7 Conclusion and Future Work

In this work, we have implemented and compared various techniques to improve the task of translation involving a low-resource English-Marathi language pair. We have shown that the pivot based transfer learning approach can significantly improve the quality of the English-Marathi translations over the baseline by using Hindi as an assisting language. We also observe that the phrases from the SMT training can help the NMT model perform better. The one (English) to many (Hindi, Marathi) multilingual model is able to improve the English-Marathi translations by leveraging the English-Hindi parallel corpus. Combined corpus experiment also uses the English-Hindi parallel corpus to improve the English-Marathi translation quality.

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 44

In future, we plan to further extend these approaches to a variety of languages to understand how the phenomenon of language relatedness can help improve the translation quality in low resource setting. We also plan to explore how multiple pivot languages can be used while translating from some source to target language pair.

Acknowledgement

References

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

Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259.

Christodouloupoulos, C. and Steedman, M. (2015). A massively parallel corpus: the bible in 100 languages. *Language resources and evaluation*, 49(2):375–395.

Dewangan, S., Alva, S., Joshi, N., and Bhattacharyya, P. (2021). Experience of neural machine translation between indian languages. *Machine Translation*, pages 1–29.

Firat, O., Cho, K., and Bengio, Y. (2016). Multi-way, multilingual neural machine translation with a shared attention mechanism. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 866–875, San Diego, California. Association for Computational Linguistics.

Gülçehre, Ç., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H., Bougares, F., Schwenk, H., and Bengio, Y. (2015). On using monolingual corpora in neural machine translation. *CoRR*, abs/1503.03535.

Haddow, B. and Kirefu, F. (2020). Pmindia - A collection of parallel corpora of languages of india. *CoRR*, abs/2001.09907.

Jha, G. N. (2010). The TDIL program and the Indian langauge corpora intitiative (ILCI). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Kim, Y., Petrov, P., Petrushkov, P., Khadivi, S., and Ney, H. (2019). Pivot-based transfer learning for neural machine translation between non-English languages. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 45

(EMNLP-IJCNLP), pages 866–876, Hong Kong, China. Association for Computational Linguistics.

Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. (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.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Kunchukuttan, A. (2020). The IndicNLP Library. `https://github.com/anoopkunchukuttan/indic_nlp_library/blob/master/docs/indicnlp.pdf`.

Kunchukuttan, A. and Bhattacharyya, P. (2020). Utilizing language relatedness to improve machine translation: A case study on languages of the indian subcontinent. *arXiv preprint arXiv:2003.08925*.

Kunchukuttan, A., Mehta, P., and Bhattacharyya, P. (2017). The iit bombay english-hindi parallel corpus. *arXiv preprint arXiv:1710.02855*.

Luong, M.-T., Le, Q. V., Sutskever, I., Vinyals, O., and Kaiser, L. (2015). Multi-task sequence to sequence learning. *arXiv preprint arXiv:1511.06114*.

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., and Auli, M. (2019). fairseq: A fast, extensible toolkit for sequence modeling. *arXiv preprint arXiv:1904.01038*.

Papineni, K., Roukos, S., Ward, T., and 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*, pages 311–318.

Philip, J., Siripragada, S., Namboodiri, V. P., and Jawahar, C. (2021). Revisiting low resource status of indian languages in machine translation. In *8th ACM IKDD CODS and 26th COMAD*, pages 178–187.

Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

Sen, S., Hasanuzzaman, M., Ekbal, A., Bhattacharyya, P., and Way, A. (2018). Neural machine translation of low-resource languages using smt phrase pair injection. *Natural Language Engineering*, pages 1–22.

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Sennrich, R., Haddow, B., and Birch, A. (2016). Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 46

Senter, R. and Smith, E. A. (1967). Automated readability index. Technical report, CINCINNATI UNIV OH.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of association for machine translation in the Americas*, volume 200. Citeseer.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *CoRR*, abs/1706.03762.

Zoph, B. and Knight, K. (2016). Multi-source neural translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 30–34, San Diego, California. Association for Computational Linguistics.

Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.

Proceedings of the 18th Biennial Machine Translation Summit
Virtual USA, August 16 - 20, 2021, Volume 1: MT Research Track

Page 47