Evaluating the Performance of Back-translation for Low Resource English-Marathi Language Pair: CFILT-IITBombay @ LoResMT 2021

Aditya Jain adityajainiitb@cse.iitb.ac.in
Shivam Mhaskar shivammhaskar@cse.iitb.ac.in
Pushpak Bhattacharyya pb@cse.iitb.ac.in
Department of Computer Science and Engineering, IIT Bombay, India.

Abstract

In this paper, we discuss the details of the various Machine Translation (MT) systems that we have submitted for the English-Marathi LoResMT task. As a part of this task, we have submitted three different Neural Machine Translation (NMT) systems; a Baseline English-Marathi system, a Baseline Marathi-English system, and an English-Marathi system that is based on the back-translation technique. We explore the performance of these NMT systems between English and Marathi languages, which forms a low resource language pair due to unavailability of sufficient parallel data. We also explore the performance of the back-translation technique when the back-translated data is obtained from NMT systems that are trained on a very less amount of data. From our experiments, we observe that the back-translation technique can help improve the MT quality over the baseline for the English-Marathi language pair.

1 Introduction

In this work, we explore various ways to perform Machine Translation (MT) in low resource settings, that is, when very less amount of parallel data is available to train the model. We also explore the performance of the back-translation technique, which is one of the data augmentation techniques to overcome the problem of low resource in neural machine translation. In our work, we focus on the Neural Machine Translation (NMT) systems, which requires a large amount of parallel training data to produce good quality translations. This is the major reason behind NMT systems to be considered as *data hungry*. The language pair for which less amount of parallel data is available is considered a low resource language pair.

As compared to parallel data, monolingual data is easier to obtain and is available in relatively large quantity. This monolingual data can also be used to improve the performance of the NMT system. We explore the back-translation technique to make use of the available monolingual data to create an augmented pseudo-parallel data which can then be used to train the NMT system. In back-translation, the monolingual data is first translated using a machine translation system. In case of low resource languages, this machine translation system is trained on very less amount of data and hence, the translation of the monolingual data produced may not be of very high quality. If the model is trained using this low quality back-translated parallel data, it can degrade the performance of the system. We explore the performance of the NMT system when it is trained using back-translation technique in which the back-translated data is generated from a NMT system which is trained on a less amount of data. From the experiments that we have performed as part of the LoResMT 2021 (Ojha et al., 2021) task, we observe that the back-translation technique gives a BLEU score improvement of up to 1.2 points over the baseline model for the English-Marathi MT task.

2 Related Work

Neural Machine Translation systems were initially based on Recurrent Neural Network (RNN) based approaches (Cho et al., 2014; Sutskever et al., 2014). But Recurrent Neural Network based architectures were not able to capture long term dependencies in long sentences. In order to overcome this problem, Attention (Bahdanau et al., 2014) mechanism was introduced. The Attention-based RNN architecture still suffered from problems like longer training time because of their sequential nature. Later Transformer architecture (Vaswani et al., 2017) was introduced which improved the performance of the NMT systems and also lead to faster training due to its non-sequential nature.

Sennrich et al. (2016) introduced the technique of back-translation in which monolingual data is used to create augmented pseudo-parallel data. Sen et al. (2018) used Statistical Machine Translation (SMT) system by extracting phrases generated during SMT training and using them along with the training data for NMT systems. Zoph et al. (2016) introduced a transfer learning techniques in which a parent NMT model is initially trained on a high resource language pair and then the parameters of this parent model are used to initialize a child model, which is then trained on a low resource language pair. Kim et al. (2019) introduced a transfer learning technique based on a pivot language in which a pivot language is used to pre-train the encoder and decoder of a NMT model, which are then used to initialize the encoder and decoder of the final NMT model which is then fine-tuned on the low resource language pair. Multi-lingual NMT models (Zoph and Knight, 2016; Firat et al., 2016; Johnson et al., 2017) that can translate to or from multiple languages have shown performance improvements in the case of low resource language pairs when the system also includes high resource language pairs.

3 Approaches

In this section, we discuss the various techniques that we have used to implement our English-Marathi and Marathi-English MT systems.

3.1 Baseline Model

In our Baseline English-Marathi and Marathi-English MT models, we train a NMT system using the given English-Marathi parallel corpus for 1600 epochs and we save the model for every 200 epochs starting from 200, to test the performance.

3.2 Back-Translation

Back-translation technique makes use of monolingual data of source or target language to generate source-target parallel sentences using a trained NMT system. From the provided English-Marathi parallel data and Marathi monolingual data, we first train a Marathi-English NMT system using the English-Marathi parallel data. We then use this Marathi-English model to translate the Marathi monolingual data to get the corresponding English output. Finally we combine the given English-Marathi parallel data and this back-translated English-Marathi pseudo-parallel data to train our English-Marathi back-translation NMT system.

4 Experiments

In this section, we discuss the various experiments that we have performed as a part of this work.

4.1 Dataset

Type of Data	**Number of sentences**
Parallel	20,933
Monolingual	21,902

Table 1: Dataset

We used the English-Marathi parallel corpus provided by the LoResMT 2021 organizers, which consisted of 20,933 English-Marathi parallel sentences. Further, for our back-translation experiment we used the Marathi monolingual corpus provided by the LoResMT 2021 organizers which consisted of 21,902 Marathi sentences.

4.2 Training Setup

For all of the NMT systems discussed in this paper, we have used the transformer-based architecture which we have implemented using the fairseq Ott et al. (2019) library. This transformer-based architecture consisted of 6 encoder layers and 6 decoder layers. The number of encoder and decoder attention heads used were 4 each. We used encoder and decoder embedding dimension of 512 each. For training the system, the optimizer used was Adam optimizer with betas (0.9, 0.98). The inverse square root learning rate scheduler was used with initial learning rate of 5e-4 and 4,000 warm-up updates. The criterion used was label smoothed cross entropy with label smoothing of 0.1. The dropout probability value of 0.3 was used.

5 Results and Analysis

Model	**English-Marathi**	**Marathi-English**
Baseline-200	11	16.8
Baseline-400	10.4	17.1
Baseline-600	—	17.2
Baseline-800	10.6	17.2
Baseline-1000	10.5	16.6
Baseline-1200	10.7	16.3
Baseline-1400	10.5	16.2
Baseline-1600	10.8	—
Back-translation	12.2	—

Table 2: BLEU scores of English-Marathi language pair where for a model named Baseline-X, X represents the number of epochs for which the model was trained.

Table 2 shows the results of the different techniques used to implement the MT systems for the English-Marathi language pair. We used BLEU (Papineni et al., 2002) metric to measure the performance of the MT systems. The baseline English-Marathi system produced a BLEU score of 11 and the baseline Marathi-English System produced a BLEU score of 17.2. We observe that the English-Marathi and Marathi-English NMT were trained using the same English-Marathi parallel data, the Marathi-English system produced higher BLEU scores than that produced by the English-Marathi system. We also observe that the English-Marathi system gives the best BLEU score after 400 epochs and after that the scores decrease and fluctuate between a small range. The Marathi-English model gives the best score after 600 epochs and after that the scores starts decreasing. We have used this Marathi-English NMT system to translate the Marathi monolingual data to English. Then we trained the English-Marathi back-translation system using the given English-Marathi parallel data and the English-Marathi back-translated pseudo-parallel data. This back-translation system produced a BLEU score of 12.2. We observe that even though the Marathi-English back-translated data was produced using a machine translation system which was trained on very low amount of data, there is still an increase in BLEU score of around 1.2 points over the baseline model.

6 Conclusion

In this work, we implement various English-Marathi and Marathi-English baseline NMT systems and use the given monolingual Marathi data to implement the back-translation technique for data augmentation. From our experiments, we observe that the technique of back-translation can help improve the MT quality over the baseline for the English-Marathi MT task for which less amount of parallel data is available.

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.

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.

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

Ojha, A. K., Liu, C.-H., Kann, K., Ortega, J., Satam, S., and Fransen, T. (2021). Findings of the LoResMT 2021 Shared Task on COVID and Sign Language for Low-Resource Languages. In *Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages.*

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.

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. (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.

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.