RL-NMT: Reinforcement Learning Fine-tuning for Improved Neural Machine Translation of Burmese Dialects

Ye Kyaw Thu [1,2] Thazin Myint Oo [2] Thepchai Supnithi [2]

[1]Language Understanding Lab. (LU Lab), Myanmar [2]National Electronics and Computer Technology Center (NECTEC), Thailand

Motivation

Challenges in Low-Resource Dialects: Low-resource dialects present unique translation challenges due to data scarcity and linguistic complexities.

Linguistic Variations: While Beik, Rakhine, and Burmese dialects share a common written script in Burmese characters, they exhibit distinct spoken variations, emphasizing the need for specialized translation approaches.

Gap in Research: Despite their linguistic similarities in writing, there is a lack of research on applying RL-based MT to address the spoken variations of these dialects, highlighting the importance of this study.

Burmese: ငါ မနက်ဖြန် နိုင်ငံခြား သွား မယ်
Beik: ငါ မောလင်း နိုင်ငံခြား သော မယ်
(English: Tomorrow, I will go to a foreign country.)

Burmese: အဘွား ဈေး က ဆပ်ပြာ ဝယ် လာ တယ်
Rakhine: အဘောင်သျင် ဆိုး က သပ္ပဲ ဝယ် လာ တယ်
(English: The grandmother buys soap at the market.)

Figure 1. Example sentences in Burmese, Beik, and Rakhine dialects

Reinforcement Learning (RL)

- **RL Framework**: In Neural Machine Translation (NMT), Reinforcement Learning (RL) treats the source sentence as the system's initial state. Each target language word choice is an action, and the cumulative selection forms the translation. The objective is to maximize a reward, indicating translation quality compared to a reference.

The integration of RL into NMT is a process marked by several key aspects:

- **Reward Policy**: Defines translation evaluation criteria using metrics like GLEU and BLEU, guiding the model towards contextually accurate and fluent translations.

- **Learning from Rewards**: The model optimizes parameters based on translation quality rewards, improving over time.

- **Exploration vs. Exploitation**: Balancing exploration for better translations and exploitation of known high-reward actions is crucial.

- **Challenges**: Challenges include managing exploration-exploitation trade-offs, ensuring stable learning, and handling natural language complexity [1, 3].

Corpus

For Beik-Burmese and Burmese-Beik, the training data consists of 7.8K pairs of sentences with about 68.0M Beik and Burmese words, respectively [4]. We used the 1.3K pairs of sentences as the development or validation set and 1.0K pairs of sentences as the test sets.

For Burmese-Rakhine and Rakhine-Burmese, the training data consists of 15.5K pairs of sentences with about 123.0K Burmese and Rakhine words, respectively [4]. We used the 1.0K pairs of sentences as the development or validation set and 1.8K pairs of sentences as the test sets.

Burmese dialects are low-resource languages and therefore it is dificult to develop a good word segmentation tool based on machine learning techniques. Moreover, manual word segmentations in the developing corpora are also not consistant. Burmese and Burmese dialect sentences are written as contiguous sequences of syllables with no characters delimiting the words. And thus, in this paper, we did syllable segmentation for all language pairs (see Table. 1) by using `sylbreak.pl` [5].

Table 1. Statistics of the parallel train/dev/test data for initial training and finetuning with Reinforcement Learning. Values represent the number of sentences.

Language Pair	Train	Dev	Test
Burmese (my) - Beik (bk)	7871	1390	1037
Burmese (my) - Rakhine (rk)	15561	1000	1811

Setup

We evaluated RL approach (i.e. initial training with MLE and RL with MRT) on four dialect translation tasks: Beik-Burmese, Burmese-Beik, Burmese-Rakhine and Rakhine-Burmese. The evaluation metric is BLEU as calculated by the `multi-bleu.perl` script.

Implementation Details

We use Simple-NMT [2] python library to train both LSTM (Seq2Seq) and Transformer models. The library implements important algorithms including Maximum Likelihood Estimation, Dual Supervised Learning and Reinforcement learning [1] for fine-tuning like Minimum Risk Training. We have performed our experiments using the default optimizer: Adam and reward: GLEU score. Applying RL techniques with the the Transformer architecture was failed several times during baseline experiments due to lack of memory of the GPU. After playing hyperparameters for both Seq2Seq and Transformer, we selected the following hyperparameters for the experiments:

Table 2. The hyperparameters for Seq2Seq and Transformer

Hyperparameters	Seq2Seq	Transformer
Batch Size	64	16
Dropout	0.2	0.2
Word Vector Size	128	512
Hidden Size	128	32
No. of Layers	4	6
No. of Splits	-	8

Results and Discussion

Table 3 showcases the performance of Seq2Seq models trained for 100 epochs, revealing a distinctive pattern in BLEU scores, which illustrates the challenges and opportunities inherent in the sequence-to-sequence approach for dialect translation.

Table 3. The highest BLEU scores of Seq2Seq and Transformer baseline models trained with 100 epochs. (The best epochs in parentheses)

	Baselines	
Source-Target	Sequence to Sequence	Transformer
Beik → Burmese	22.62 (95 epoch)	17.58 (99 epoch)
Burmese → Beik	18.99 (81 epoch)	15.80 (98 epoch)
Burmese → Rakhine	74.92 (96 epoch)	59.35 (100 epoch)
Rakhine → Burmese	74.84 (66 epoch)	55.68 (98 epoch)

The integration of Reinforcement Learning (RL) for model fine-tuning, as depicted in Table 4, demonstrates a nuanced impact on translation quality, measured in BLEU scores.

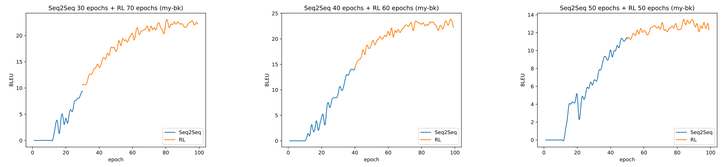
Table 4. The highest BLEU Scores of the initial (Seq2Seq) and finetuned (RL) models for Beik to Burmese and Burmese to Beik translations. For each setting, underline indicates a higher BLEU score than initial pretrained model and ↑ means a higher BLEU score than the baseline model trained for 100 epochs.

No. of Epoch	Beik → Burmese		Burmese → Beik	
Seq2Seq, RL	Seq2Seq	Seq2Seq-RL	Seq2Seq	Seq2Seq-RL
30, 70	9.54	13.43	9.36	22.82 ↑
40, 60	8.70	12.46	14.15	23.72 ↑
50, 50	10.75	12.52	11.42	13.50
60, 40	11.17	11.96	10.29	11.50
70, 30	17.92	21.48	13.40	15.72

Table 3 illustrates the limitations of the Transformer models when trained solely for 100 epochs for dialect translations, evidenced by the BLEU scores. However, integrating Reinforcement Learning (RL) for model fine-tuning, as delineated in Tables 5, significantly elevates the BLEU scores, portraying a more consistent performance across different epoch divisions.

Table 5. The highest BLEU Scores of the initial (Transformer) and finetuned (RL) models for Burmese to Rakhine translation. For each setting, underline indicates a higher BLEU score than initial pretrained model and ↑ means a higher BLEU score than the baseline model trained for 100 epochs.

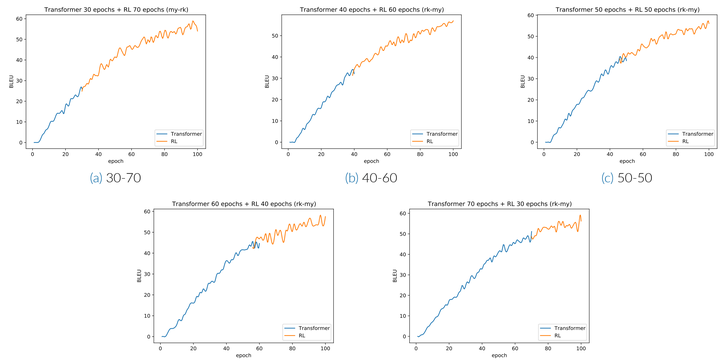
No. of Epoch	Burmese → Rakhine		Rakhine → Burmese	
Transformer, RL	Transformer	Transformer-RL	Transformer	Transformer-RL
30, 70	25.29	58.69	25.23	56.72 ↑
40, 60	35.33	59.13	34.17	56.95 ↑
50, 50	40.85	56.79	40.40	56.43 ↑
60, 40	47.13	58.22	45.70	58.26 ↑
70, 30	50.51	58.20	51.15	57.20 ↑



(a) 30-70 (b) 40-60 (c) 50-50

(d) 60-40 (e) 70-30

Figure 2. BLEU score improvements of Seq2Seq+RL (Minimum Risk Training) for Burmese-Beik (my-bk) language pair



(a) 30-70 (b) 40-60 (c) 50-50

(d) 60-40 (e) 70-30

Figure 3. BLEU score improvements of Transformer+RL (Minimum Risk Training) for Rakhine-Burmese language pair

References

[1] Leshem Choshen, Lior Fox, Zohar Aizenbud, and Omri Abend.
On the weaknesses of reinforcement learning for neural machine translation.
In *International Conference on Learning Representations*, 2020.

[2] Ki-Hyun Kim.
simple-nmt, 2019.

[3] Julia Kreutzer.
Rl in nmt: the good, the bad and the ugly.
`http://www.cl.uni-heidelberg.de/statnlpgroup/blog/rl4nmt/`, 2018.

[4] Thazin Myint Oo, Thitipong Tanprasert, Ye Kyaw Thu, and Thepchai Supnithi.
Transfer and triangulation pivot translation approaches for burmese dialects.
IEEE Access, 11:6150–6168, 2023.

[5] Ye Kyaw Thu.
sylbreak.
`https://github.com/ye-kyaw-thu/sylbreak`, 2017.