Neural Machine Translation on Myanmar Language

Hnin Aye Lwin and Thinn Thinn Wai

University of Information Technology, Yangon, Myanmar,

Abstract. This work covers neural machine translation from Myanmar (Burmese) to English languages. Rule-based syllable breaking approach is used for Myanmar sentences as pre-processing stage. Recurrent neural network Encoder-Decoder architecture with attention mechanism is implemented to evaluate BLEU score on Myanmar English translation results. Batch size 64 can give better BLEU score than other batch sizes from our experimental results. However, increasing training epoch doesn't have much effect on BLEU score.

Keywords: Neural machine translation, recurrent neural network, attention mechanism, syllable segmentation, normalization, rule-based approach, semi-automated approach, Myanmar language

1 Introduction

Different people use different languages all over the world. The path to bilingualism, or multilingualism, can often be a long, never-ending one. In these cases, translation helps people to overcome language barriers and surpass international boundaries. As translation becomes important, the demands for professional translators are growing. Some of the translation works are difficult and challenging for normal person who have basic knowledge on languages. But much of it is tedious and repetitive. Automatic or machine translation system is proposed to use both as translation memories to remember key terms and as the fastest way to translate context into a new language.

Machine translation is the task of automatically converting text from one source language into another target language. Given a sequence of text in a source language, there is no one single best translation of that text to another language. Due to natural ambiguity and flexibility of human language, machine translation is one of the most challenging artificial intelligence tasks in natural language processing. Machine translation becomes one of the important research areas and translation for Myanmar language turns out to be active demanding problem.

Machine translation can be done with approaches like rule-based, phrase-based, hybrid and neural which is reviewed in machine translation literature review paper [1]. Although statistical approaches are used for machine translation before, neural machine translation becomes popular after achieving state-of-the-art results with deep neural networks in various fields. The main reason neural machine translation influences over statistical machine translation is that it can give more fluent translation results because neural machine translation system considers entire sentences. As neural machine translation system is jointly trained as a single system, it can learn complex relationships between languages than statistical machine translation systems whose translation, reordering and language models are independently trained. Zhongyuan Zhu evaluates neural machine translation on English to Japanese task and it surpasses all statistical machine translation baseline models [2].

In this paper, neural machine translation system from Myanmar to English language is proposed. FastText word embedding is used for better capturing of word similarities. Neural machine translation depends on neural network models to develop statistical models for the purpose of translation so that two recurrent neural networks are used as encoder and decoder architecture.

The remainder of the paper is organized as follows. In Section 2, an introduction to Myanmar language is described and the grammar structure of Myanmar and English languages is discussed. The background theories of this paper are provided in section 3. The proposed system is discussed in section 4. Section 5 provides evaluation method for measuring accuracy of translation output. In section 6, training details and results are presented. Section 7 summarizes the note and describes future work.

2 Myanmar Language Script and Its Grammar Structure

The Myanmar script is an abugida (alpha syllabary) in the Brahmi family. It is composed of 33 consonants, 11 basic vowels, 11 consonant combination symbols, and extension vowels, vowel symbols, devowelizing consonants, diacritic marks, specified symbols and punctuation marks.

Myanmar has mainly 9 parts of speech: noun, pronoun, verb, adjective, adverb, particle, conjunction, post-positional marker and interjection [3] [4]. The words in Myanmar language can be defined as follows: simple words, complex words , compound words and loan words [3] [4] [5].

Myanmar language mostly use SOV grammatical arrangement and sometimes use OSV format. Because of its use of postposition (wi.Bat), it can be defined as postpositional language whereas English, which uses its syntax as SVO, can be defined as prepositional language. Myanmar language is also a free word order language for everyday speaking. The lack of regularities in sentence structure leads to very complex Myanmar language processing in order to obtain satisfactory translation results [6].

3 Background Theories

3.1 Neural Machine Translation

Neural machine translation is a newly emerged approach to machine translation. It attempts to build and train a single, large neural network which can be called as end-to-end translation. The system reads a sentence and outputs a correct translation. It consists of two components named encoder and decoder, both of which are built with recurrent neural networks. Encoder encodes a variable length source sentences into a fixed length context vector and decoder decodes that vector into a variable length target sentences.

3.2 Recurrent Neural Network

Recurrent neural network can be used in many applications like image classification, sentiment analysis, video classification, image captioning. However, it is the analogous neural network for text data. Recurrent neural network is designed for sequential data and applies the same function to the words or characters of the text. A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor. Recurrent neural network remembers the past and its decisions are influenced by what it has learnt from the past.

4 Proposed System

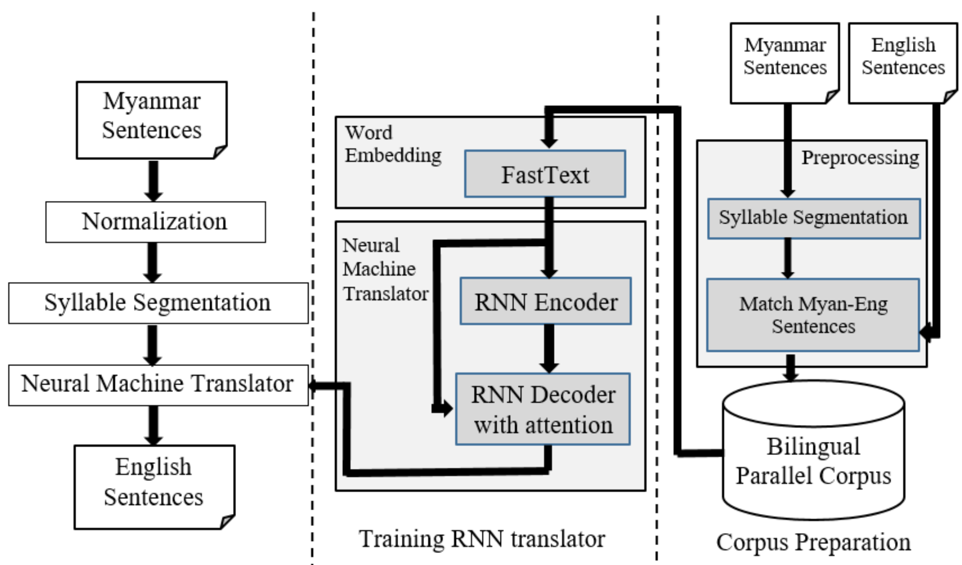

Fig. 1. Proposed System Process Flow

Firstly, normalization is performed on system's input which is Myanmar sentences. Then, Syllable segmentation is done on those normalized sentences before using neural machine translator. Translator outputs English sentences processed from segmented Myanmar sentences at last.

Bilingual parallel corpus is prepared to use in translation system. Sentences from parallel corpus are converted to vectors by using fastText pretrained model at word embedding stage. And then, neural machine translator is trained by using RNN encoder decoder architecture with attention mechanism. Overview of proposed system flow is shown in Fig. 1.

4.1 Normalization

Normalization is a process that transforms a list of words to a more uniform sequence. In normalization process for Myanmar to English translation, Myanmar sentences are firstly detected and converted to Unicode format for further processing. Myanmartools is used to detect language format and conversion is done on rule-based approach. Semi-automated approach is used to normalize Myanmar sentences. After converting from Zawgyi to Unicode format in Myanmar language, unseen code points and repeated characters are removed by using rule-based approach and some of the wrong spellings are corrected manually.

4.2 Syllable Segmentation

For Myanmar language, spaces are used for better reading but the language doesn't have standard format for breaking words or syllables like English. Further rules or algorithms are needed to break Myanmar words as well as syllables. Syllable segmentation is the ability to identify how a syllable can be written in a language. Rule-based syllable breaking is performed to separate Myanmar syllables in this paper.

Segmentation rules are created based on the characteristics of Myanmar syllable structure and they follow the rules of Myanmar Unicode consortium. All six rules are shown in Table 1.

4.3 Corpus Preparation

To train Myanmar - English neural machine translator, bilingual parallel corpus is required. In this paper, Myanmar sentence aligned corpus from MCF NLP is used. It has more than 100,000 English-Myanmar parallel sentences. As a preprocessing step, syllable segmentation is performed for Myanmar language. Only sentences with less than 50 syllables are used to train translation model. Finally, Bilingual parallel corpus is built by matching segmented Myanmar sentences to corresponding English sentences.

4.4 Word Embedding

Word embedding is a vector form of word representations that bridges the human understanding of language to that of a machine. They are capable of capturing

Table 1. Syllable Segmentation Rules

Rules	Explanation
Single character rule (R1)	A character can be defined as a syllable in Myanmar language
Special ending character rule (R2)	Some characters which represent the end of a syllable can be defined as a syllable
Second consonant rule (R3)	When a syllable has two consonants, the second consonant should come with either Athat (Killer) or Htutsint (Kinzi)
Last character rule (R4)	Last character in a sentence, a phrase or input file can be regarded as the end of a syllable
Next starter rule (R5)	This rule breaks up the syllable when it sees 'tha way htoe' appearing after a complete syllable
Miscellaneous rule (R6)	This rule covers breaking of numbers, special characters and non-Myanmar characters

context of a word in a document, syntactic and semantic similarity, relations with other words, etc. Different word embedding techniques such as GloVe, Word2vec and fastText are used to learn features of text in Natural Language Processing.

FastText is created by Facebook AIs research lab and uses shallow neural network for word embedding. It supports training both continuous bag of words (CBOW) and skip-gram (SG) models. The thing that FastText incoporate sub-word information allows it to support out-of-vocabulary words which is known from fastText paper [7]. FastText can give meaning to some unknown words because semantic knowledge of the sub-word can help provide a bit more semantic information to that unknown word.

FastText has pretrained word vectors for 294 different languages which includes also for Myanmar language. Pamela Shapiro pointed that pretrained models are useful for low-resoure languages like Myanmar to improve neural machine translation results [8]. So, FastText pretrained model is used in this paper for embedding both Myanmar and English.

4.5 RNN Encoder-Decoder

RNN encoder - decoder architecture is the standard neural machine translation method which encodes a variable length input sentences into vector representations and then decodes back into output translations. In the RNN encoder decoder framework, an encoder reads the input sentence, a sequence of vectors m = $m_1, ..., m_T$ into a vector v.

$$h_t = f\left(m_t, h_{t-1}\right) \quad (1)$$

$$v = q\left(h_1, ..., h_{Tx}\right) \quad (2)$$

where h_t is a hidden state at time t, and v is a vector generated from the sequence of hidden states. F and q are nonlinear functions.

The decoder is trained to predict the next word $e_{t'}$ given the context vector v and all the previously predicted words $e_1, ..., e_{t'-1}$.

$$p\,(e) = \prod_{t=1}^{T} p\,(e_t \mid \{\, e_1, ..., e_{t-1}\}, v) \tag{3}$$

context vector v is computed as a weighted sum of annotations h_i, to which an encoder maps the input sentence and weights α_{ij}.

$$v_i = \sum_{j=1}^{T_x} \alpha_{ij}\, h_j \tag{4}$$

4.6 Attention Mechanism

Long term dependencies can be difficult to learn in RNNs because input sequences are encoded to one fixed length vector. Attention mechanism can solve one apparent disadvantage which is incapability of the system to remember long term dependencies. It also helps alignment problem that identifies which parts of the input sequence are relevant to each word in the output. Minh-Thang Luong shows that attention-based NMT models are superior to non-attentional ones in many cases such as translating names and handling longer sentences [9].

5 Evaluation

Various evaluation methods like Word Error Rate (WER), Translation Error Rate (TER) and Bilingual Evaluation Understudy (BLEU) are used to measure the results of machine translation. Among them, BLEU, which is de-facto standard in measuring translation output, will be used in this paper. It works by counting n-grams in the generated sentence to n-grams in the reference sentence. Comparison of matching is made regardless of word order [10]. BLEU uses 4-grams measure to calculate scores and its equation is as follows:

$$BLEU = BP\, \cdot\, exp\,(\sum_{n=1} N\, w_n\, log\, p_n) \tag{5}$$

where BP is brevity penalty, w_n is positive weights summing to one and p_n is modified n-grams precision. Brevity penalty can be computed as:

$$BP = \begin{cases} 1, & \text{if } r < c \\ e^{(1-\frac{r}{c})}, & \text{if } r \geq c \end{cases} \tag{6}$$

where r is the length of reference (actual, human) translation and c is the length of candidate (model, predicted) translation. P of each gram can be calculated as:

$$p_n = \frac{\sum_{C\, \epsilon\, \{Candidates\}} \sum_{n-gram\, \epsilon\, C} Count_{clip}\,(n-gram)}{\sum_{C'\, \epsilon\, \{Candidates\}} \sum_{n-gram'\, \epsilon\, C'} Count_{clip}\,(n-gram')} \tag{7}$$

6 Training Details and Results

GRU memory cell are used for RNN with attention mechanism. As an optimization setup, adaptive optimizer, Adam with softmax cross entropy loss is applied to the network. Teacher forcing method is used during training to use actual output as input to next time step. Attention score is computed with tanh function and attention weights are calculated with softmax activation function. With embedding units 1024 and embedding dimension 256, model with three different batch sizes 32, 64 and 128 are compared for RNN. All experiments are run on NVIDIA tesla K80 with 12GB RAM (google colab).

20,000 and 100 parallel sentences are used for training and testing respectively. There are 1,987 syllable vocabulary for Myanmar language and 14,236 for English. The model is trained for 10 epoches. BLEU score results for RNN with batch sizes 32, 64 and 128 for Myanmar to English translation are shown in Table 2.

Table 2. BLEU score for Myanmar to English translation

Batch size	RNN (myan-eng)
32	0.752121
64	0.775993
128	0.675887

Best score is achieved by batch size 64. So, further experiments on epochs changing is done only with batch size 64. Loss is reduced per epoch in training time which can be seen in Fig. 2.

Although loss decreases every epoch, BLEU score per epoch don't arise significantly. The best score 0.776 is obtained by training over 10 epoches.

7 Conclusion and Future Work

In this paper, sequence to sequence translation of Myanmar to English languages is proposed. Syllable segmentation is done as preprocessing for Myanmar sentences. Processed Myanmar sentences are matched with corresponding English sentences to prepare bilingual parallel corpus.

That corpus is used to train translation system. The system applies two recurrent neural networks as encoder-decoder framework. As grammatical structures of Myanmar and English languages are different, direct translation cannot help improve translation accuracy. Attention mechanism is used to handle long term dependencies reordering problem. The proposed system results the best BLEU score of 0.776 with batch size 64 by training for 10 epoches.

As future works, English to Myanmar or bidirectional translation system can also be built based on the theories of this paper. As neural machine translation on other languages are also experimented in various researches, multilingual

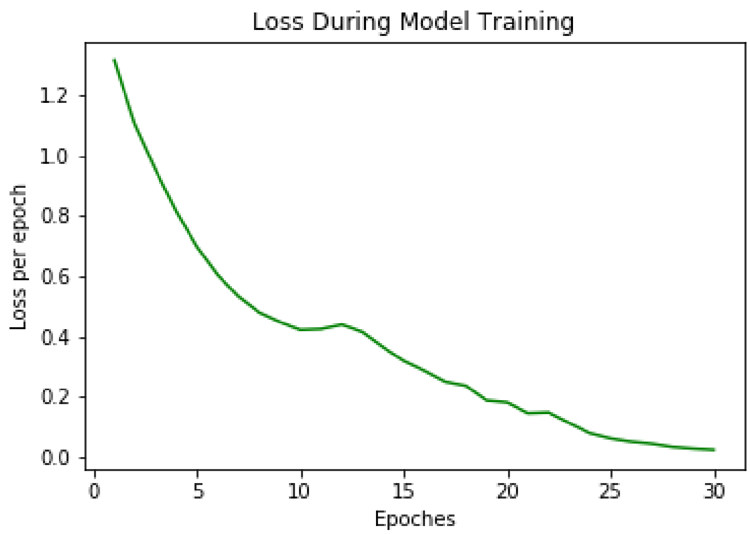

Fig. 2. Loss per epoch in Model Training

translation system can be trained by combining multiple bilingual systems as a single system. The BLEU score of proposed system can also be improved by considering other linguistic features like Part Of Speech or using transfer learning in the training of network.

Acknowledgements It is hard to find words to express my gratitude to Myanmar Computer Federation for giving open source Myanmar-English parallel corpus. I would also like to thank my supervisor with kindest personal regards for guiding me warmly.

References

1. Ankush Garg, Mayank Agarwal, Machine Translation: A literature Review, arXiv: 1901.01122v1 [cs.CL], (2018)
2. Zhongyuan Zhu, Evaluating Neural Machine Translation in English-Japanese Task, WAT (2015)
3. Myanmar Grammar, Department of the Myanmar Language Commission, Ministry of Education, Union of Myanmar (2005)
4. Adoniram Judson, Grammatical Notices of the Burmese Langauge, American Baptist Mission Press, (1842)
5. Tun Tint, Features of Myanmar language, (2004)
6. Win Win Thant, Tin Myat Htwe, Ni Lar Thein, Syntactic Analysis of Myanmar Language (2011)
7. Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov, Bag of Tricks for Efficient Text Classification, arXiv:1607.01759v3 [cs.CL] (2016)

8. Pamela Shapiro, Kevin Duh, Morphological Word Embeddings for Arabic NMT in Low-Resource Settings, Proceedings of the Second Workshop on Subword/ Character Level Models, New Orleans (2018)
9. Minh-Thang Luong, Hieu Pham, Christopher D. Manning, Effective Approaches to Attention-based Neural Machine Translation, Proceedings of the 2015 Conf on Empirical Methods in NLP, Portugal (2015)
10. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, BLEU: A Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) (2002)
11. Qiang Li, Derek F. Wong, Lidida S. Chao, Muhua Zhu,Tong Xiao, Jingbo Zhu, Min Zhang, Linguistic Knowledge Aware Neural Machine Translation, IEEE/ACM Transactions and audio, speech, and language processing (2018)
12. Andrei Popescu-Belis, Context in Neural Machine Translation: A Review of Models and Evaluations, arXiv: 1901.09115v1 [cs.CL] (2019)
13. Juyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio, Empirical Evaluation of Gated Recurrent Neural Network on Sequence Modeling, arXiv:1412.3555v1 [cs.NE] (2014)
14. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention is All You Need, 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA (2017)
15. Yukio Matsumura, Takayuki Sato, Mamoru Komachi, English-Japanese Neural Machine Translation with Encoder-Decoder-Reconstructor, arXiv (2017)
16. Philipp Koehn, Rebecca Knowles, Six Challenges of Neural Machine Translation, Proceedings of the first workshop on NMT (2017)
17. Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer, Deep Contextualized Word Representations, Proceedings of NAACL-HLT 2018, New Orleans, Louisiana (2018)
18. Ye Qi, Devendra Singh Sachan, Matthieu Felix, Sarguna Janani Padmanabhan, Graham Neubig, When and Why are Pre-trained Word Embeddings Useful for NMT?, Proceedings Conference of the North American Chapter of the Association for Computational Linguistics (2018)
19. Jiatao Guy, Hany Hassanz, Jacob Devlin, Victor O.K. Li, Universal NMT for Extremely Low Resource Languages, Conference of the North American Chapter of the Association for Computational Linguistics (2018)
20. Pamela Shapiro, Kevin Duh, Morphological Word Embeddings for Arabic NMT in Low-Resource Settings, Proceedings of the Second Workshop on Subword/ Cha Level Models, New Orleans (2018)
21. Eric Greenstein, Daniel Penner, Japan to English Machine Translation Using Recurrent Neural Network, Stanford University (2015)
22. Andi Hermanto, Teguh Bharata Adji, Noor Akhmad Setiawan, Recurrent Neural Network Language Model for English-Indonesian Machine Translation: Experimental Study, Conference on Science in IT (ICSITech) (2015)
23. IIya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to Sequence Learning with Neural Network, Google, NIPS' Proceedings of the 27th International Conference on Neural Information Processing Systems (2014)