Neural Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)

Thazin Myint Oo [‡]   Ye Kyaw Thu [†]   Khin Mar Soe [‡]

[‡] Natural Language Processing Lab., University of Computer Studies Yangon, Myanmar

[†] Language and Semantic Technology Research Team (LST), NECTEC, Thailand
Language and Speech Research Research Lab., Waseda University, Japan

Outline

- Abstract
- Introduction
- Rakhine and Myanmar Language
- Corpus Preparation
- Segmentation
- Experimental Methodology
- Experiments
- Results and Discussion
- Conclusion

Abstract

- Explore the neural machine translation (NMT) between Myanmar (Burmese) and Rakhine (Arakanese)
- Developing Myanmar-Rakhine parallel corpus was used
- Experiments were carried out using three different NMT approaches:
  - Recurrent neural network (RNN),
  - Transformer
  - Convolutional neural network (CNN)

Abstract (con't)

- two different segmentation schemes for word embedding were studies,
  - Word-BPE
  - Syllable-BPE
- Make a comparison between Statistical Machine Translation (SMT) and NMT for Myanmar-to-Rakhine and Rakhine-to-Myanmar Machine Translation (MT)
- The highest quality NMT and SMT performances are obtained with Syllable-BPE segmentation for both types of Translations.
- Focusing on NMT , transformer with Word-BPE segmentation outperforms CNN and RNN for bi-directional translation
- CNN with Syllable-BPE segmentation obtains a higher score than the RNN and transformer

Introduction

- Myanmar language includes a number of mutually intelligible Myanmar dialects, with a largely uniform standard dialect used by most Myanmar standard speakers
- MT has so far neglected the importance of properly handling the spelling, lexical and grammar divergences among language varieties.
- **Contrubutions**

  - corpus building (Myanmar-Rakhine parallel corpus)
  - the 1st NMT results of three neural network architectures (RNN, Transformer, CNN)
  - presenting appropriate hyper-parameters (batch size, learning rate, cell type and activation function)
  - Now, we can make a comparison between SMT and NMT for my-rk, rk-my MT

Rakhine State

Rakhine Language

•Rakhine language is largely monosyllabic and analytic language, with a Subject Object Verb (SOV) word order and uses the Myanmar script.

•It is considered by some to be a dialect of the Myanmar language.

•Comparing with Myanmar language, the speech of the Rakhine language is likely to be closer to the written form. Rakhine language notably retains an /r/ sound that has become /j/ in Myanmar language. Rakhine pronounce the medial “ျ” as “Yapint” (i.e. /j/ sound) and the medial “ြ” as “Rayit" (i.e. /r/ sound)

•differences between Rakhine and Myanmar languages are in their pronunciations and their vocabularies.

my: လုံချည် တစ် ထည် ဘယ်လောက်လဲ ။
rk: ဒယော တစ် ထည် ဇာလောက်လေး ။
(“How much for a longyi?” in English)

rk : ကလေချေ တိ ဘောလုံး ကန် နေတယ် ။
my: ကောင်လေး တွေ ဘောလုံး ကျောက် နီရေ ။
(“Boys are playing football” in English)

rk : ဇာ ပြော နီချင့် ယင်းသူရို့ ။
my : သူတို့ ဘာ ပြော နေတာလဲ ။
(“What are they talking about” in English)

rakhine : အဘောင်သျှင် စျီး က သပုံ ဝယ် လာတယ် ။
myanmar: အဘွား စျေး က ဆပ်ပြာ ဝယ် လာတယ် ။
(“Grandmother buy soap from market” in English)

Corpus Preparation and statistics

- Currently, there is no parallel corpus for MSL.
- used 18,373 Myanmar sentences (without name entity tags) of the ASEAN-MT Parallel Corpus
- Contains six main categories: people, survival, food, fun, resource and nightlife
- Manual translation into the Rakhine language was done by native Rakhine students form two universites and the translated corpus is checked by the editor of a Rakhine newspapers.
- Word segmentation for Rakhine was done manually and there are exactly 123,018 words in total
- 14,076 sentences for training
- 2,485 sentences for development
- 1,812 sentences for evaluation

Segmentation

- Rakhine word“စား ဗျာယ်”, “စား ပီးဗျာယ်”, “စား ဖို့ဗျာယ်”. Here, “စား” (“eat” in English) is a root word and the others are suffixes for past and future tenses.
- Rakhine word “ကလိန့်မေချေ တိ” (ladies) is segmented as two words “ကလိန့်မေချေ” and the particle “တိ”.
- Rakhine compound word “ဖေ့သာ + အိတ်” (“money” + “bag” in English) is written as one word “ဖေ့သာအိတ်” (“wallet” in English).
- Rakhine adverb words such as “အဂယောင့်” (“really” in English), “အမြန်” (“quickly” in English) are also considered as one word.

Syllable Segmentation

- Generally Myanmar word composed of multiple syllable.
- Focus on consonant-based syllables, the structure of the syllable can be described with Backus normal form (BNF) as follow:

Syllable := *CMW*[*CK*][*D*]c

we used RE-based Myanmar syllable segmentation tool named "sylbreak."

Byte Pair Encoding Segmentation

- (Sennrich et al.,2016) proposed a method to enable open-vocabulary translation of rare and unknown words as a sequence of subword units representing BPE algorithm
- new word “rocket,”ဒုံးပျံ may be segmented as ဒ@@ ုံး ပျံ after looking up the learnt vocabulary, assuming ဒ and ုံး ပျံ as BPE units learnt during the training.

Experimental Methodology

- **Phrase-based Statistical Machine Translation (PBSMT)** – [Kohen et al., 2003]
- **Encoder-Decoder Models for NMT**
  - **Stacked Recurrent Neural Network (RNN) with Attention** - [Schwenk, 2012, Kalchbrenner and Blunsom, 2013, Sutskever et al., 2014, Bahdanau et al., 2014, Luong et al., 2015]
  - **Self-attentional Transformer** - [Vaswani et al., 2017]
  - **Fully Convolutional Models (ConvSeq2Seq)** - [Gehring et al., 2017]

Experiments

Moses SMT System

•used the Phrase-based SMT provided by the Moses toolkit [Kohen et al., 2003]

Framework for NMT

•used the Sockeye [Hieber et al., 2017] toolkit, which is based on MXNet [Chen et al., 2015] to train NMT models

Training Details

•initial learning rate is set to 0.0001 to 0.0005

•Drop out rate of embedding Transformer is 0.1 and RNNs is 0.2

•size of embeddings and hidden states is 512

•trained all models for maximum epoch using the SGD, Adagrad and Adam optimizers

•size of the training batches were set to 128,256 and 512

•byte pair encoding (BPE) models were trained with a vocabulary size of 8,000

•all experiments are run on NVIDIA Tesla K80 24GB GDDR5.

Result and discussion

Batch Size	RNN		Transformer		CNN	
	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my
128	79.86	81.44	79.64	82.01	**80.82**	**83.59**
256	80.76	82.94	79.47	81.37	80.33	83.54
512	80.00	82.26	79.47	80.79	79.86	81.38

Table 1. : BLEU scores of Syllable-BPE segmentation with different batch sizes for three NMT models

CNN achieved highest score

Result and discussion

Batch Size	RNN		Transformer		CNN	
	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my
128	60.02	44.44	72.70	72.82	69.03	72.24
256	60.31	46.47	**73.39**	72.45	65.61	68.26
512	42.76	34.93	73.30	**72.95**	67.89	71.68

Table 2: BLEU scores of Word-BPE segmentation with different batch sizes for three NMT models

Transformer architecture achieved top score

Result and discussion

Learning rate	RNN				Transformer			
	GRU		LSTM		GRU		LSTM	
	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my
0.0001	79.47	81.37	79.48	80.88	80.76	82.94	80.26	83.02
0.0002	79.82	81.65	**82.85**	82.07	80.88	81.54	80.90	82.99
0.0003	80.22	82.23	80.24	82.13	80.92	82.63	81.78	83.30
0.0004	80.65	82.66	80.85	82.33	81.25	82.54	**81.92**	**84.06**
0.0005	80.41	81.46	81.98	**83.86**	80.57	82.30	80.65	82.51

Table 3: BLEU scores for batch size 256 of Syllable-BPE segmentation with different learning rates and two memory cell types on RNN and the transformer

LSTM gave the highest NMT performance

Result and discussion

Learning rate	Batch Size 128				Batch Size 256			
	ReLu		Soft-ReLu		ReLu		Soft-ReLu	
	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my
0.0001	**81.37**	83.29	80.00	81.97	80.26	81.97	80.03	81.08
0.0002	81.01	82.24	79.89	82.50	80.07	82.29	80.01	81.51
0.0003	80.99	81.59	80.11	**83.34**	81.16	81.69	82.14	**84.08**
0.0004	N/A	N/A	N/A	N/A	79.74	80.87	**83.75**	83.06
0.0005	N/A	N/A	N/A	N/A	79.05	82.43	81.44	83.31

Table 4: BLEU scores for batch sizes 128 and 256 of Syllable-BPE segmentation with different learning rates and two activation functions on CNN

Soft-Relu achieved highest score

Result and discussion

Segmentation	OSM		RNN		Transformer		CNN	
	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my	my-rk	rk-my
Syllable-BPE	82.71	**84.36**	82.03	83.98	82.85	82.65	**83.75**	84.08
Word-BPE	77.12	75.27	60.31	46.47	73.39	72.95	69.03	72.24

Table 5: Comparison of SMT and NMT (top BLEU scores) on two segmentation schemes

CNN achieved top BLEU score in Myanmar to Rakhine MT.

Conclusion

- the first study of the neural machine translation between Standard Myanmar and Rakhine (a spoken Myanmar dialect)
- implemented three NMT systems (RNN, Transformer and CNN) with our developing Myanmar-Rakhine parallel text
- investigated word segmentation schemes (Word-BPE and Syllable-BPE).
- The highest quality NMT and SMT performances are obtained with Syllable-BPE segmentation for both types of Translations.
- Focusing on NMT , transformer with Word-BPE segmentation outperforms CNN and RNN
- for bi-directional translation
- CNN with Syllabel-BPE segmentation obtains a higher score than the RNN and transformer

Future Work

- We plan to extend our study with focus on NMT models with one more subword Segmentation scheme sentencePiece for Myanmar-Rakhine NMT.

- intend to investigate SMT and NMT approaches for other Myanmar dialect languages, Such as Myeik and Dawei.

Thank you