OPTIMIZE THE LEARNING RATE OF NEURAL ARCHITECTURE IN MYANMAR STEMMER

Yadanar Oo and Khin Mar Soe

Natural Language Processing Lab, University of Computer Studies, Yangon, Myanmar

ABSTRACT

Morphological stemming becomes a critical step toward natural language processing. The process of stemming is to reduce alternative forms to a common morphological root. Word segmentation for Myanmar Language, like for most Asian Languages, is an important task and extensively-studied sequence labelling problem. Named entity detection is one of the issues in Asian Language that has traditionally required a large amount of feature engineering to achieve high performance. The new approach is integrating them that would benefit in all these processes. In recent years, end-to-end sequence labelling models with deep learning are widely used. This paper introduces a deep BiGRU-CNN-CRF network that jointly learns word segmentation, stemming and named entity recognition tasks. We trained the model using manually annotated corpora. State-of-the-art named entity recognition systems rely heavily on handcrafted feature built in our new approach, we introduce the joint model that relies on two sources of information: character level representation and syllable level representation.

KEYWORDS

Myanmar word stemmer, Sequence labelling, Conditional random fields, Neural architecture, word segmentation

1. INTRODUCTION

Myanmar Language is characterized by its rich and complex morphology based on root pattern schemes. Morphological stemming is one of the most essential topics in natural language processing applications such as information retrieval, text summarization, machine translation, etc. Word segmentation is the task of deciding word boundaries in a segmented text. In the English language, word boundaries are easily determined because of the presence of white spaces or punctuation between words. In Myanmar Language, segmenting sentences into words is an important task because sentences are clearly defined by a sentence boundary marker but words are not always delimited by spaces. Spaces may sometimes be added between words and even between a root word and the associated post-position. It is because there are no indicators such as blank spaces to show the word boundaries in Myanmar text. The same phenomenon does not happen only to Myanmar language but also many other Asia languages such as Japanese, Chinese, and Thai.

Therefore, in order to understand the Myanmar text, the first thing that we need to do is to cut the sentences into word segments. Although it sounds easy to cut a sentence into a word sequence, however from past experience, we know that it is not a trivial task. During the process of Myanmar word segmentation, two main problems are encountered: segmentation ambiguities are dealt with known words, i.e. words found in the dictionary. An unknown word is defined as a word that is not found in the system dictionary. In other words, it is an out-of-vocabulary word. For any language, even the largest dictionary will not be capable of registering all geographical names, person names, organization names, technical terms and some duplication words, etc. This paper presents an approach of morphologically extract Myanmar root word,

DOI: 10.5121/ijnlc.2019.8504

through the removal of prefixes and the suffixes but word segmentation is required for Myanmar Language because word boundaries are not indicated by white spaces.

Normally, segmentation is considered as a separate process from stemming and named entity recognition. In our approach, we implement word segmentation, stemming and named entity recognition as a joint process. We introduce a deep BiGRU-CNN-CRF network that jointly learns word segmentation, stemming and named entity recognition tasks. We trained the model using a manually annotated corpus. The NCRF++ toolkit [9] was used to build neural sequence labelling architecture for the joint process of Myanmar word. Our main contributions are (i) proposing a neural model that jointly extracts stem word, detect word boundary and detect named entity (ii) giving empirical evaluations of this model on a different configuration. (iii) we introduce the BiGRU-CNN-CRF network for Myanmar morphological stemming.

In order to model the character sequence information of a syllable, Bi-directional Gated Recurrent Unit encodes the character sequence of each syllable and concatenates the left-to-right and right-to-left as character sequence representations. Convolutional neural networks (CNNs) have shown its great effectiveness to extract morphological information such as prefix and suffix of a word. For sequence labelling tasks, the interaction between labels in surroundings is considered and jointly decode the best chain of labels for a given input sentence. For example, in this approach of stemming suffix words are absolutely followed by a root word and standard BIO annotation I-R cannot follow I-Suf. Therefore, an inference layer with a linear-chain Conditional Random Field (CRF) is used. This classifier is beneficial for tasks with strong dependencies between token tags.

The remainder of the paper is organized as follows. Section 2 describes the Myanmar language formation. Section 3 describes the literature review. Section 4 reports the neural sequence labelling model. Section 5 proposed system architecture. Section 6 explains the task specifications. Section 7 discusses the experimental results. Section 8 draws conclusions and outlook.

2. FORMATION OF MYANMAR LANGUAGE

The Myanmar language, Burmese, belongs to the Tibeto-Myanmar language group of the Sino-Tibetan family. It is also a morphologically rich and agglutinative language. Myanmar words are postpositionally inflected with various grammatical features [2]. In the Myanmar language, there is no white space between words and words are difficult to define. Normally, to produce the stem word or named entity, the word segmentation task is a pre-processing stage of stemming. Segmentation is considered as a separate process from stemming.

The basic order of the Myanmar languages is subject-object-verb. There are nine Part-of-Speech classes for all Myanmar words. These are Noun, Pronoun, Verb, Adjective, Adverb, Conjunction, Postpositional Marker, Particles and Interjection. Noun is the content word that can be used to refer a person, place, thing. Noun is the stem word in a sentence. “ရေအိုးစင် ” [water pot], “ နိုင်ငံရေးသမား ” [politician], “ ဘောလုံးသမား ” [footballer] etc. Noun in Myanmar language can be combined with particles to form the plural by suffixing the particle “တွေ” [-twe], “များ ” [-myar]. Noun can also suffix with “များ ” [-myar], “တိုင်း ” [-tine]. e.g., “ အသင်းကြီး ” [club], “ နေရာတိုင်း ” [place]. In the word “ မြို့များ” [cities], the stem word is “ မြို့ ” [city] and “များ ” [-myar] is the suffix. The word “ဝင်လာသူ ” [comer] is the noun and they are also the stem words in Myanmar language.

Stem verb is always suffixed with at least one particle to form a tense politeness, mood, etc. The stem verb remains unchanged when they have the particle suffixed to them. For instance, “စုံဖြူးဆဲ

” [developing], “ ဖွံ့ဖြိုးဗိုး ” [developed] have different verb particles. They have the same stem verb “ ဖွံ့ဖြိုး ”[develop]. Verb is negated by the particle “ မ ”[-ma], which is a prefix to the verb to form the negative verb. Some verbs also negated by particle and also have suffix but they also unchanged the meaning of the stem verb. These verbs are between the particle “ မ ” [not] and “ ဘဲ ” [-bal] , “ မ ” [not] and “ ခင် ” [khin]. For example, “ မပြုလုပ်ဘဲ ” [not do] and “ မရှိရှာ ” [not have].

Adjective is used to modify the noun. Myanmar adjectives can be formed by combining verbs and particles. For example, “ပေါ်ထွက်လာခဲ့သော” [appeared] is the adjective that combines the verb “ပေါ်ထွက် ” [appear] and adjectives suffix “ခဲ့သော” [-kae thaw] . Some of the adjective are combined with “အ” [-a] “ဆုံး ” [-sone], for example, “ အကြီးဆုံး ” [the heaviest] and “ အကြီးအမြတ်ဆုံး ” [the nobel]. A word that modifies the verb is an adverb.

Myanmar adverb is always before the verb and there can be more than one adverb for one verb. Adverb also has suffix “စာ” [-swar]. Their stem form remains unchanged when suffix removal. Reduplication occurs in Myanmar sentences and most of the reduplicated words are Adverbs and their stem forms are Adjective. Many Myanmar words, especially adjectives or verbs with two syllables, such as “ ရိုးသား ” [honest], “ယုံကြည် ” [believe] can be reduplicated as “ ရိုးရိုးသားသား ” [honest], “ယုံယုံကြည်ကြည် ” [believe]. In this reduplication case, our approach cannot segment when one particle mixes with a verb and an adverb form “ အ ” [-a] “ အ ” [-a] “အပြေးအလွှား ” [scurry], “တ ” “တ ” “တရင်းတနှီး ” [familiar], “အ ” “တ ” “ အရောတဝင် ” [intimate], etc. Particles are words serving quality nouns, pronouns, adjectives, verbs, and adverbs. Some of the particles are used as type classifiers, for example “ ၃၉ယောက် ” [39 persons], “တစ်ဂိုး ” [one goal]. Some are used as a numerical modifier “ ခြောက်ကြိမ်မြောက် ” [sixth time], “ ဆယ်သိန်းကျော် ” [ten lakh]. Post positional marker is used to indicate time, mood, object and subject “ ၃ရက်တိုင်တိုင် ” [3 days], “ ဥပဒေနှင့်အညီ ” [abiding the law]. Myanmar conjunction is used to connect words, phrases or clauses. “ ကဲ့သို့သော ” [as], “ပြီးနောက် ” [after], “ထို့ပြင် ” [therefore]. Interjection expresses sudden emotions which may find utterance in expressions of feeling is one of admiration, delight, dislike, angry or desire, etc. “ အလို့လေး ” [oh], “ အောင်မလေး ” [uh].

Named Entities (NEs) have a unique status in Natural Language Processing (NLP) and they are not found in the dictionary or lexicon. In Myanmar Language, proper nouns are person name, city name, organization name, country name, etc. Proper nouns are tagged with NE (Named Entity). Proper nouns cannot combine with suffix or prefix. But, some of the proper nouns have the suffix. For example, “တရုတ်များ” means many [Chinese]. “ တရုတ် ” [Chinese], that combine with “များ ”[-myar]. For the person name, “ ဦးမြင့်ဦး ”[U Myint Oo], “ ဦး ”[U] is the prefix of the name “ မြင့်ဦး ”[Myint Oo]. It means that “ မြင့်ဦး ”[Myint Oo] is the male. In Myanmar Language, prefix “ ဦး ”[U] and “ ဒေါ် ”[Daw], “ ကို ”[Ko] and “ မ ”[Ma] is used to separate male or female. Moreover, it shows that the person name and it cannot separate “ ဦး ”[U] and “ မြင့်ဦး ”[Myint Oo]. Some numbers can be name entity, for example, “ ၃၃လမ်း ” [33th street]. In this case, 33 is not just a number. It is the name of the street. Proper name can also exist in front of “ အဖွဲ့ ” [organization], “ ဦးစီးဌာန ” [Department], “ မြို့ ”[city]. e.g. “ ကမ္ဘာ့ကျန်းမာရေး အဖွဲ့ ” [World Health Organization].

3. LITERATURE REVIEW

Word embedding has been very progressive in recent years at improving performance across a variety of NLP tasks. Word vector pre-trained on large text corpora have been released on [10] "Learning Word Vectors for 157 Languages " that trained on 3 billion words from Wikipedia and Common Crawl using Continuous bag-of-words (CBOW) 300-dimension. To train the word vector, they use Skip-gram and CBOW models.

In [8] W.P.Pa, N.L.Thein, February 2008, "Myanmar Word Segmentation Using Hybrid Approach" Word Segmentation system consists of four components, sentence splitting, tokenization, initial segmentation by Maximum Matching Algorithm and statistical combined model (bigram model and modified word juncture model) for final segmentation.

In [9], Ye Kyaw Thu, Win Pa Pa, Andrew Finch, "Word Boundary Identification for Myanmar Text Using Conditional Random Fields". Conditional random field is used to identify Myanmar word boundaries within a supervised framework. CRF approach is compared against a baseline based on maximum matching using a dictionary from Myanmar Language Commission Dictionary (word only) and the manually segmented subset of the BTEC1 corpus.

In recent research literature, neural models can be challenging. In (Jie Yang, Shuailong Liang and Yue Zhang, 12 Jul 2018), "Design Challenges and Misconceptions in Neural Sequence Labelling" explored three neural model designs: character sequence representation, word sequence representation, and inference layer. Experiments show that character information improves model performance. In our approach, such joint work is performed as a syllable-based neural sequence labelling architecture.

In [13], Zhenyu Jiao, Shuqi Sun, Ke Sun proposed Chinese Lexical Analysis with Deep Bi-GRU-CRF Network. They introduced a deep Bi-GRU-CRF network that jointly models word segmentation, part-of-speech tagging and named entity recognition tasks. Their main purpose was jointly accomplishing three tasks. The model worked in a full end-to-end manner and it is effective and efficient.

In [14], they proposed a character-based model for joint segmentation and POS tagging for Chinese that use bidirectional RNN-CRF architecture with novel vector representations of Chinese characters that capture rich contextual information and sub-character level features. In addition to utilizing the pre-trained character embedding, they proposed a concatenated n-gram representation of the characters. They converted rich local information in the character vectors via utilizing the incrementally concatenated n-gram representation.

4. NEURAL SEQUENCE LABELING MODEL

The neural sequence labelling framework contains three layers, i.e., a character sequence representation layer, a word sequence representation layer and an inference layer.

4.1. CHARACTER SEQUENCE LAYER

Character features such as prefix, suffix, and capitalization can be automatically extracted by encoding the character sequence within the word. Character sequence layer integrates several neural encoders GRU for character-level information of a word into its character-level representation. If a character sequence representation layer is used, word embedding and character sequence representations are concatenated for word representations.

4.2. WORD SEQUENCE LAYER

Character-level information combines with word embedding and feeds them into different networks to model context information of each word. Similar to character sequences, the word sequence layers can model word sequence information through CNN structures. Word CNN utilizes the same sliding window as character CNN.

4.3. INFERENCE LAYER

The inference layer takes the extracted word sequence representations as features and assigns labels to the word sequence. CRF inference layer is examined. CRF considers the correlations between labels in neighbourhoods and jointly decode the best chain of labels for a given input sentence.

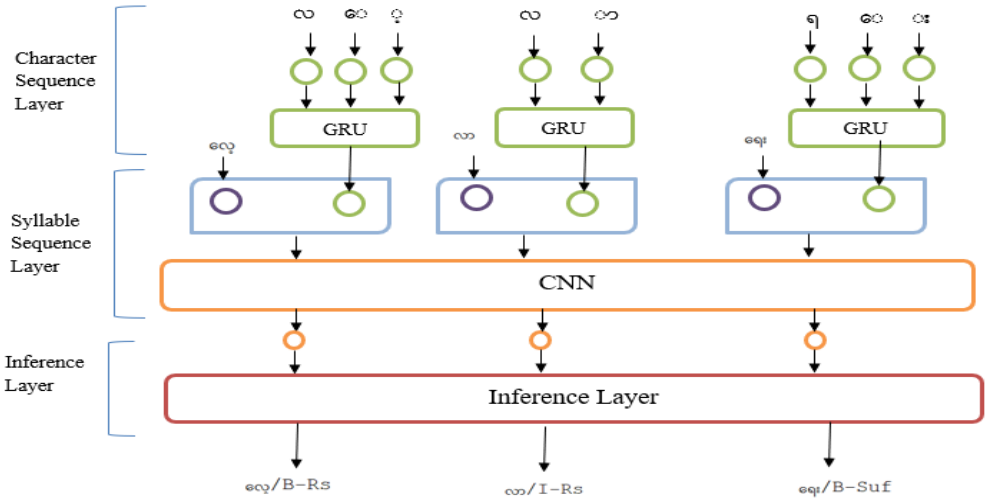

Figure 1. Architecture of neural sequence labeling model

The architecture of the neural sequence labelling model is shown in figure 1. Neural sequence labelling architecture for the word " ".. Green, purple, blue, and orange represent character embedding, syllable embedding, character sequence representations and syllable sequence representations, respectively.

In this approach, pre-trained embedding layers have been applied to improve the performance of neural network architectures for NLP tasks. The main target of word embedding model is to convert word to the form of numeric vectors. Most existing word embedding results are generally trained on data source such as news pages or Wikipedia articles. In English language, word embedding model can be applied for data preprocessing well but there is a very little amount of work done in Myanmar language. Information about word morphology and shape is normally overlooked when learning word representations.

The first step to process a sentence by neural architecture is to transform characters into embedding. This transformation is done by lookup embedding table. A character lookup table $M_{char} \in R^{|V_{char}| \times d}$ where $|V_{char}|$ denotes the size of the character vocabulary and d denotes the dimension of embeddings is associated with all characters. Given a sentence $S = (c_1; c_2; ...; c_L)$, after the lookup table operation, we obtain a matrix $X \in R^{L \times d}$ where the i^{th} row is the character embedding of c_i. When we apply pre-trained embedding with own training data, the performance improves. One of the key points of this architecture to take advantage of better pre-

trained embedding. Experiment on different dimension of word embedding that influence the accuracy of joint model on this task.

5. SYSTEM ARCHITECTURE

In the system, there are four phases. Data collection and syllable segmentation is performed in the Dataset preparation phase. And then, as a pre-processing, manually tag the syllable. In the training phase, the joint process is trained on Neural architecture. In the final stage, untagged data are test.

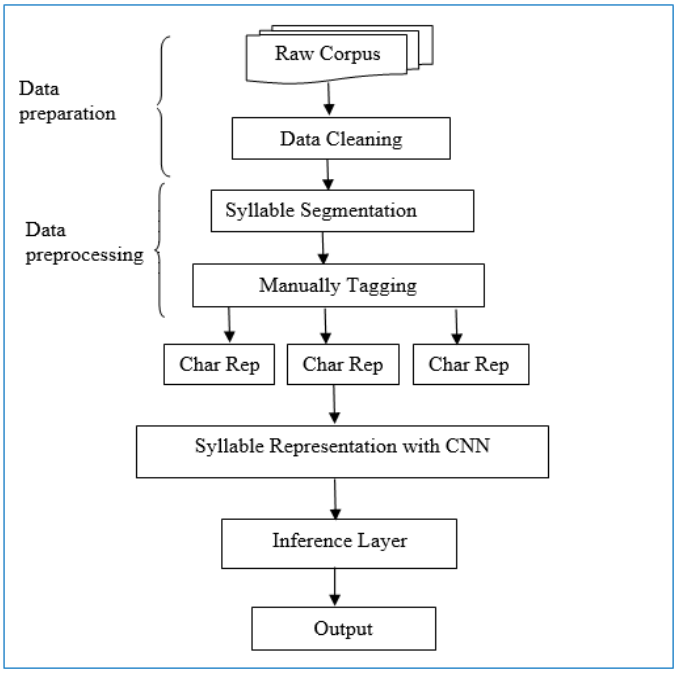

Figure 2. Overview of system architecture

The framework of the proposed system is shown in Figure 5.2. In the system, there are four phases. Data collection and data cleaning is performed in the data preprocessing phase. And then, input sentence is segmented into syllable. After syllable segmentation, each syllable is tagged manually as a preparation of data. Because there is no standard corpus for joint word segmentation and stemming in our language, Myanmar. In order to train the joint word segmentation, stemming and named entity detection model, each syllable is need to tagged manually. And then, in the training phase, joint model is trained on Neural architecture. Training data firstly through the character representation layer. And then, character-level representation and syllable-level embedding are combined at syllable sequence layer. Subsequently, the inference layer assigns labels to each word using the hidden states of words sequence representations. In the final stage, untagged data are tested.

6. TASK SPECIFICATION

In Myanmar Language, "word" is difficult to define normal, to produce the stem word or NER, word segmentation task is a pre-processing stage of stemming and so far, segmentation is considered as a separate process from stemming. In this system, our new approach is integrating them that would benefit in all processes. This approach focuses on syllable-based boundary tagging and proposed approach for stemming and then recognize the named entity at the same time.

6.1. SENTENCE SEGMENTATION

There is no white space between words, but the sentences are delimited by sentence end marker called "။" pote-ma. So, separate the sentence by using sentence end marker.

6.2. SYLLABLE SEGMENTATION

Syllable is a sound unit. A word consists of one or more syllables. In this research, use the algorithm of "A Study of Myanmar Word Segmentation Schemes for Statistical Machine Translation"(Thu, Y.K., Finch, A., Sagisaka, Y., Sumita, E.). The task of joint word segmentation and stemming is to assign word type labels to every syllable in a sentence. In order to indicate the word boundaries, BIO format is represented where every syllable is labelled as B-label if the syllable is the beginning of a word, I-label if it is inside a word but not the first token within the word, or O otherwise. A single word could span several syllables within a sentence. The sentence is first segmented into syllable. Then, from the output, syllable boundary tagging is used to classify the word type and detect the boundary of words.

For stemming, each syllable is tagged with one of the five-word types: Root word (R), Single word(S), Prefix (Pre), Suffix (Suf) and Named Entity (NE). Each syllable is tagged with 'Rs' (for example " ဇလှ/B-Rs ") represents the sub syllable of root word that is also beginning of a root word. Syllable tagged with 'Rs' (for example " လာ/I-Rs ") represents the intermediate word or the last syllable of the root word. Therefore, the one-word type contains one or more syllables. For example root word contains six segmented sub syllables which are "သ/B-Rs ဘာ/I-Rs ဝ/I-Rs ပတ်/I-Rs ဝန်း/I-Rs ကျင်/I-Rs ".

7. RESULTS AND DISCUSSIONS

This section, it is explained the results of research and at the same time is given the comprehensive discussion. Firstly, we represent about data setup. In the corpus, there are 30K sentences. The training corpus is divided into two sets: the first 80% of the data to training and 10% each to test set and development set. The syllable is annotated by one kind of label, such as "B-Rs", "I-Rs", "B-Pre", "I-Pre" "B-Suf", "I-Suf", "B-S", "I-S", "B-NE", "I-NE". In the experiment, we used Newspapers data (Thit Htoo Lwin, 7-Days News, Eleven News Journal).
This is an example of tagging the sentence.

ရန်ကုန်မြို့ သုဝဏ္ဏအားကစားကွင်းတွင် ဒီဇင်ဘာလ ၁၅ရက်မှ ၂၅ရက်အထိ တိုင်းနှင့်ပြည်နယ် ဆောင်းရာသီ အားကစားပြိုင်ပွဲ ကျင်းပခဲ့သည်။

ရန်/B-NE **ကုန်**/I-NE **မြို့**/B-R **သု**/B-NE **ဝဏ္ဏ**/I-NE **အား**/B-R **က**/I-R **စား**/I-R **ကွင်း**/I-R **တွင်**/B-S **ဒီ**/B-NE **ဇင်**/I-NE **ဘာ**/I-NE **လ**/B-R **၁**/B-S **၅**/I-S **ရက်**/B-R **မှ**/B-S **၂**/B-S **၅**/I-S **ရက်**/B-R **အ**/B-S **ထိ**/I-S **တိုင်း**/B-R **နှင့်**/B-S **ပြည်**/B-R **နယ်**/I-R **ဆောင်း**/B-R **ရာ**/I-R **သီ**/I-R **အား**/B-R **က**/I-R **စား**/I-R **ပြိုင်**/B-R **ပွဲ**/I-R **ကျင်း**/B-R **ပ**/I-R **ခဲ့**/B-Suf **သည်**/I-Suf **။**/O

In this example, In this example, "**ရန်ကုန်**" [Yangon] is the named entity. So, it is assigned as NE tag. In this named entity "**ရန်**" is beginning of the name "B-NE" and "**ကုန်**" is end of the named entity "I-NE" "**ရန်**/B-NE **ကုန်**/I-NE". "**မြို့**" [city] is root word, it is tagged as "R" and it is also the only one word and beginning of the root word "**မြို့**/B-R". In the word "**သုဝဏ္ဏ**"

[Thuwana] also the named entity and "သူ" is the beginning of the named entity "B-NE" and "ဝဏ္ဏ" is the end of the named entity "I-NE". "အားကစားကွင်း" [stadium] is the root word and "အား" is the beginning of the root word "B-R", က is the intermediate word of the root "I-R", "စား" also the intermediate word of the root "I-R" and "ကွင်း" is the end of the root word "I-R" "အား/B-R က/I-R စား/I-R ကွင်း/I-R". The word "တွင်" is the Postpositional marker and is the assigned as a single word "တွင်/B-S". "ဒီဇင်ဘာ" is the name of the month, it is identified as named entity and "ဒီ" is the beginning of the named entity "B-NE", "ဇင်" is the middle word of the named entity "I-NE" and "ဘာ" is the last word of the named entity "I-NE" "ဒီ/B-NE ဇင်/I-NE ဘာ/I-NE". "လ" means month and it is only one root word "လ/B-R". ၁၅ is the numerical number so it is added as a single word "၁/B-S ၅/I-S". "ရက်" means day and it is root word "ရက်/B-R". "မှ" is the postpositional marker and label with single word "မှ/B-S". "၂၅" also means numerical number "၂/B-S ၅/I-S". "အထိ" is postpositional marker "အ/B-S ထိ/I-S". In the word "တိုင်းနှင့်ပြည်နယ်" [state and division], it is separated into three words "တိုင်း", "နှင့်" and "ပြည်နယ်". "တိုင်း" is the root word "တိုင်း/B-R", "နှင့်" is postpositional marker "နှင့်/B-S" and "ပြည်နယ်" is the root word "ပြည်/B-R နယ်/I-R". "ဆောင်းရာသီ" [winner] is the root word "ဆောင်း/B-R ရာ/I-R သီ/I-R". "အားကစား" [sport] "ပြိုင်ပွဲ" [contest] also root word "အား/B-R က/I-R စား/I-R ပြိုင်/B-R ပွဲ/I-R". "ကျင်းပခဲ့သည်" [celebrated] is past tense verb and the word "ကျင်းပ" [celebrate] is root verb "ကျင်း/B-R ပ/I-R" and "ခဲ့/B-Suf သည်/I-Suf" is past tense suffix. "။" assign as other word, it means that it is not root word, single word or suffix. It is just a symbol so it is assigned as other word "။/O".

The dropout rate is 0.5 and epoch 100 for training. In each epoch, we divide the whole training data into batches and process one batch at a time. It is evaluated on batch size 20 in the experiments. The joint model is trained on an Intel Xeon E5-2697 processor, training takes about 12 hours while tagging the test set takes about 60 seconds for CoNLL 2003.

In experiments, empirical evaluations of joint models on different configurations are evaluated. We evaluated different setups like the importance of learning rate and pre-trained word embedding that have a large impact on performance.

7.1. PRE-TRAINED EMBEDDING

Pre-trained embedding is a type of vector representation that admits words with the same meaning to have the same vector representation. It is influenced by various NLP research fields including document classification, author identification, sentiment analysis, etc. Actually, it is a class of techniques where individual words are represented as real-valued vectors in a predefined vector space. Each word is mapped to one vector and the vector values are learned in a way that relates a neural network, and the technique is associated with a deep learning approach. To train word vector, we use Global Vectors for Word Representation (GloVe) models between 100 to 600 dimensions. GloVe embedding trained on 10 million tokens and 27K vocabulary size. In this approach, GloVe embedding is used for both character and syllable embedding. In Myanmar Language, information about word morphology and shape is normally overlooked when learning word representations. However, for tasks like stemming, intra-word

information is intensely useful, especially when dealing with morphologically rich languages. Text pre-processing as word embedding is an important part to build a neural network and it is a significant effect on final results.

Table 1. The performance of joint model on different dimension of pre-trained embedding.

Dimension	Precision	Recall	F-Measure
d-100	87.80	88.28	88.04
d-200	89.57	89.77	89.67
d-300	**91.36**	**91.14**	**91.25**
d-400	88.94	88.30	88.62
d-500	90.27	88.78	89.52
d-600	90.24	89.10	89.66

We perform two types of hyper-parameter optimization and selected the best settings based on development set performance. In this experiment, we evaluate the performance with different dimensions between 100-600 on the joint model. According to the experimental results, the difference in terms of performance can be as large as 3.1 percentage points in F1-score for the same hyperparameter setting with different dimensions of word embedding. The best result is GloVe embedding with 300-d in CNN based model. It achieves the F1 score 91.25. The second best embedding is 200-d and 600-d embedding. Based on the outcome of the different runs, the worst F1-score is word embedding with 100-d.

7.2. LEARNING RATE

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is a challenging task because the too small value can cause the system in a long training process and too large value can result in learning a sub-optimal set of weight too fast or unstable training process. So, the learning rate is an important parameter when configuring the neural network model. In this approach, the BiGRU-CNN-CRF model is trained for word segmentation, stemming and check the named entity across the different learning rate in the joint model. Due to the time constraints, we did not perform every point of learning rate tuning. So, it is divided into two types of settings for the learning rate. The first part is the learning rate from 0.001 to 0.009. Different learning rate is tuned and GloVe 300-dimension is used for both character and syllable embedding.

Table 2. The performance of the joint model on a different dimension of the pre-trained embedding.

Learning rate	Precision	Recall	F-Measure
lr 0.001	88.16	87.48	87.82
lr 0.002	89.44	89.36	89.40
lr 0.003	**91.36**	**91.14**	**91.25**
lr 0.004	90.42	89.18	89.80
lr 0.005	90.55	89.85	90.20
lr 0.006	90.40	89.39	89.89
lr 0.007	90.98	90.08	90.53
lr 0.008	87.55	87.58	87.56
lr 0.009	90.59	90.14	90.37

In the experiment, we evaluate learning rate hyper-parameter settings. Then, we took the same setting and tuned the learning rate. According to the experimental result, the selection of the learning rate has a large impact on the performance of the system. On most tasks, F1-score is around 89 and 90% by learning rate from 0.001 to 0.009. Learning rate 0.003 gives the best performance compared to the others. The worst learning rate is 0.008 that has F1-score 87.56. But the performance increase to 3% in the learning rate of 0.009. Learning rate 0.007 gives the second-best performance.

8. CONCLUSIONS

In this research, we consider stemming as a typical sequence tagging problem over segmented word, while segmentation and named entity recognition also can be modelled as a syllable-level tagging problem via predicting the labels that identify the word boundaries and name entities. Our new approach proposed a simple and effective neural sequence labelling model for joint Myanmar word segmentation, Stemming and Named Entity Recognition. This paper performs embedding as a pre-processing step in the CNN-based model which learns character and syllable-level representation of syllables for Myanmar word stemmer that also detect segmentation boundaries and named entities at the same time.

In this paper, different learning rate and different dimension of pretrained embedding is evaluated for each BiGRU-CNN-CRF model. In our proposed system, the hyper parameters that need to be tuned is the dimension, d, of embedding table and learning rate. So, we made a lot of experiments such as choose d=100 to 600 to find that best d for our CNN based model. This is empirically justified in the experiments where d = 300 is standard for published word representation and learning rate 0.003 gets the best F-Measure. Tuning the learning rate significantly impact model performance.

This paper explores the effectiveness of neural network on Myanmar lexical analysis and conducted a systematic comparison between different dimension of pretrained embedding and different learning rates. This exploration of using neural networks for Myanmar lexical analysis is the first work to apply neural network and joint lexical analysis approach.

In future work, we will increase the size of the input file to train embedding model and manually segmented corpus. Joint word segmentation process would like to use in further processing such as parsing, chunking and machine translation. Moreover, stemmer also uses in text summarization, information retrieval and text categorization processes.

REFERENCES

[1] Pa. W.P., N.L.: "Myanmar Word Segmentation using Hybrid Approach.", presented at ICCA, Yangon, pp.166-170, 2008.

[2] Win Win Thant, Tin Myat Htwe and Ni Lar Thein, "Grammatical Relations of Myanmar Sentences Augmented by Transformation Based Learning of Function Tagging", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011

[3] W.P.Pa, Y.K.Thu, A.Finch and E.Sumita, "Word Boundary Identification for Myanmar Text Using Conditional Random Field", Springer, Switzerland, 2016

[4] Thu, Y.K., Finch, A., Sagisaka, Y., Sumita, E.: "A Study of Myanmar Word Segmentation Schemes for Statistical Machine Translation". In Proceedings of 12th International Conference on Computer Applications, Yangon, Myanmar, pp.167-179, 2014.

[5] Jie Yang, Shuailong Liang, and Yue Zhang. "Design challenges and misconceptions in neural sequence labeling". In COLING, 2018.

[6] Jie Yang and Yue Zhang. NCRF++: An Open-source Neural Sequence Labeling Toolkit. arXiv:1806.05626v2[cs.CL] 17 Jun 2018.

[7] Nils Reimers and Iryna Gurevych. 2017a. Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. arXiv preprint arXiv:1707.06799.

[8] Edouard Grave, Piotr Bojanowski, Prakhar Gupta1, Arma for 157 Languages" arXiv:1802.06893v2, 28 Mar 2018.

[9] Jiaqi Mu, Suma Bhat, and Pramod Viswanath. 2017. All-but-the-top: Simple and effective postprocessing for word representations. arXiv preprint arXiv:1702.01417.

[10] Xuezhe Ma and Eduard Hovy. "End-to-end sequence labeling via Bidirectional LSTM-CNNs-CRF". In ACL. volume 1, pages 1064–1074, 2016

[11] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors withSubword Information.arXiv preprint arXiv:1607.04606.

[12] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014) (Vol. 14, pp. 1532–1543). Retrieved from https://nlp.stanford.edu/projects/glove/

[13] Zhenyu Jiao, Shuqi Sun, Ke Sun, "Chinese Lexical Analysis with Deep Bi-GRU-CRF Network". arXiv preprint arXiv:1807.01882. Jul 2018.

[14] Y. Shao, C. Hardmeier, J. Tiedemann, J. Nivre. "Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF." arXiv preprint arXiv:1704.01314. Apr 2017.

AUTHORS

I am **Yadanar Oo.** I am Ph.D candidate of University of Computer Studies, Yangon. I am interested in the field of Natural Language Processing. Currently, I research on Myanmar Stemming, Named Entity Detection and Segmentation.