International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 2, April 2020, pp. 2023~2030
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i2.pp2023-2030

Improving accuracy of part-of-speech (POS) tagging using hidden markov model and morphological analysis for Myanmar Language

Dim Lam Cing, Khin Mar Soe
Natural Language Processing Lab, University of Computer Studies, Myanmar

Article Info

Article history:

Received Sep 23, 2019
Revised Oct 25, 2019
Accepted Nov 2, 2019

Keywords:

Natural language processing
hidden markov model
Morphological analysis

ABSTRACT

In Natural Language Processing (NLP), Word segmentation and Part-of-Speech (POS) tagging are fundamental tasks. The POS information is also necessary in NLP's preprocessing work applications such as machine translation (MT), information retrieval (IR), etc. Currently, there are many research efforts in word segmentation and POS tagging developed separately with different methods to get high performance and accuracy. For Myanmar Language, there are also separate word segmentors and POS taggers based on statistical approaches such as Neural Network (NN) and Hidden Markov Models (HMMs). But, as the Myanmar language's complex morphological structure, the OOV problem still exists. To keep away from error and improve segmentation by utilizing POS data, segmentation and labeling should be possible at the same time.The main goal of developing POS tagger for any Language is to improve accuracy of tagging and remove ambiguity in sentences due to language structure. This paper focuses on developing word segmentation and Part-of- Speech (POS) Tagger for Myanmar Language. This paper presented the comparison of separate word segmentation and POS tagging with joint word segmentation and POS tagging.

Corresponding Author:

Dim Lam Cing,
Natural Language Processing Lab,
University of Computer Studies,
No.4, Main Road, ShwePyiThar Township, Yangon, Myanmar.
Email: dimlamcing@ucsy.edu.mm

1. INTRODUCTION

In numerous uses of characteristic language handling, Part-of-Speech (POS) labeling is an essential assignment for each language. So, to have high precision tagger is one of the importance tasks for NLP applications. Handling ambiguous and unknown words are the challenge of POS tagging [1, 2]. For every NLP application such as machine translation, information extraction, speech recognition, grammar checking and word sense disambiguation, etc are needed to do word segmentation and Part-of-speech (POS) tagging of a fundamental process of natural language processing application. There are many methods for development of POS taggers. The most using techniques are rule based method, statistical based method and neural network based method. In the rule-based approach, rules are developed according to the nature of the language to define precisely how and where to assign the various POS tags [3-5]. This methodology has just been utilized to build up the POS tagger for Myanmar Language. In the factual methodology, measurable language models are manufactured, refined and used to POS label the info message naturally. Most commonly used statistical approaches are Hidden Markov Models based approach, Support vector machine based, Conditional Random Field based and Maximum Entropy based approach [6, 7].

This paper describes Hidden Markov Models (HMM) and the proposed system for word segmentation and part-of-speech tagging for Myanmar language. Myanmar Language is morphologically

rich, complex, and agglutinative in nature, expressions of which are arched with numerous linguistic highlights. POS labeling [8] is a significant issue in the field of NLP and one of the fundamental preparing ventures for any language in NLP. i.e., the capability of a computer to automatically POS tag a given sentence. Normally, the first step of processing is to divide the input text into units called tokens where each is either a word or something else like a number. The main clue used in space-delimited language like English is the white space. In major East-Asian languages such as Japanese, Chinese, Thai and Myanmar, there is no spaces between words. Myanmar language, its writing style does not use any delimiter between words.

In word segmentation and POS tagging, the structure of morphological words is the main source of information to get the correct process of tagging. By using the morphological structure of words, eliminate irrelevant tags can be removed and find the suitable tag for the word [9-11]. So, morphological analysis is an important part of language engineering applications especially for morphologically rich and complex language like Myanmar.

There has been very few research conducted on various language processing tasks including morphological analysis for Myanmar language compare to English, France, Chinese, India, and Thai., etc. Since high level language processing tasks such as POS tagging, machine translation, semantic analysis, syntactic analysis, sentiment analysis, information retrieval, classification, clustering system, etc. all process on smallest language unit; words. The morphology of the language through a systematic linguistic study is important in order to reveal words that are significant to users such as historians, linguists, etc.

Most of the current researches on Myanmar language done used a lexicon or dictionary or corpus which lists all the words forms for word segmentation as an initial stage of processing. To get correct segmentation, we need an exhaustive lexicon or corpus. Myanmar language[12-16] has been classified by linguists as a monosyllabic or isolating language with agglutinative features. Its writing style does not use any delimiter between words and so there is no way of knowing whether a word form of syllables is group, or is just a separate group of monosyllabic words. Every syllable has a meaning of its own. The Myanmar Language have complex morphotactic structures and has the ambiguous word segmentation. Therefore, segment the sentence to generate lexical and semantic of word sequences is a challenging task. Thus, this paper aim to addresses this shortcoming by proposing a language model that consider joint word segmentation and POS tagging. The rest of this paper is organized as follows. In Section 2, we discussed Literature Review. Section 3 described Aspect of Myanmar Language. Section 4 presented Design of Proposed System. Section 5 provides the Evaluation. Finally, we described the conclusion of the paper.

2. LITERATURE REVIEW

Part-of-Speech Tagger that using supervised learning approach for Myanmar Language is presented in [17]. For disambiguous of the POS tags, Baum-Welch algorithm and Viterbi algorithm with HMM model is used for training and decoding. For tagging a word, Myanmar lexicon is used with its all possible tags. The examination results show that the strategy got high precision (over 90%) for various testing input. Myanmar Word Segmentation [18] used Hybrid Approach and the sentences are segmented in syllable and matched by longest words. In the using of Longest matching method, the words that are known from a dictionary are first segmented and the unknown words are guest from an n-gram model [19]. The major issue of this technique is comes from the vagueness in the longest coordinating procedure, since words can be showed up in numerous structures.

The porposed of Y. Zhang and S. Clark [20], that got a lower mistake rate contrasted with a two stage baseline system. The large combined search space for this method is a challenge and it is very hard in decoding. For reason for at the same time word division and POS labeling, a solitary straight model is utilized, and for joint preparing and pillar search of unraveling, the summed up perceptron calculation is picked. The joint model lessens a mistake pace of exactness for division to 14.6% and a blunder decline in labeling precision of 12.2%, contrasted with the conventional pipeline strategy. A Persian POS tagger, the Persian sentences are tagged by implementing a blend of measurable and principle-based technique. To tag unknown words, a morphological analysis probabilistic method is used. Persian morphological rules that are knowledge base and that the probabilities is worked by a corpus is the second result of the research. Trial results show that their approach increase the labeling execution and exactness [11].

3. ASPECT OF MYANMAR LANGUAGE

Myanmar language is highly agglutinative and is morphologically rich and complex. Moreover, to separate each word, the Myanmar writing style do not use spaces and there is no chance to get of knowing whether a gathering of syllables structure a word, or is only a group of separate monosyllabic words.

Every syllable has its own meanings. In Myanmar words consist of one or more syllables which are compound in different ways. Depend on the way of the words structures from syllables, these can be classify into three types single simple words, complex words and reduplicative words [21, 22]. For example, ေပါင်း (steam) + အိုး(pot) =>ေပါင်းအိုး (rice cooker), မီး(fire) + ပူ (hot) => မီးပူ (iron), ပန်း(flower) + ချီ (carry) => ပန်းချီ (painting), all have their referential meaning and each monosyllable within words also has their own meaning. In Myanmar morphology processes include inflection, derivation, and compounding.

3.1. Inflection morphology

Myanmar inflection morphology of nouns, verbs and adjectives is mostly achieved by suffixation. The inflection morphology remains the same POS tags with the original words but by adding the inflection morpheme -တို့, -များ can make the plural on nouns and the inflectional morpheme -ခဲ့ make the past tense on verbs. For example: ေကျာင်းသားများ (students) -> ေကျာင်းသား (student) + များ; သွားခဲ့ (went) -> သွား (go) + ခဲ့.

3.2. Derivation morphology

Myanmar morphology derivation occurs by means of prefixation or suffixation. Derivation can change the POS tag of word forms. Derivation of nouns, verbs and adjectives are also achieved by suffixation but a circumfix also occurs in the Myanmar language. For example: အလုပ် (work) -> အ (Prefix) + လုပ် (do); ေ◌ြ ပး ◌ြ ခင်း (running) -> ေ◌ြ ပး (run) + ◌ြ ခင်း (Suffix). But အ- is not prefix bound morpheme in some nouns and verbs and cannot be splitted; for example: if the words အေမ(mother) is splitted, it has not meaning.

3.3. Compounding

Myanmar words contain many compound words. They are noun compound words, verb compound words, adjective compound words and also noun, verb and adjective are compound. For example: compound noun: ေစျးနှုန်း (price)-> ေစျး(market) + နှုန်း(rate); compound verb: ◌ြ ဖတ်ပိုင်း (voucher) -> ◌ြ ဖတ်(cut) + ပိုင်း(divide); compound adjective: ခိုင်မာ (firm) -> ခိုင်(firm) + မာ(rigid); compound noun,verd and adjective: လူနာတင်ကား(ambulance) -> လူ(human) + နာ(painful) + တင်(placed) + ကား(car). By compounding the words some words POS is the same to the original and some words got a new POS tag.

4. DESIGN OF PROPOSED SYSTEM

The structure of the proposed framework is shown in Figure 1. There are two modules: preparing and testing modules. In the training phase, the collection of segmented and tagged-sentences are used to develop the proposed HMM model. This model is used in the testing phase. In testing phase, the input Myanmar sentences are identified into each sentence using the sentence end marker called pote-ma '။'. After that, word segmentation and POS tagging is performed

4.1. Corpus creation

Part-of-Speech tagged corpora are one of the essential resources for developing state-of-the-art POS Tagger in Myanmar. There are several steps to create tagged corpus. The following list demonstrates steps needed corpus building.

- Collecting raw text
- Hand-annotating and preparing training data

We collect and normalize raw text from online journals, newspapers and e-books. Since, documents used various Myanmar font styles; these are converted to standard Unicode format and and make cleaning such as spelling checking. We assign tags in un-annotated text manually and finally, we have got the training data for statistical method. If the number of tags is large, the complexity will be increased and the performance will be decreased. According to Myanmar grammar books and dictionary book [12-16], there are nine Part-of-Speech tags in Myanmar language. We have annotated every word with appropriate basic POS tags and created a POS tag Corpus. Moreover, we added another three POS tags Number, Symbol and Abbreviation in our research. The tagset is described in Table 1.

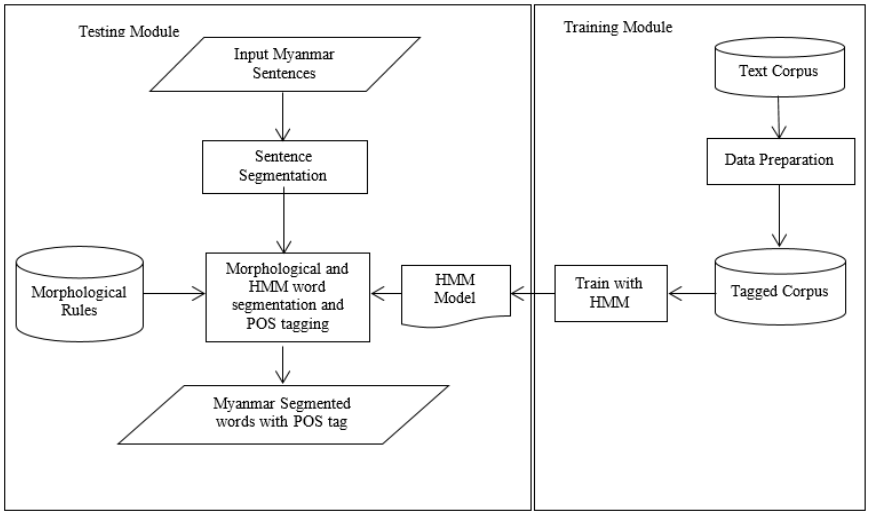

Figure 1. Framework of the proposed system

Table 1. Tagset

No.	Tag	Description	Example
1.	NN	Noun	ပန်း(flower)
2.	PN	Pronoun	ကျွန်မ(I)၊ သင်(you)
3.	V	Verb	ဝယ်(buy)၊ စား(eat)
4.	Adj	Adjective	ပူ(hot)
5.	Adv	Adverb	ေလးစားစွာ (respectfully)
6.	PPM	Postpositional Marker	က၊ ကို
7.	Conj	Conjunction	ထို့အခါ၊ ၍
8.	Part	Particles	များ၊ ခဲ့
9.	Interj	Interjection	အဲ၊ အမယ်လေး
10.	Number	Number	၁၂၂၀
11.	Symbol	Symbol	( ) / % + - = ၊ ။
12.	Abbrev	Abbreviation	အထက၊ဖဆပလ၊ေအဘီအမ်

4.1.1. Corpus statistic

For our experiments, the corpus consists of sentences from Myanmar grammar books, Myanmar text books, some Myanmar history and websites. Corpus informations are described in Table 2. The font used for this research is Unicode. There are total 39716 sentences covering 690258 words and each sentence has an average of 18 words. The vocabulary size is 27043 words.

Table 2. Distribution of POS tags

POS Tags	No. of words
NN	25%
PN	4%
V	15%
Adj	2%
Adv	2%
PPM	17%
Conj	5%
Part	22%
Interj	0.03%
Number	1%
Symbol	7%
Abbrev	0.09%

4.2. Training hidden markov model

To get training data, we have to compute probabilities for each tag in the tagged corpus. Since we have developed a model, it produces two results. The results of the training phase are transition probabilities and emission probabilities.

4.2.1. Estimating probabilities

POS tagging using HMM, the probabilities are calculated from a tagged training corpus instead of using the full power of HMM learning. The probabilities of tag transition $P(t_i|t_{i-1})$ is the probability of a tag given in the previous tag. Estimation of transition probability is computed by counting the times that the first tag in a tagged corpus, how often the first tag is followed by the second.

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1},t_i)}{C(t_{i-1})}$$

The emission probabilities, $P(w_i|t_i)$ given a tag, it will be associated with a given word [23]. The emission probability is

$$P(w_i|t_i) = \frac{C(t_i,w_i)}{C(t_i)}$$

4.3. Joint Myanmar word segmentation and POS tagging

The input sentences are firstly separated by pote-ma "။". The words in each sentence is segmented and assigned POS with the proposed tagsets in Table 1 by using HMM probabilistic models. In Myanmar Language, since words are formed by combining more than one syllable that is one word can have one or more syllables and one syllable has more than one character, syllable identification must be done before word level segmentation [24]. For example, the input is as follows in Table 3:

ြကာပန်းသည်ရထဲတွင်ပေါက်သည်။ (Lotus grows in water.)

After Syllable Identification, the right output is come out as follows:

ြကာ|ပန်း|သည်|ေရ|ထဲ|တွင်|ေပါက်|သည်

Table 3. N-gram word segmentation for input sentence

N-gram (N=1,2,3,4,5)	Word Segmentation
Unigram	ြကာ|ပန်း|သည်|ေရ|ထဲ|တွင်|ေပါက်|သည်
Bigrams	ြကာပန်း|ပန်းသည်|သည်ေရ|ေရထဲ|ထဲတွင်|တွင်ပေါက်|ေပါက်သည်
Trigrams	ြကာပန်းသည်|ပန်းသည်ေရ|သည်ေရထဲ|ေရထဲတွင်|ထဲတွင်ပေါက်|တွင်ပေါက်သည်
4-grams	ြကာပန်းသည်ေရ|ပန်းသည်ေရထဲ|သည်ေရထဲတွင်|ေရထဲတွင်ပေါက်|ထဲတွင်ပေါက်သည်
5-grams	ြကာပန်းသည်ေရထဲ|ပန်းသည်ေရထဲတွင်|သည်ေရထဲတွင်ပေါက်|ေရထဲတွင်ပေါက်သည်

A typical strategy to do word division and POS simultaneously is to utilize the N-gram (5-grams) which sweeps an information sentence from left to right, and recover the word with its everything potential labels with the likelihood from emanation record. If all 5-grams words have not been contained in the emission probability file, the system used 4-grams, trigrams, bigrams and unigram. Word segmentation for input sentence as per the longest N-gram technique

ြကာပန်း|သည်|ေရထဲတွင်|ေပါက်|သည်

Word probabilities and language model probabilities is calculated by using relative frequency count. If there are more than one POS options for word, the system selected POS option with highest word probability as described in Table 4.

Table 4. All possible word, tag and probability

Word Segmentation	POS	Language Model Probability	Selected POS
ြကာပန်း (Lotus)	NN	1	NN
သည် (null)	PPM	0.4	PPM
	Part	0.3	
	PN	0.2	
	Adj	0.1	
ေရ (water)	NN	0.6	NN
	V	0.2	
	Part	0.2	
ထဲတွင် (in)	PPM	1	PPM
ေပါက်(grow)	Part	0.2	
	V	0.7	V
	NN	0.1	
သည်(null)	PPM	0.4	PPM
	Part	0.3	
	PN	0.2	
	Adj	0.1	

4.4. Morphological rules approach

The internal structures of words are defined by using morphological rules [11]. These rules comprise of three sections: prefix (ဘခ), stem and suffix (များ). The common syntax is as follows:

prefix + stem + suffix → POS tag

In the above syntax, sometime both of prefix and suffix are contain in the string. In some syntax, one of prefix or suffix is empty string. There are three types' morphological rules for Myanmar Language: inflectional, derivational rules and compounding. In this system, morphological rules (68 rules) are characterized [25] and utilized. The rules are drawn out from Myanmar Grammar book [12-16]. The uses of inflectional, derivational and compounding are described in Section 3.

5. EVALUATION

To appraise the testing result for POS labeling, the framework utilized the parameters of Recall, Precision and F-score. These parameters are characterized as pursues:

$$Recall, R = \frac{Number\ of\ correct\ POS\ tag\ assigned\ by\ the\ system}{Number\ of\ words\ in\ the\ test\ set}$$

$$Precision, P = \frac{Number\ of\ correct\ POS\ tag\ assigned\ by\ the\ system}{Number\ of\ POS\ tag\ assigned\ by\ the\ system}$$

$$F_{score}, F = \frac{2PR}{P+R}$$

5.1. Experimental setup

For testing the proposed model, we divided our corpus into two corpuses as follows in Table 5. We collect 500 new sentences for open testing. In our experiments, we compare the separate word segmentation and POS tagging using HMM , joint word segmentation and POS tagging using HMM and joint word segmentataion and POS tagging using HMM with morphological rules in Table 6. For the comparative purpose, we used Bigram Part-of-Speech Tagger for Myanmar Language [17] as based line system. The proposed system and base line system used same training corpus and test data.

Table 5. Statistic of the dataset

Data	No.of Sentence	No. of words
Corpus 1	29680	547969
Corpus 2	39716	690258

Table 6. Accuracy of system on different test cases using HMM and morphological rules

Corpus Size (sentences)	Separate word Segmentation and POS tag			Joint word segmentation and POS tag			Joint word segmentation and POS tag + morphological rules		
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
29680	68%	67%	67%	78%	76%	77%	90%	88%	89%
39716	77%	75%	76%	85%	83%	84%	94%	92%	93%

5.2. Results and discussion

Table 6 shows the experiment results for Myanmar word segmentation and POS tagging with different training data sizes. Conforming to the table, the proposed technique starts to get a few progressions over the correlation standard. When the measure of preparing information sentences is increased and using of morphology rules also has good increased compared with the corresponding baselines. The accuracy of the tagger is appraised by using testing data which contains different kinds of words. Testing words can be defined as known words, unknown words and ambiguous words for the tagger. “Known words” are the words contain in the training corpus and “Unknown Words” are the words which are not containing in the training corpus. “Ambiguous words” are the known words which are tagged wrong because of segmentation error and it is needful to solve for disambiguating that tag is the correct tag for these words. In proposed system, most “Unknown Words” occur in Proper Noun (name of person, name of location), different position of Particle and Postpositional marker in segmentation can cause ambiguous in POS tagging. There is no training data to cover all Proper Nouns. Including of disambiguous words and unknown words make decrease in the performance of the tagger. To solve the disambiguation of ambiguous words is to use the morphological rules. By using morphological rules, the system reduced ambiguous in Particle and Postpositional markers.

6. CONCLUSION

This paper presents a joint word segmentation and POS tagging in Myanmar using HMM and morphological rules. In our experiments, we compare the separate word segmentation and POS tagging with our proposed joint word segmentation and POS tagging using HMM. Then, we found that there is a significant improvement in joint word segmentataion and POS tagging using HMM with morphological rules. We also describe the distribution of words in the corpus. Until now, there are unknown words in our experiments. The future work will be to improve the exactness of word segmentation and POS tagging. We also need a larger corpus for training. By using a large training and morphological rules, the assignment of POS tag will be more accurate and will be reduced the unknown words, incorrect tag and ambiguous words. The paper has shown that word segmentation and POS tagging in Myanmar can be improved by using lager training corpus and combining the morphological analysis of Myanmar Language.

REFERENCES

[1] T. Mikolov, A. Deoras, D. Povey, L. Burget, J. H. Cernocky, "Strategies for training large scale neural network language models," *IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 196-201, 2011.
[2] A.J.P.M.P. Jayaweera, N. G. J. Dias, "Hidden markov model based part of speech tagger for sinhala language," *International Journal on Natural Language Computing (IJNLC)*, vol. 3(3), 2014.
[3] Sirajuddin Y. Hala, Sagar H. Virani, "Improve accuracy of parts of speech tagger for Gujarati language," *International Journal of Advance Engineering and Research Development*, vol. 2(5), 2015.
[4] P.M Bhatt, A. Ganatra, "Analyzing & enhancing accuracy of part of speech tagger with the usage of mixed approaches for Gujarati," *International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878*, vol. 8(1), 2019.
[5] K. Mohnot, N. Bansal, S.P. Singh, A. Kumar, "Hybrid approach for part of speech tagger for Hindi language," *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol. 4(1), 2014.
[6] S. AlGahtani, J. McNaught, "Joint Arabic Segmentation and Part-of-Speech Tagging," *Proceedings of the Second Workshop on Arabic Natural Language Processing ©2014 Association for Computational Linguistics*, pp. 108-117, 2015.
[7] A. F. Wicaksono, A. Purwarianti, "HMM based part-of-speech tagger f or Bahasa Indonesia," *On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop,* 2010.
[8] S. HOON N. A., "Conditional random fields for Korean morpheme segmentation and POS tagging," *ACM Transactions on Asian Language Information Processing*, vol. 14(3), 2015.
[9] Z. H. Pozveh, A. Monadjemi, A. Ahmadi, "Persian texts part of speech tagging using artificial neural networks," *Journal of Computing and Security*, vol. 3(4), pp. 233-241, 2016.
[10] C. Lyu, Y. Zhang, D. Ji, "Joint word segmentation, POS-tagging and syntactic chunking," *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, 2016.

[11] H. Fadaei, M. Shamsfard, "Persian POS tagging using probabilistic morphological analysis," *Int. J. Computer Applications in Technology*, vol. 38(4), pp. 264-273, 2010.
[12] P. Hopple, "The structure of nominalization in Burmese," *Ph. D Dissertation. University of Texas, Arlington,* 2003.
[13] Department of the Myanmar Language Commission ,"Myanmar grammar," *Ministry of Education. Myanmar,*2006.
[14] "Myanmar-English dictionary," *Ministry of Education,* Myanmar.
[15] Grammar. Burmese language. http://en.wikipedia.org/wiki/Burmese_Language
[16] Department of the Myanmar Language Commission, "Myanmar grammar," *Ministry of Education, Myanmar,* 2016.
[17] P. H. Myint, T. M. Htwe, N. L. Thein, "Bigram part-of-speech tagger for Myanmar language," *2011 International Conference on Information Communication and Management, IPCSIT,* vol. 16, 2011.
[18] W. P. Pa, N. L. Thein, "Myanmar word segmentation using hybrid approach," *Proceedings of 6th International Conference on Computer Applications,* 2008.
[19] W. P. Pa, Y. K. Thu, A. Finch, E. Sumita, "Word boundary identification for Myanmar text using conditional random fields," *Genetic and Evolutionary Computing, Springer International Publishing Switzerland,* p. 447,2016
[20] Y. Zhang, S. Clark, "Joint word segmentation and POS tagging using a single perceptron," *Proceedings of ACL-08: HLT,* pp. 888-896, 2008.
[21] T. M, Htwe, D. L. Cing, "A neural probabilistic language model for joint morphological segmentation and POS tagging," *The Seventh International Conference on Science and Engineering(ICSE),* pp. 9-10, 2016.
[22] T. T. Zin, K. M. Soe, N. L. Thein, "Myanmar phrases translation model with morphological analysis for statistical Myanmar to English translation system," *25th Pacific Asia Conference on Language, Information and Computation,* pp. 130-139, 2011.
[23] D. Jurafsky, James H. Martin, "Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition," Copyright 2006, Draft of June 25, 2007.
[24] https://github.com/ye-kyaw-thu/sylbreak
[25] D. L. Cing, K. M. Soe, "Joint word segmentation and part-of-speech (POS) tagging for Myanmar language," *17th International Conference on Computer Application,* 2019.

BIOGRAPHIES OF AUTHORS

Dim Lam Cing received M.C.Sc in Computer Science from Computer University (Kalay) in 2010. She is a PhD candidate in University of Computer Studies, Yangon (UCSY). Her research interest includes Natural Language Processing and Machine Learning.

Khin Mar Soe received M.C.Sc and Ph.D degree in Information Technology from University of Computer Studies, Yangon (UCSY) in 2000 and 2005 respectively. She is currently a full professor from Natural Language Processing (NLP) Lab in UCSY. Her main research interest includes Natural Language Processing and Artificial Intelligence.