

Emotion detection on social media status in Myanmar language

Thiri Marlar Swe, Naw Lay Wah

Faculty of Information Science, University of Computer Studies, Mandalay, Myanmar

Article Info

Article history:

Received Dec 14, 2022

Revised Jan 9, 2023

Accepted Feb 4, 2023

Keywords:

Emotion classifier

Emotion detection

Myanmar language

Myanmar word-emotion

Lexicon

Naïve Bayes emotion classifier

Recurrent neural network

ABSTRACT

Many social media emerged and provided services during these years. Most people, especially in Myanmar, use them to express their emotions or moods, learn subjects, sell products, read up-to-date news, and communicate with each other. Emotion detection on social users makes critical tasks in the opinion mining and sentiment analysis. This paper presents the emotion detection system on social media (Facebook) user status or post written in Myanmar (Burmese) language. Before the emotion detection process, the user posts are pre-processed under segmentation, stemming, part-of-speech (POS) tagging, and stop word removal. The system then uses our preconstructed Myanmar word-emotion Lexicon, M-Lexicon, to extract the emotion words from the segmented POS post. The system provides six types of emotion such as surprise, disgust, fear, anger, sadness, and happiness. The system applies naïve Bayes (NB) emotion classifier to examine the emotion in the case of more than two words with different emotion values are extracted. The classifiers also classify the emotion of the users on their posts. The experiment shows that the system can detect 85% accuracy in NB based emotion detection while 86% in recurrent neural network (RNN).

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Thiri Marlar Swe

Faculty of Information Science, University of Computer Studies

Mandalay, Myanmar

Email: thirimarlarswe@gmail.com

1. INTRODUCTION

In this age, social media has become rapidly popular to communicate with each other around the world. Most people daily use social media, such as Facebook and Twitter. They express their emotions, moods, share knowledge, get up-to-date local or global news, sell products by posting or sharing status with texts, videos, photos and sometimes with the feeling tag that provided by social media application. Emotions perform a key role in human cognition, particularly in human interaction, human intelligence, perception, and rational decision-making. Emotions of social media are important in some application areas such as medical treatment, pre-election result estimation, and opinion on the specific topic.

An emotion is a psychological state which includes a subjective experience, a behavioral or expressive response, and a physiological response [1]. Emotion detection is the process of identifying a person's feelings or emotions. Detecting emotion from text is a new classification task and advances in textual analysis and becomes an interest in natural language processing (NLP) field. Psychologist Ekman [2] suggested that surprise, fear, anger, sadness, and happiness as six basic emotions.

Machine learning-based and Lexicon-based approaches are the main approaches in emotion detection from the text. Many researchers do studies by applying different approaches. In a lexicon-based approach, researchers create language-related emotion lexicons which can be directly applied in emotion detection from text. In the lexicon, the words are mapped to their emotional state. This process uses a bag of words noted with their emotional bearings and clearly calculates the total number of words or masses the

values included in texts. This paper presents an emotion detection system which applies both machine learning-based and lexicon-based approaches. The system aims to provide emotional information in some application areas such as election period, mental health care, criminal cases, and so on. In a lexicon-based approach, the system uses a pre-constructed Myanmar word-emotion lexicon (M-Lexicon) proposed by Swe and Myint [3] in order to extract the emotion words from the segmented part-of-speech (POS)-tagged post.

If there is one and only one emotion word or words with the same emotion value, the system can easily detect the emotion of the post depending on the emotion word value. If a post has more than one emotion word with different emotion values, the system needs to examine the emotion from such kinds of words. The problem states which word is strongly represented to express the emotion of the post. To solve this problem, rule-based detection can be applied. Rule-based detection has some facts such as there may need to define strong and complete rules, to have more new rules to get more accurate results. Machine learning-based approaches can overcome this fact.

Machine learning-based has more benefits than rule-based; the system can simply and easily detect the emotion of the social media user. So, the system applies machine learning-based, naïve Bayes (NB) and/or recurrent neural network (RNN), classifiers in the emotion detection process. This paper presents the NB emotion classifier and RNN emotion classifier to examine the emotion of the user when there is more than one extracted emotion word in the post, with different emotion values. The emotion detection system will also detect different user groups' emotions from their posts.

The rest of this paper is formed as: the next section describes the related works. It is followed by the method of the emotion detection system in detail. Before concluding remarks in section 5, results are shown in section 4.

2. RELATED WORKS

Most of emotion recognition work had been done for English and their mother languages (Korean [4], Vietnamese [5], France [6], and Arabic [7]). The machine learning based and lexicon based approaches can complete their works [8]. In this section, a short review of studies efforts in which lexicon and machine learning approaches are applying in emotion detection from text will be presented.

Two weighting schemes and mutual information are proposed to create Turkish emotion lexicon (TEL) [9] based on a TREMO dataset. The lexicon was enriched by using bigram and concept hierarchy methods. By applying their proposed lexicon, the performance of machine learning and the lexicon-based approaches were compared. Their experiments had shown that the utilization of their proposed lexicon orderly produced similar outcomes in emotion examination in Turkish text.

The restaurant business has achieved remarkable popularity in recent years. The positive review on the restaurant services is important to attract new customer as well as to maintain the existing customer. Thus, sentiment analysis on the positive or negative reviews on the restaurant service become important facts. A sentiment classification system [10] had been proposed to analyze negative and positive classes of the customer reviews based on their hazy feedback in Bengali language. They use multinomial NB classifier in restaurant service classification. Their classifier can classify 84% of positive and 74% of negative reviews correctly. They used small dataset.

Deep learning-based emotion analysis on Arabic tweets are developed [11]. They used deep convolutional neural networks (CNN) for Arabic sentence categorization tasks. They compared their approach with three machine learning algorithms like support vector machines (SVM), multi-layer perceptron (MLP), and NB. They used count-vectorized and term frequency-inverse document frequency (TF-IDF) schemes in calculating the accuracy. They applied Arabic tweets dataset grant by SemiEval with four emotion categories: fear, anger, sadness, and joy. Their way attained exceptional results with 99.82% validation accuracy, and 99.90% training accuracy.

RNN is applied on the sentiment analysis of Spanish Tweets [12]. They proposed a modified version of feedforward RNN, namely long short-term memory (LSTM) cells. Their architectures used the combination of sentiment lexicon values and input word vector. They tested their architectures on three test datasets: InterTASS, 1k TASS, and full TASS. Their experiments have shown that sentiment analysis performance can be enhanced by their combined features but not best performance. Their methods needed to perform better preprocessing of input at word level.

3. STUDIES METHOD

This paper presents an emotion detection system on Facebook status/post written in Myanmar (Burmese) language. The detection system uses Facebook user's posts and detects the emotion of the user from the post. The system additionally provides the emotion of the user and groups of users on their personal

profile information such as educational status, occupation or work, birth date, and gender. The system first determines whether the post is a feeling status. If the status is with the feeling tag, the system examines the emotions from the feeling word using the feeling-emotion lexicon, which is a sub-lexicon in M-Lexicon. If not, the system performs the emotion detection process using classification methods.

The emotion detection includes three main parts: data pre-processing, M-Lexicon-based emotion words extraction, and emotion classification using machine learning-based approaches. The collected Facebook status is preprocessed with three steps such as word segmentation, stemming word and POS tagging, and stop word removal. The words from the preprocessed step are used to extract the emotion words. The extracting process is completed by finding the word in the lexicon. The extracted words with the related emotion values are used to examine the emotional label for the post. The general architecture of the detection system is illustrated in Figure 1.

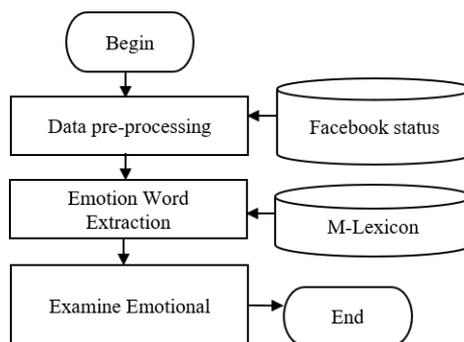


Figure 1. Architecture of the system

3.1. Data collection and pre-processing

Facebook has become the most popular social medium around the world. Facebook had 2.92 billion members worldwide by the end of 2021 [13]. Facebook permits their users to pick the particular “networks” that they owned, such as geographic location, education, and occupation/work. Each user retains a “profile”, that shows the fundamental facts like hometown, graduation and also personal information: username and whether the user is “in a relationship” or single (i.e., “relationship status”) and so on. The system collects 27 Facebook user profile information such as education, graduation, work, date of birth, and gender. Facebook provides several communication ways between their users. Users communicate with each other by sending private messages. In addition, the users can express their feelings, emotions, photos and videos by posting status on each other’s Walls [14]. Facebook posts can contain the feeling status. A Facebook post includes the following information: post owner (with the tag friend-in some post), posted time, post content and reactions to the posts. There are six types of animated reactions: ‘classic like’, ‘love’, ‘haha’, ‘sad’, ‘wow’, and ‘angry’. The system stores the text posts and posts with feeling tag of each Facebook user. The reaction counts of the post are also kept with it. The collected data is preprocessed before the emotion word extraction process is performed. The pre-processed steps contain three main processes: segmentation on Myanmar text post, the stemming word and POS tagging process, and stop word removal process as last step.

3.1.1. Word segmentation

Myanmar (Burmese) language is the Sino-Tibetan language spoken by 30 million people and the official language in Myanmar. In Myanmar script, the sentences are written in left to right style and obviously divided by a sentence boundary marker. In spite of the fact that inter-phrase spacing may sporadically to be employed, there is no typical inter-word spacing [15]. In Myanmar writing, white spaces do not use between words or between syllables. A Myanmar text is a characters string which does not have explicit word boundary delimiters. So, the word segmentation is an important stride preceding the emotion detection from the text.

The word segmentation can be done by using a segmentation tool, namely Myan-Word-breaker [16]. Myan-Word-breaker performs a segmentation process in syllable segmentation and syllable merging steps. By using rule-based heuristic approach, the syllable segmentation process is acted in which the sentence is segmented by defined six rules [17]. By using the dictionary-based statistical approach, the segmented syllables are combined into words afterwards the syllable segmentation is done. In some cases, the segmented result is re-merged manually to better segmented results. The result of the system segmentation process is shown in Table 1.

Table 1. Word segmentation on a post

Original Post	မနေ့ကစိတ်ညစ်နေပေမယ့်အခုတော့ဝမ်းသာစရာတွေနဲ့သာယာနေတာ (Yesterday I was sad, but now I am happy with my joys)
Syllable Segment	မ_နေ့_က_စိတ်_ညစ်_နေ_ပေ_မယ့်_အ_ခု_တော့_ဝမ်း_သာ_စ_ရာ_တွေ_နဲ့_သာ_ယာ_နေ_တာ
Merged words	မနေ့_က_စိတ်ညစ်_နေ_ပေမယ့်_အခု_တော့_ဝမ်းသာ_စရာ_တွေ_နဲ့_သာယာ_နေ_တာ

3.1.2. Stemming and part-of-speech tagging

Segmented word may contain the repeated word or need to reduce it with the same root to an accepted form. For example, in English, the words chocolates, chocolatey and choco as being the same and derived from chocolate. The stemming process is mostly done in three ways: rule-based stemming, statistical stemming, and hybrid stemming. In this emotion detection system, a rule-based stemming way is applied in the stemming word. This process, the word stemming process, is performed before the POS tagging process. Table 2 shows some stemming rules of the system. The algorithm of the stemming process is described in Figure 2. The system finds root word of each segmented word within the stemming rules. If there is a matched rule, the system replaces the searched word with the root word of it. The system continues the process until the end of segmented words.

Table 2. Stemming rules

Words	အကြိုက်၊ အကြိုက်ဆုံး၊ ကြိုက်ကြိုက်၊ ကြိုက်ဆုံး	စိတ်တွေညစ်	ပျော်ပျော်ရွှင်ရွှင်
Root Word (Stem)	ကြိုက်	စိတ်ညစ်	ပျော်ရွှင်

```

Input: seg_text, stem_rules
Output: seg_stem_text
Begin
b_texts ← break the seg_text
For each w in b_texts
    root_word ← look_up w in stem_rules words
    If root_word <> NULL Then
        substitute w with root_word in b_texts
    End If
End For
seg_stem_text ← combine b_text
End
    
```

Figure 2. Stemming process steps

After segmentation and stemming word processes are performed, the system applies POS tagging so as to continue the finding and removing the stop words as the next step. POS tagging is a process to classify words in the text in association with a particular part of speech, depending on the meaning of the word and its context. The POS tagger sets a POS tag to each word in a given sentence. This emotion detection system uses Myanmar POS (myPOS) corpus [18], [19] and ripple down rules-based part-of-speech tagger (RDRPOS) tagger [20], a fast and simple to use and language free toolkit, to tag the segmented Facebook post body.

3.1.3. Stop word removal

There may have words in the POS tagged text which are not important or not required to use in the lexicon creation and emotion detection process because it does not have much meaning to express the emotion of user. Thus, the system removes such kind of words, stop words, after the POS tagging process. Stop words, by definition, are unmeaning words with very low differentiation value [21]. Pronouns, conjunctions, prepositions, and particles are usually stop words. Those words that are not associated with an emotion do not contribute to detection process, and therefore they are unnecessary. These words list will be predefined for this system. There are 275 predefined stop words that are listed. In addition to this, the system also defined POS tag set for the stop word list. If a word is in the stop word list, the system removes it from the POS tagged text. If not, the system also finds the word in the stop word POS tag set. If the word is in the set, the system removes it. The stop word removal results of the system are shown in Table 3.

Table 3. Stop word removal result

After POS Tagged	မနေ့/n က/ppm စိတ်ညစ်/v နေ့/part ပေမယ့်/conj အခု/n တော့/part ဝမ်းသာ/v စရာ/part တွေ/part နဲ့/ppm သာယာ/adj နေ့/part တာ/part
After Stop word remove using List	မနေ့/n က/ppm စိတ်ညစ်/v နေ့/part အခု/n တော့/part ဝမ်းသာ/v စရာ/part တွေ/part နဲ့/ppm သာယာ/adj နေ့/part တာ/part
After Stop word remove using Tag Set	စိတ်ညစ်(sad) ဝမ်းသာ(joy) သာယာ(happy)

3.2. Emotion word extraction

After the pre-processing steps are done, the next step is the emotion word extracted by using the Myanmar word-emotion lexicon. Myanmar word-emotion lexicon, M-Lexicon that involves words with the related six emotions values. The words from the pre-processing result list are matched with the words of the M-Lexicon. If a word is settled in the lexicon, the correlated emotion value is also extracted with it and marked as an extracted emotion word. If the word is not in the M-Lexicon, it will be removed from the list. Figure 3 shows the emotion word extraction process of the system.

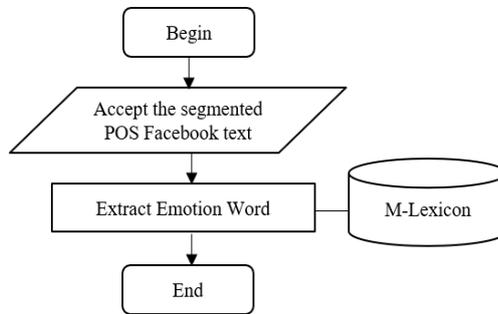


Figure 3. Emotion word extraction process

3.3. Emotion detection

After all, words are matched with the Myanmar word-emotion lexicon words, the system checks, and the extracted word is either a single list or the same emotion value list. If so, the system skips the emotion detection process using machine learning approaches and processes to examine the emotional label step. If there is more than one emotion word with different emotion values, the system goes through the emotion examining process using machine learning approaches. The emotion detection process is shown in Figure 4.

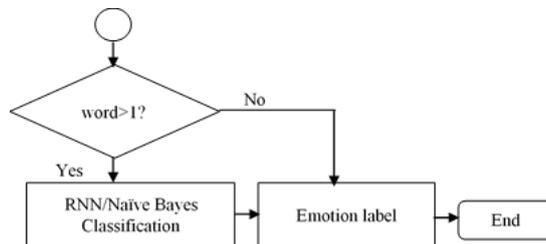


Figure 4. Emotion detection process

3.4. Machine learning based emotion detection

As described in previous section, this process step is worked only when there is more than one emotion word with different emotion values. The system applies RNN emotion classifier and multinomial NB classifier to detect the emotion from the status that contains two or more words with different emotion values. This section will describe that how the two-machine learning based methods are used in emotion detection.

3.4.1. Multinomial naïve Bayes emotion classifier

The system applies multinomial naïve Bayes classifier to examine the emotion from the extracted emotion word for the post and user. NB is a probabilistic classifier which returns the class C^* . The class has the maximum posterior probability given the post as in (1). To return the emotion class, the most probable class C^* is computed by giving some post by selecting the class which has the highest product of the prior probability of the class $P(c)$ and the likelihood of the post $P(x_1, x_2, \dots, x_n | c)$. The value of the count of word with a given class $count(x, c) + 1$ is divided by the sum of the class count $count(c)$ and total words count $|V|$ as in (2). The result of total number of a given class is divided by the total count of all class is to obtain the prior probability of class as in (3).

$$C^* = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{i=1}^n P(x_i | c) \quad (1)$$

$$P(x | c) = \frac{count(x, c) + 1}{count(c) + |V|} \quad (2)$$

$$P(c) = \frac{N_c}{N} \quad (3)$$

The classifier accepts a status and its emotion words extracted by using M-Lexicon and creates a dataset where each status contains the words. The dataset contains the emotion class label (A-F for six emotions) on each status. The classifier creates a text-post vectorized matrix using the extracted emotion words by counting the times of emotion words that appear in each post. The classifier then calculates prior probability, posterior probability, likelihood and finally performs the emotion class label classification and sets the class label for the status. Figure 5 shows the naïve Bayes emotion classifier process.

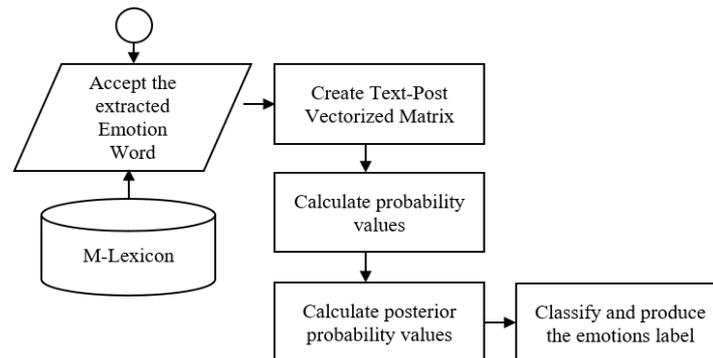


Figure 5. Naïve Bayes emotion classifier process

3.4.2. Recurrent neural network (RNN) emotion classifier

RNN [22] is the extension of the artificial neural network (ANN) that contains memory to understand sequential data. The output depends on previous inputs and performs the same task for all input and output data recurrently. There are three types of RNN architecture such as one to many, many to one, and many to many. This detection system uses many to one RNN architecture where the extracted emotion words are used as input. The input size is equal to the words count. The hidden size is one in which the neuron size is smaller than or equal to the input size. The output layer produces the emotion class label. The rectified linear unit (ReLU) [23] activation function (5) is used in (4) for calculating the hidden state output.

$$h^t = \sigma(W_{in}x^t + W_{rec}h^{t-1}) \quad (4)$$

$$RELU(x) = \max\{0, x\} \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \quad (5)$$

The SoftMax activation function (7) is used in (6) to calculate the output (y) of the output layer.

$$y_t = f(W_{out}h^t) \quad (6)$$

$$SoftMax(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \tag{7}$$

$$y' = argmax(y_i) \tag{8}$$

The input words (x) are encoded by using one-hot encoding [24] and the input weight matrix (W_{in}) is obtained by finding corresponding vector in the pre constructed word embedding matrix. The word embedding matrix was constructed by using word to vector (Word2Vec) [25] continuous bag of words model (CBOW) with the words of M-Lexicon. The recurrent weight matrix (W_{rec}) and output weight matrix (W_{out}) are constructed in which the matrix value is filled by He initialization [26] method. The RNN emotion classifier, shown in Figure 6, takes the extracted emotion words as inputs. The classifier encodes the input words and creates weight matrices. The classifier calculates each hidden state and produces the classifier result by using (8).

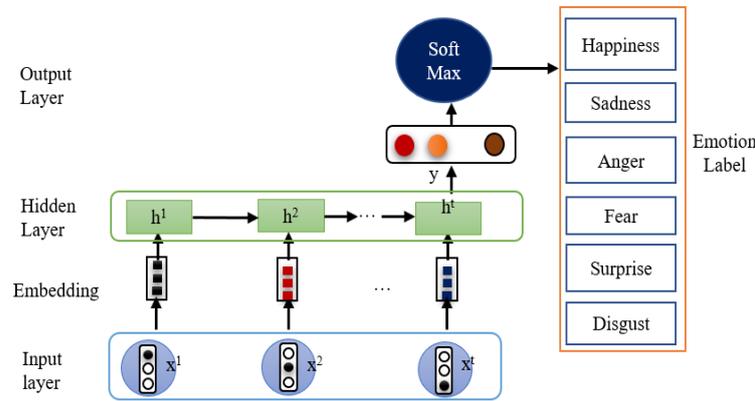


Figure 6. RNN emotion classifier process

3.4.3. Emotion detection result

As discussed in section 3.3, the emotion detection system accepts the segmented POS tagged post as an input. The system finds and deletes the stop words from the input. The system then extracts the emotion words from the input by applying M-Lexicon. If there are more than one emotion word with the different corresponding emotion values (disgust, surprise, anger, fear, sadness, and happiness), for instance ဝမ်းသာ (happiness) (1, 0, 0, 0, 0, 0), စိတ်ညစ် (sadness) (0, 1, 0, 0, 0, 0), the system applies the machine learning-based emotion classifiers. The system produces the detected result from the output of the applied classifiers.

4. RESULT AND DISCUSSION

After implementing the emotion detection system on social media status written in Myanmar language, the system is tested by using 2,000 Facebook status of 27 users. The system defined the reaction-emotion mapping as ‘Love’ and ‘Haha’ reactions are concerned with the ‘Happiness’ emotion, and the ‘Sad’, ‘Wow’ and ‘Angry’ reactions are referred to the ‘Sadness’, ‘Surprise’ and ‘Anger’ respectively. The mapping value is used in the M-Lexicon creation process and emotion labeling process by the expert.

The system experiments by calculating precision, recall, F-measure, and accuracy. Each post is manually labeled as the emotion class by experts which is to be used in performance calculation. Precision is the number of true positive class predictions, and the value is the number of accurately labeled posts retrieved by the process divided by the total number of posts retrieved by the process. The recall statistic is a measure of how well the classifier performs on positive examples in the status dataset. The value is the total number of posts truly labeled and retrieved by the process divided by the number of correctly annotated posts. The F-measure or F1 score is calculated by the results of the precision and recall, and it is the harmonic mean of precision and recall taking both metrics into account. According to Tables 4 and 5, the experiment result shows that average 85% accuracy in NB emotion classifier while 86% accuracy in RNN emotion classifier. The study also tested the time performance of both two machine learning methods for emotion detection which result indicates that the NB emotion classifier is lighter than RNN emotion classifier.

Table 4. Experiments results on naïve Bayes

Emotion	Precision	Recall	F1-score	Accuracy
Happiness	0.928	0.960	0.944	90%
Sadness	0.922	0.952	0.937	89%
Fear	0.863	0.961	0.909	84%
Anger	0.858	0.954	0.904	86%
Surprise	0.857	0.825	0.841	76%
Disgust	0.891	0.932	0.911	85%

Table 5. Experiments results on RNN

Emotion	Precision	Recall	F1-score	Accuracy
Happiness	0.948	0.961	0.954	92%
Sadness	0.938	0.957	0.948	91%
Fear	0.953	0.879	0.915	85%
Anger	0.886	0.916	0.901	85%
Surprise	0.882	0.852	0.867	78%
Disgust	0.936	0.917	0.926	87%

5. CONCLUSION

This paper presented an emotion detection on social media (Facebook) user status. Before examining the emotion of the user from the status, the status content is preprocessed. In the preprocessing step, the content is segmented, stemmed, and performed POS tagging process using myPOS and ripple down rules-based part-of-speech tagger (RDRPOS) tagger to continue the stop word removal process. This system applies our constructed M-Lexicon to extract the emotion words from the status. If there is more than one emotion word with different emotion values, the system applied machine learning based approaches: NB classifier and compared with the RNN emotion classifier. The system produces the emotion class upon the classification result. The system directly produced the emotion class label upon a single word or emotion words with same emotion value. The system tested on 2,000 Facebook status of users. The system states that 86% correctly annotated emotion class by RNN when 85% accuracy in NB classifier. This system does not provide emotion detection from the emoticon, slang words, and negative words.

REFERENCES

- [1] D. H. Hockenbury and S. E. Hockenbury, *Discovering psychology*, 5th edition. Worth Publishers, 2011.
- [2] P. Ekman, "Basic emotions," in *Handbook of Cognition and Emotion*, Chichester, UK: John Wiley & Sons, Ltd, 2005, pp. 45–60.
- [3] T. M. Swe and P. H. Myint, "Word-emotion Lexicon for Myanmar language," in *Studies in Computational Intelligence*, Springer International Publishing, 2020, pp. 157–171.
- [4] H. J. Do and H.-J. Choi, "Korean twitter emotion classification using automatically built emotion lexicons and fine-grained features," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation: Posters*, 2015, pp. 142–150.
- [5] S. Trinh, L. Nguyen, M. Vo, and P. Do, "Lexicon-based sentiment analysis of Facebook comments in Vietnamese language," in *Recent Developments in Intelligent Information and Database Systems*, Springer International Publishing, 2016, pp. 263–276.
- [6] A. Abdaoui, J. Azé, S. Bringay, and P. Poncelet, "FEEL: a French expanded emotion Lexicon," *Language Resources and Evaluation*, vol. 51, no. 3, pp. 833–855, Sep. 2017, doi: 10.1007/s10579-016-9364-5.
- [7] M. Abdullah, M. AlMasawa, I. Makki, M. Alsolmi, and S. Mahrous, "Emotions extraction from Arabic tweets," *International Journal of Computers and Applications*, vol. 42, no. 7, pp. 661–675, Oct. 2020, doi: 10.1080/1206212X.2018.1482395.
- [8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, Jun. 2011, doi: 10.1162/COLLA_00049.
- [9] M. A. Toçoğlu and A. Alpkoçak, "Lexicon-based emotion analysis in Turkish," *Turkish Journal of Electrical Engineering and Computer Sciences*, pp. 1213–1227, Mar. 2019, doi: 10.3906/elk-1807-41.
- [10] O. Sharif, M. M. Hoque, and E. Hossain, "Sentiment analysis of Bengali texts on online restaurant reviews using multinomial naïve Bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934655.
- [11] M. Baali and N. Ghneim, "Emotion analysis of Arabic tweets using deep learning approach," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0252-x.
- [12] O. Araque, R. Barbado, J. F. Sanchez-Rada, and C. A. Iglesias, "Applying recurrent neural networks to sentiment analysis of Spanish tweets," *TASS 2017: Workshop on Semantic Analysis at SEPLN*, pp. 71–76, 2017.
- [13] J. Clement, "Number of monthly active Facebook users worldwide as of 4th quarter 2019," *Statistica*, 2020.
- [14] T. A. Pemppek, Y. A. Yermolayeva, and S. L. Calvert, "College students' social networking experiences on Facebook," *Journal of Applied Developmental Psychology*, vol. 30, no. 3, pp. 227–238, May 2009, doi: 10.1016/j.appdev.2008.12.010.
- [15] M. language Commission, *Myanmar-English dictionary*, 11th edition. 2nd ed, Department of Myanmar Language Commission, Ministry of Education, Myanmar, 2019, 2011.
- [16] N. L. Aung, "Myan-word-breaker," Github, Accessed: Mar. 14, 2023. [Online]. Available: <https://github.com/stevenay/myan-word-breaker>.
- [17] T. T. Thet, J.-C. Na, and W. K. Ko, "Word segmentation for the Myanmar language," *Journal of Information Science*, vol. 34, no. 5, pp. 688–704, Oct. 2008, doi: 10.1177/0165551507086258.
- [18] Z. Z. Hlaing, Y. K. Thu, M. M. N. Wai, T. Supnithi, and P. Netisopakul, "Myanmar POS resource extension effects on automatic tagging methods," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAInLP)*, Nov. 2020, pp. 1–6, doi: 10.1109/ISAInLP51646.2020.9376835.
- [19] K. W. W. Htiike, Y. K. Thu, Z. Zhang, W. P. Pa, Y. Sagisaka, and N. Iwahashi, "Comparison of six POS tagging methods on 10K sentences Myanmar language (Burmese) POS tagged corpus," *Proceedings of the CICLING*, 2017.
- [20] D. Q. Nguyen, D. Q. Nguyen, D. D. Pham, and S. B. Pham, "RDRPOSTagger: A ripple down rules-based part-of-speech tagger," in *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, pp. 17–20, doi: 10.3115/v1/E14-2005.
- [21] R. T.-W. Lo, B. He, and I. Ounis, "Automatically building a stopword list for an information retrieval system," in *Journal on Digital Information Management: Special Issue on the 5th Dutch-Belgian Information Retrieval Workshop (DIR)*, 2005, pp. 17–24.

- [22] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, Mar. 1990, doi: 10.1207/s15516709cog1402_1.
- [23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.
- [24] B. Lantz, "Finding groups of data-clustering with k-means," in *Machine Learning with R*, Packt Publishing Limited, 2013, pp. 73–74.
- [25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *Proceedings of Workshop at ICLR*, Jan. 2013, doi: 10.48550/arXiv.1301.3781.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1026–1034, doi: 10.1109/ICCV.2015.123.

BIOGRAPHIES OF AUTHORS



Thiri Marlar Swe    received the B.C.Sc. degree in Computer Science from Computer University, Mandalay in 2009 and the M.C.Sc. degree in Computer Science from Computer University, Mandalay in 2012. Currently, she is a Lecturer at the Faculty of Computer Science, University of Computer Studies (Banmaw). She is currently pursuing a Ph.D. degree in Information Technology in University of Computer Studies, Mandalay. She can be contacted at email: thirimarlarswe@gmail.com.



Naw Lay Wah    holds a Ph.D. in Information Technology from University of Computer Studies, Mandalay. She is currently at the Faculty of Information Science, University of Computer Studies, Mandalay. She can be contacted at email: nawlaywah@ucum.edu.mm.