

## Sentiment Analysis on Political News

<sup>1</sup>Aye Yu Zin, <sup>2</sup>Nyein Thwet Thwet Aung\*

<sup>1</sup>University of Information Technology, Yangon, Myanmar

<sup>2</sup>University of Information Technology, Yangon, Myanmar  
[ayzin@uit.edu.mm](mailto:ayzin@uit.edu.mm), [nyeinthwet@uit.edu.mm](mailto:nyeinthwet@uit.edu.mm)

**Abstract**— Usage of social media has been increased day by day. People share their thoughts, activities, feelings and knowledge by using social media and they make interactions by commenting or updating status to communicate each other on social media websites. In Myanmar, Facebook is widely used for all types of discussion on every kind of topics. Those discussions can express the opinion of Myanmar people. Those kinds of opinion are very important for decision making. The proposed system makes opinion mining or sentiment analysis on Myanmar people about the current condition of Changing Policies of the Basic Rule of our government. Opinions are extracted from comments of Facebook news articles which are written in Myanmar language. The implemented method uses the hybrid approach of sentiment analysis that is the combination of rule based approach and machine learning approach. At first, collection of positive and negative words is made for construction of sentiment lexicon and then the pre-processing state of text mining is used to extract sentiment words. After that, Naïve Bayes classifier is used to classify the comments into positive or negative comments on political news. Result outcomes are evaluated by accuracy and precision. The proposed system challenges in sentiment analysis on Myanmar Language and hope to provide for future developments in Myanmar Language.

**Keywords:** Myanmar Language, Sentiment lexicon, Machine learning, Sentiment analysis, Opinion mining

### I. INTRODUCTION

Nowadays, use of the internet is rapidly growing and users of internet are simultaneously doing their works on the internet. Internet users share their knowledge and reviews by using

social media networks. Online reviews by users are available on the Internet and it can help to decide and ensure the quality of any item, product or entity. Researchers can make sentiment analysis by using those online reviews of users to provide an important decision. Sentiment is the way of expressing an opinion or feeling towards something i.e. a product, organization or person. Sentiment analysis or opinion mining can be made with the help of Natural Language Processing (NLP). Sentiment analysis can help to identify the opinion, feeling or emotion of users based on the reviews or comments on the social networks. Currently, there are many social network websites such as Twitter, Instagram, Facebook, LinkedIn, etc. Using social network websites is becoming a part of daily life of people. People can write or share their knowledge, opinion, perceptions for any kind of discussion on any topic i.e. education issues, entertainment, political issues, product reviews, restaurant reviews or movies reviews. This kind of showing opinion of users becomes a very dataset for researchers.

In Myanmar, Facebook is the most popular social media network. Most of the Myanmar people share their thoughts, opinions, feeling, knowledge and experiences by updating status or writing comments on Facebook posts using Myanmar Language. The proposed system intends to analyze the opinions of Myanmar people by making sentiment analysis on Myanmar political news, especially, the current condition of Changing Policies of the Basic Rule of our government. Opinions are classified into positive or negative from comments of Facebook news articles which are written in Myanmar language. In this proposed system, comments from Facebook are classified into positive or negative comments by using lexicon based approach and machine learning approach.

The details of the proposed system will be described in the following sections. The related works and background knowledge will be

described in Section 2. In section 3, the proposed system will be explained in detail. Evaluation methods for the results are shown in section 4 and finally, conclusion is given in section 5.

## II. RELATED WORKS

Many studies about sentiment analysis have been carried out. Sentiment analysis includes text analytics, NLP and Linguistics computation to classify sentiment polarities. Sentiment analysis is a special text mining task used to determine and extract subjective attitude or opinion of people from a given text [7]. Sentiment classification is a basic task in sentiment analysis to classify people opinion which is expressed in the way of text format into different sentiment polarity classes that is positive class and negative class [7].

In [5], the creation of Myanmar sentiment lexicon for food and restaurant domain is proposed. Dictionary based approach of lexicon based sentiment analysis is used for analysis of opinion word extraction in food and restaurant domain. 500 customers' reviews were used to extract the opinion words by using the proposed 872 Myanmar Senti-Lexicon of food and restaurant domain.

In [4], an opinion lexicon for Myanmar movies using a bootstrapping approach is constructed. An initial sentiment lexicon containing a small number of opinion words is manually created. Comment is classified as positive or negative using the initial lexicon. The opinion words in the movie domain are chosen from the candidates, based on the statistics of the syllable n-grams, and added to the sentiment lexicon.

In [3], a system for assigning polarity scores to Facebook Myanmar movie comments is proposed. The polarity scores of each positive and negative word in the movie domain-specific polarity lexicon are calculated. And then, the polarity scores to each comment of the plain text movie corpus are assigned.

In [1], sentiment analysis system is proposed for classification of positive and negative information about various cosmetic products. At first, the system identify whether the review comments are subjectivity or not by comparing to the words in cosmetic sentiment lexicon. After this process, the subjectivity comments are kept to determine to be the feature set for the

machine learning process. It used Naïve Bayes classifier to build the classifier. According to experimental results, it is necessary to use positive and negative data in a similar proportion to train the classification application for efficiency result.

In [2], lexicon construction for sentiment analysis is described. At first, pre-processing on raw data comments that are extracted from Facebook news media pages. And then, a lexicon is constructed for classification works. For generating sentiment lexicon, word correlation and chi-square statistic are applied.

In [6], an opinion mining system is proposed to mine opinion of Thai people about Thai government revolution. Traditional pre-processing of text-mining and sentiment lexicon construction are used to extract features from Facebook Status. Comparative experiments among Naïve Bayes, SVM, Decision Tree and KNN are performed.

## III. THE PROPOSED METHODOLOGY

The overall process of the proposed method is shown in Figure 1.

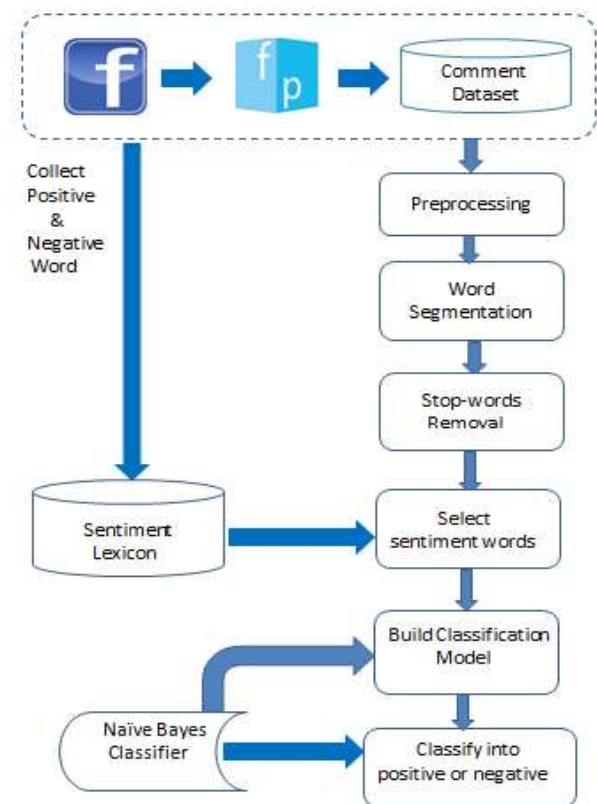


Figure 1. The proposed system

Firstly, data collection from Facebook is made. There are Facebook pages that update news

articles about Myanmar such as Mizzima, Eleven Media Group, The Ladies News, Myanmar Online News, etc. From those Facebook news articles pages, news article post about Changing Policies of Basic Rule of our government is selected and comments from those posts are collected using Facepager Tool. Those comments are about the current opinions of Myanmar people on Changing Polices of Basic Rule. After the data collection, positive and negative sentiment words are collected manually for the construction of sentiment lexicon. The following steps will be performed after dataset collection and constructing sentiment lexicon.

**A. Pre-processing**

Data preprocessing can reduce the complexity of computational processing and can help to produce a higher quality of text classification. Procedures for preprocessing contain the following steps.

**i. Remove Unnecessary Comments:** When Facebook users write comments on Facebook posts, there are also unnecessary comments for the posts. For examples: “Hi”, “Hello”, “ဟိုတယ်”. Those kinds of unnecessary comments for sentiment classification will be removed because it can reduce time complexity of system processing.

**ii. Font Conversion:** In Myanmar Language, there is an issue with fonts. Most of the Facebook users of Myanmar people use Zawgyi font while the development of applications in the technology use Unicode. The comments in the dataset that is collected from Facebook are written with Zawgyi fonts. Therefore, it needs to convert from Zawgyi fonts to Unicode using an online converter.

**iii. Remove Emoticons:** When Facebook users write comments on Facebook posts, users write text comments with emoticons. For examples: ☺, ☹, :-D, T-T, etc. Only emoticons are removed but the text comments are kept.

**iv. Translate meaningful English to Myanmar:** There are some English words that can translate to similar Myanmar words. For example: ‘Like’ = ‘ကြိုက်သည်’ ‘support’= ‘ထောက်ခံသည်’ ‘we will be rooting for you’= ‘အထောက်အပံ့မေ့မယ်’ And there are also some English acronyms that can be translated as sentiment words. For example: ‘LOL’= ‘ရယ်ရယ်အရယ်အရယ်’

**v. Translate Myanglish words:** There are also Myanmar words that are written in English pronunciation or Myanglish words. For example: ‘အရမ်းအတုတ်ယပ်’ (‘very good’ in English) = ‘အရမ်းအထင်အတတ်ယပ်’ Some Myanglish comments are useful for sentiment analysis so those kinds of comments are translated into Myanmar Language. For example, ‘taw taw soe soe war war pl’ (‘very bad’ in English) = ‘တောင်တောင်ပိတ်ပိတ်စစ်စစ်’

**vi. Correct Wrong spelling words:** In comments, there are wrong spelling words. Therefore, it needs to change into correct form of words.

These preprocessing stages are necessary because if the analysis of data that has not been carefully investigated can produce wrong outputs and misleading results. Therefore, the above preprocessing rules are needed and it can help to reduce the computational complexity of system.

**B. Word Segmentation & Stop words Removal**

Word segmentation is the process of determining word boundaries in a piece of text. Word segmentation is an essential step prior to natural language processing in the Myanmar language, because a Myanmar text is a string of characters without explicit word boundary delimiters. For word segmentation, Myan-Word-Breaker which is developed by Nay Lin Aung is used (<https://github.com/stevenay/myan-word-breaker#myan-word-breaker>). It is a word segmentation tool for Myanmar Sentences. After word segmentation, it needs to remove stop words of Myanmar Language since it can help to reduce time complexity of system processing.

**C. Sentiment Words Selection**

After doing stop words removal, sentiment words selection is performed. All the words of comments are compared to vocabularies in the constructed sentiment lexicon to identify if there is sentiment word or not in the comments. Even if there is at least one sentiment word in the comment, that comment will be kept to perform the classification process. If there is no sentiment word in the comment, that comment will be discarded.

### D. Building Classification Model

To classify the comments into positive or negative comments, the preprocessed datasets are divided into the training dataset and the testing dataset. The training dataset is used to build the classifier based on Naïve Bayes method. Naïve Bayes classifier can provide high performance for the probabilities of words found in the comments to classify the comments into negative or positive comments.

$$P(w_k|+) = \frac{n_k + 1}{n + |\text{Vocabulary}|} \quad (1)$$

$$P(w_k|-) = \frac{n_k + 1}{n + |\text{Vocabulary}|} \quad (2)$$

For training dataset calculation, the probability of positive sentiment word is calculated by (1) where  $n$  is the number of the positive case and  $n_k$  is the number of times word  $k$  in the positive case. And also, the probability of negative sentiment word is calculated by (2) where  $n$  is the number of the negative case and  $n_k$  is the number of times word  $k$  in the negative case. After that, the following equation is used:

$$V_{NB} = \underset{v_j \in V}{\text{argmax}} P(v_j) \prod_{w \in \text{words}} P(w|v_j) \quad (3)$$

For the classification of the new sentence is calculated according to (3) where  $V_{NB}$  is the value of Naïve Bayes classifier and the  $\text{argmax}$  means the value that gives us the highest value. If a sentimental word gives the greater negative value than positive value, the comment which contains that word will be classified as a negative comment.

### IV. EVALUATION

To test the performance of the implemented methods of the system, evaluation can be calculated according to accuracy, precision and recall.

$$\%Accuracy = \frac{\text{The total number of correct predicted comments} * 100}{\text{The total number of all comments}} \quad (4)$$

$$Precision = \frac{\text{The number of correct predicted comments}}{\text{The number of predicted comments}} \quad (5)$$

$$Recall = \frac{\text{The number of correct predicted comments}}{\text{The number of actual comments}} \quad (6)$$

The accuracy of positive and negative comments classification are calculated by (4) and the results will be shown in percentage. While making evaluation of the machine learning based classification system, it is not enough by evaluating only with accuracy. Therefore, precision and recall of the system will be calculated according to (5) and (6), respectively to identify the performance of the system.

### V. CONCLUSION

The use of Internet becomes really intensive, especially; using social media becomes a daily life style of Myanmar people. Therefore, there is an opportunity for making application developments or researches in technology fields. And also it is really necessary to mature the development of Myanmar Language for further developments. Myanmar language is still needed to develop in every domain and we have to try to improve the development of Myanmar language. In this proposed system, sentiment lexicon is manually created at first. And the preprocessing stages of text mining are performed for the data cleaning. After that, Naïve Bayes classifier is built for classification of political news comments into negative or positive. We hope the proposed system can be a help for further development of Myanmar language.

### ACKNOWLEDGMENT

I would like to thank Prof. Saw Sanda Aye, Rector of University of Information Technology (UIT), Yangon, Myanmar and Prof. Swe Zin Hlaing and aslo my supervisor Dr. Nyein Thwet Thwet Aung for their valuable support for me to make this system.

### REFERENCES

- [1] Pakawan Pugsee, Patcharaporn Sombatsari, Rujipom Juntiwakul, *Satisfactory Analysis for Cosmetic Product Review comments*, DMCT'17, May 25-27, 2017.
- [2] Zin, T.T, Yar, K.T, Htay, S.S, Htwe, K.K, Aung, N.T.T and Thant, W.W, *Domain-Specific Sentiment Lexicon for Classification*
- [3] Yar, K.T, Htay, S.S, Htwe, K.K, Aung, N.T.T and Thant, W.W, *Assigning Polarity Score to Facebook Myanmar Movie Comments*, International Journal of Computer Application(0975-8887), Volume 177-No.6, November 2017
- [4] Thant, W.W and Kiyooki Shirai, *Automatic Acquisition of Opinion Words from Myanmar Facebook Movie Comments*.

- [5] Aye, Y.M and Aung, S.S, *Senti-Lexicon and Analysis for Restaurant Reviews of Myanmar Text*, IJAEMS [Volume 4, issue-5, May 2018], ISSN:2454-1311
- [6] Panida Songram and Chatklaw Jareanpon, *Opinion Mining of Thai Politics on Facebook Status Updates*, ICISIP(2015)
- [7] Nialm V.Kolekar, Gauri Rao, Sumita Dey, Madhavi Mane, Veena Jadhav and Shruti Patil, *Sentiment Analysis and Classification using Lexicon-based approach and Addressing Polarity Shift Problem*, JATIT, 15<sup>th</sup> August 2016, Vol 90, No.1
- [8] L. Yang, X.H. Yu, Z.S. Cheng and L. Bing, *Sentiment Analysis of Sentences with Modalities*, 2013 ACM 978-1-4503-2415-1/13/10, Unstructured NLP'13
- [9] R.B. Hubert, E. Estevez, A. Maguitman and T. Jamoswki, *Examining Government-Citizen Interactions on Twitter Using Visual and Sentiment Analysis*, dg.o'18, May 30-June 1. 2018, Delft, Netherlands.
- [10] Z. Zhang and B. Varadaraja, *Utility Scoring of Product Reviews*, 2006 ACM 1-59593-433-2/06/0011
- [11] <https://github.com/stevenay/myan-word-breaker#myan-word-breaker>