Sentiment Analysis on Myanmar News Articles

Thet Thet Zin, Su Su Htay, Nyein Thwet Thwet Aung, Win Win Thant, KayThiYar, KhineKhineHtwe
University of Information Technology (UIT)
Myanmar
thetthetzin@uit.edu.mm, susuhtay@uit.edu.mm, nyeinthwetthwetaung@uit.edu.mm, winwinthant@uit.edu.mm

Abstract— **Sentiment Analysis (SA) is one of the most active research areas in Natural Language Processing (NLP), web/social network mining and text/multimedia data mining. Traditional news agencies on online social media allow news consumers to express their opinions about the news articles. The analysis of news articles helps to measure and understand the social importance of many news of events like about 21st century Panglong Conference, Yangon Bus Service (YBS) Transportation, and so on. The sentiment analysis for Myanmar news upon social media is rarely conducted among researchers to the best of our knowledge. This research aims to mine opinion of Myanmar people upon news articles from Facebook news media pages written in Myanmar language. It consists of two main steps: (1) extract subjective sentences and (2) sentiment analyze and classify from extracted subjective sentences. The main outcome of the research will be positive and negative human opinion of Myanmar news comments. Therefore, this research will be not only opened a window to tap into the psychological thinking but also studied the general mind-state of communities especially for Ministries of the government. Knowing news consumer reactions are very useful for decision making of politicians and policy makers.**

Keywords—Sentiment Analysis (SA); Natural Language Processing (NLP); social media, news articles.

I. INTRODUCTION

Sentiment analysis and opinion mining are fields of study that analyze people's opinions, evaluations, attitudes, and emotions generally from written language. Because of the explosive communication and information exchanges using social media, researchers are now given the opportunity to access a tremendous amount of texts and images that express people's opinions and sentiments. Therefore, research in sentiment analysis not only has an important impact on Natural Language Processing but may also have a profound impact on management sciences, political science, economics, and social sciences as they are all affected by opinions. Human perception and user opinion has greater potential for knowledge discovery and decision support. It is a natural phenomenon that good decision can be taken on the basis of opinion of others.

Now day's social media is the best tool to know about the people opinion, advice, comment, complement and their perception about any news articles, product policy and services. We are utilizing social media for the manufacturer and consumer, service provider and user, policies and its effects. In our country, most people use social media especially Facebook to express their feeling using Myanmar language about specific news or product. Moreover most of the popular Myanmar news agencies have gone to social media to allow news consumers to express their opinions about the news articles. Our country have changed the government and set new policies in every country sectors. New government is public oriented government. So people's opinion is very important for every government activities. And need to analyze these people's opinions. Our research intended to mark people's opinion on important news articles.

Most of the researches in sentiment analysis of website, blog, forum and social media have been performed in English or European languages where words are clearly defined by word delimiter such as white space or other special symbols. However Myanmar texts are sequence of characters without word boundaries and highly ambiguous language. Thus, applying the sentiment analysis to Myanmar news users' comments, Myanmar texts need to be parsed and tokenized into individual words first before analysis. Because of this problem, this research proposes a sentiment analysis technique is based on the integration of Myanmar word extraction and sentiment analysis techniques for mining Myanmar news consumer's opinion.

II. RELATED WORK

Explicit Feedback on Myanmar news is required from social network users for getting their opinion on this news. Researches from other countries also do this work with different methodologies on different languages. The authors in [1] presented a survey regarding the presently available technique, application and problem that appeared in the field of

opinion mining. They introduced sentiment classification problem in different level i.e. document-level, sentence-level and aspect-level. Also, some techniques that have been used to solve these problems have been introduced. They described that one of the main problem in the sentiment analysis was the classification of positive and negative sentiments recorded by the users in the social group. They analyzed for Indian and Non-Indian languages. Kiran, et al. [2] provided a platform for classifying news articles using sentiment analysis and created a positive environment among the society. They used POS tagger of Standard NLP group to tag the different words in a sentence and sentiwordnet to get the sentiscore of attributes i.e. for adjectives, adverbs and verbs are focused as their features. They applied some weighting scheme such as Binary, Term-Frequency (tf), Term-Frequency Inverse Document Frequency (tf-idf) and Support Vector Machine (SVM). The authors in [3] analyzed to mine opinion of Thai people about the current government revolution. The opinion was extracted from the Facebook status updates which were written in Thai language. They tried to extract features from the status updates using traditional pre-process of text mining and from the lexicon which requires a collection of positive and negative words. They performed comparative experiments among Naïve Bayes, SVM (Support Vector Machines), KNN (K-Nearest-Neighbor) and decision trees. They showed that that KNN gave highest accuracy when using the lexicon for feature extraction. They pointed out that adding acronyms' lexicon, emoticons' lexicon, and interjections' lexicon may be improve the accuracy in the future works.

Lexicon-based approaches classify the sentiment of the text by analyzing the sense of the opinion words methods, which measure the polarity of text based on lexicons, such as WordNet , MPQAlexicon and SentiWordNet. Their approach [5] is based on integration of various lexicons and dictionary resources for sentiment. The paper is concerned with the analysis of Twitter messages called tweets. They proposed the lexicon-based framework for sentiment classification, which classified tweets as a positive, negative, or neutral. Their framework also detected and scored the slangs used in the tweets. They performed two tasks in Subjective Text Identification: (i) Extracting opinionated word from the source text (ii) Slangs detection and translation. Algorithm for sentiment scoring is described in the paper. They showed that the system needs to improve the precision in negative cases and recall in neutral cases and planned to expand their framework by testing with other datasets.

This proposed system intends to support ministries of our government to recognize the people opinion on government's activities via news on social network. To achieve this purpose, people opinions from social network are collected using specific tools and classify these opinions with SVM.

III. Challenges of the research

The hypothesis here is that the polarity of news comments can be identified. Most work on opinion mining has been carried out on users' opinion on product or services. In this case, users express their opinion freely. The situation is different in comments upon news articles. Some news consumers do not exactly know about the news. So they do not dare to express their feeling freely. Thus, they may resort to other means to express their opinion such as embedding their opinion in a more complex statements using informal or conversation words. To identify sentiment that is not expressed lexically is rather difficult. Another difference is Myanmar language has many synonym words in expressing opinion. Moreover, Myanmar language resources such as lexicon, sentiment dictionary and dataset for this research are not present yet.

IV. Motivation and contributions

"What other people think" has always been an important piece of information for most of us during the decision-making process and more and more people are making their opinions available to strangers via the Internet. With the explosion of user generated content came to the need by companies, politicians, service providers, social psychologists, researchers to mine and analyze the content for different uses. The bulk of this user generated content required the use of automated techniques for mining and analyzing since manually mining and analyzing is difficult for such a huge content.

Public opinion is important for government, companies, service provider and product manufacturer. As you know, our government has changed policies in every sector. The most important thing for new government is that what the public opinions on his changes are? To know this important fact we need to analyze public opinion systematically. Mostly news agencies express government activities and policies on their press and online media. Myanmar people also express their feeling on news articles by writing comments using Myanmar language. Today in Myanmar, most of the news agencies use Facebook pages to express their news articles and most people read news from Facebook official pages. Thus, we are encouraged to analyze people opinion on news article especially political news using Facebook pages news articles.

However, in the previous section some of the issues of this work are posed by short, informal texts, such as the comments. To achieve this research goal, this paper brings some contributions. First of all, the pre-processing of comments to normalize the

language and generalize the vocabulary employed to express sentiment. At this stage, the linguistic features of comments are taken into account from the training data with a unique label. According to the nature of Myanmar language, we need to employ tokenization on comments. Second is the use of simple heuristics to select features to be employed. The final feature selection is influenced by the data at hand and it is difficult to employ on new sets of data. This work selects the features to be employed in the classification model based on the condition that they should occur at least once in the training set. The last one is the application of supervised learning using a simple Support Vector Machines linear classifier on a set of realistic data. In the near time, the result will show that using the method described can improve the sentiment classification performance.

V. METHODOLOGY OF THE RESEARCH

Sentimental analysis is all about to get the real voice of people towards specific product, services, organization, movies, news, events, issues and their attributes. Sentiment analysis includes branches of computer science like Natural Language Processing, Machine Learning, Text Mining and Information Theory and Coding. By using approaches, methods, techniques and models of defined branches, we can categorized our data which is unstructured data may be in form of news articles, blogs, tweets, movie reviews, product reviews etc. into positive, negative or neural sentiment according to the sentiment is expressed in them.

Therefore, we would like to investigate structured models for document-level sentiment classification. Document-level sentiment classification aims to automate the task of classifying a textual review, which is given on a single topic, as expressing a positive or negative or neutral sentiment. When predicting the sentiment of a subjective document (e.g., as positive or negative), it is well known that not all sentences are equally discriminative or informative. But identifying the useful sentences automatically is itself a difficult learning problem. We would like to propose a two-level approach for document-level sentiment classification that simultaneously extracts useful (i.e., subjective) sentences and predicts document-level sentiment based on the extracted sentences. There are two types of classification techniques that have been used in document-level sentiment classification such as supervised method and unsupervised method. For classification of user's comments, we aim to use supervised approach. Figure shows the workflow of the proposed system.
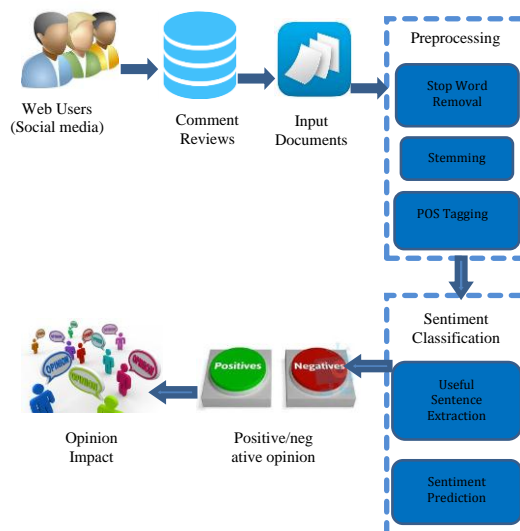

Fig. -1. The workflow of the proposed system

VI. CONCLUSION

This research proposed to mine opinion of Myanmar news articles from Facebook news pages. Two-level approach for document-level sentiment classification will be proposed to simultaneously extracts useful (i.e., subjective) sentences and predicts document-level sentiment based on the extracted sentences. For classification of documents, we aim to use supervised approach. To the best of our knowledge, sentiment analysis on Myanmar language does not present yet. This research will be useful to analysis people opinion upon news articles especially about government policies. Moreover, UIT students can get knowledge and experience by participating in this research project and we can get Myanmar language resources for future research work. In the future, we would like to extend our work to sentence level and aspect level sentiment classification.

REFERENCES

[1] K.S.Ilakiya,Mrs.M.Lovelin Ponn Felciah,IJCSMC, "Challenges and Techniques for Sentiment Analysis: A Survery", Vol.4, Issue.3, March 2015,pg. 301-307

[2] Kiran Shriniwas Doddi, Dr.Mrs.Y.V.Haribhakta, Dr. Parag Kulkarni, "Sentiment Classification of News Articles", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 4621-4623.

[3] Panida Songram, Chatklaw Jareanpon, "Opinion Mining of Thai Politics on Facebook Status Update", Proceeding of the 3rd IIAE International Conference on Intelligent Systems and Image Processing 2015.

[4] Fazal Masud Kundi, Aurangzeb Khan, Shakeel Ahmad, Muhammad Zubair Asghar, "Lexicon-Based Sentiment Analysis in the Social Web", Journal of Basic and Applied Scientific Research, ISSN 2090-4304-4(6)238-248, 2014.

[5] Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena, "Large-Scale Sentiment Analysis for News and Blogs" ICWSM 2007 Boulder, Colorado, USA.

[6] Yan Zhao, Suyu Dong and Leixiao Li, "Sentiment Analysis on News Comments Based on Supervised Learning Method", International journal of Multimedia and Ubiquitous Engineering, Vol.9, No.7 (2014), pp. 333-346.