

Assigning Polarity Scores to Facebook Myanmar Movie Comments

Win Win Thant

Natural Language Processing
Laboratory
University of Information
Technology (UIT), Yangon,
Myanmar

Nyein Thwet Thwet Aung

Natural Language Processing
Laboratory
University of Information
Technology (UIT), Yangon,
Myanmar

Su Su Htay

Natural Language Processing
Laboratory
University of Information
Technology (UIT), Yangon,
Myanmar

Khine Khine Htwe

Natural Language Processing Laboratory
University of Information Technology (UIT), Yangon,
Myanmar

Kay Thi Yar

Natural Language Processing Laboratory
University of Information Technology (UIT), Yangon,
Myanmar

ABSTRACT

User-generated texts such as reviews, discussions or comments are valuable indicators of users' preferences. Apart from binary classification (positive or negative) of the reviews, some researchers calculated polarity scores that give a very concise summary and provide more information of the reviews. In this paper, a system for assigning polarity scores to Facebook Myanmar movie comments is proposed. Myanmar is a language with underdeveloped electric resources. As this is pioneering work for this combination of language and sentiment analysis, the polarity scores of each positive and negative word in the movie domain-specific polarity lexicon is calculated. And then the polarity scores to each comment of the plain text movie corpus are assigned. The proposed system achieves 89% and 85% accuracy on positive and negative opinion words respectively in the evaluation of polarity score lexicon. We also make the comment polarity for 3-class evaluation and 5-class evaluation based on the scores of comments.

General Terms

Sentiment analysis

Keywords

Polarity score lexicon, Plain text corpus, Facebook movie comments

1. INTRODUCTION

Opinion mining and sentiment analysis have become popular in linguistic resource rich languages. Reviews, comments and opinions of the people play an important role in determining whether a given population is satisfied with a product or a service or in judging their response to the specific event. Data consisting of such reviews or opinions has a very high potential for knowledge discovery. One of the basic tasks in sentiment analysis is to predict the polarity of a given sentence, to find out if it expresses a positive or negative feeling about a certain topic [1]. The sentiment may be represented in several possible forms, some of which are [2]:

- Fixed categorization into positive or negative,
- A finite number of graded sets such as strongly positive, mildly positive, neutral, mildly negative, strongly negative,

- A real value denoting sentiment strength in an interval such as [-1; +1].

While sentiment classification in movie reviews has been extensively studied for English (movie reviews datasets were among the earliest to use for the task), it has not been tried for Myanmar so far. Given the lack of previously developed movie datasets or sentiment polarity lexicons for Myanmar, we had to create them ourselves. In addition to assigning a score to the dataset of movie comments, we further focused on building a sentiment polarity lexicon for Myanmar. To our knowledge, no prior work had been done exclusively on assigning polarity scores to Facebook Myanmar movie comments.

We will relate our work to other works in Section 2. We will give resources of the system in Section 3. We will discuss the major components in more detail, the data cleaning, the polarity classification and the polarity score calculation and generating the scores of comments in Section 4. In Section 5, evaluation results for the analysis system will be presented. The conclusion will be presented in Section 6.

2. RELATED WORK

In this section, we will give the related work about lexicon building approaches, popular lexicons and lexicon building of other languages.

2.1 Lexicon Building Approaches

In [3], the authors discussed three main approaches for opinion lexicon building: manual approach, dictionary-based approach, and corpus-based approach. The major drawback of the manual approach is the cost (time and effort) to hand select words to build such a lexicon. There is also the possibility of missing important words that could be captured with automatic methods. Dictionary-based approaches work by expanding a small set of seed opinion words, with the use of a lexical resource such as the WordNet [4]. The main shortcoming of these approaches is that the resulting lexicon is not domain-specific. Corpus-based approaches can overcome these problems by learning a domain-specific lexicon using a domain corpus of labeled reviews.

2.2 Lexicons

SentiWordNet [5] is a lexical resource devised to support Sentiment Analysis applications. It provides an annotation

based on three numerical sentiment scores (positivity, negativity, neutrality) for each WordNet synset [6]. Clearly, given that this lexical resource provides a synset-based sentiment representation, different senses of the same term may have different sentiment scores.

WordNet-Affect [7] is a linguistic resource for a lexical representation of affective knowledge. It is an extension of WordNet which labels affective-related synsets with affective concepts defined as A-Labels (e.g. the term euphoria is labeled with the concept positive-emotion, the noun illness is labeled with physical state, and so on).

MPQA Subjectivity Lexicon [8] provides a lexicon of 8,222 terms (labeled as subjective expressions), gathered from several sources. This lexicon contains a list of words, along with their POS-tagging, labeled with polarity (positive, negative, neutral) and intensity (strong, weak).

SenticNet [9] is a lexical resource for concept-level sentiment analysis. It relies on the sentic computing, a novel multi-disciplinary paradigm for sentiment analysis. Differently from the previously mentioned resources, SenticNet is able to associate polarity and affective information also to complex concepts such as accomplishing goal, celebrate special occasion and so on. At present, SenticNet provides sentiment scores (in a range between -1 and 1) for 14,000 common sense concepts.

2.3 Lexicon Building of Other Languages

There is a large body of work concerning sentiment analysis using lexicon. This is because lexicons are crucial for sentiment analysis.

The researchers [10] developed a sentiment lexicon for Sinhala Language with the aid of English sentiment lexicon (SentiWordNet 3.0). It is generated by mapping the English words in the English/Sinhala dictionary to the sentiments in SentiWordNet 3.0. The advantage of the method is it does not require a translation tool or software, unlike other multilingual sentiment approaches. They have achieved acceptable results maximum of 60% in Naïve Bayes classification with a Sinhala sentiment lexicon.

The authors [11] presented a graph based WordNet expansion method to generate a full (adjective and adverb) subjective lexicon. They used synonym and antonym relations to expand the initial seed lexicon. They created the resources for Hindi language, Hindi subjective lexicon. They also developed an annotated corpus of Hindi Product Reviews. For dataset in Hindi language, translated pre-annotated Amazon product reviews from English to Hindi using Google translate. They explored how the synonym and antonym relations can be exploited using simple graph traversal to generate the subjectivity lexicon.

They [12] produced the first publicly available large scale Standard Arabic sentiment lexicon (Ar-SenL) using a combination of existing resources: English WordNet (EWN), Arabic Word-Net (AWN 2.0), English SentiWordNet (ESWN 3.0) and the Standard Arabic Morphological Analyzer (SAMA 3.1) [13]. They showed that using English-based linking produces, on average, superior performance in comparison to using the WordNet-based approach. A union of the two resources was better than either and outperformed a high-quality manually-derived adjective sentiment lexicon for Arabic.

For our language, raw corpus and domain specific lexicon are manually created. This is the first work for identifying scores to the comments written in Myanmar language.

3. RESOURCES

Lexical resources such as corpus and lexicon are very important for sentiment analysis. A text corpus, an unstructured set of texts, is used to do sentiment analysis and checking occurrences in the proposed system. Lexicon is a database of lexical units for a language along with their sentiment orientations.

3.1 The Plain Text Corpus

The 12,600 Myanmar movie comments were randomly collected from the Facebook. The plain text corpus includes positive, negative and neutral comments and the numbers of these comments are nearly equal. The minimum length of the comment is 2 words, the maximum comment length is 90 words and the approximate average length is 17 words. For example, 2,310 comments are collected from the Facebook movie link of OKMPmovie (အုပ်ထွေးမြတ်ငါးလို့). The following table shows the example web pages link that the data is collected.

Table 1: Example Data Collected Sites

| No. | Facebook Movie Web Pages Link | No. of comments |
|-----|---|-----------------|
| 1 | https://www.facebook.com/Kyikyikyalkyal/ | 1,950 |
| 2 | https://www.facebook.com/achit.sit.movie/ | 1,460 |
| 3 | https://www.facebook.com/minlebokay/ | 1,120 |
| 4 | https://www.facebook.com/ttndonut/ | 2,140 |
| 5 | https://www.facebook.com/OKMPmovie/ | 2,310 |

3.2 Polarity Lexicon

A lexical resource for sentiment analysis also referred to as a sentiment lexicon, is a database of lexical units for a language along with their sentiment orientations. There are two broad approaches to creation of sentiment lexicons – manual and automated. Creation of a sentiment lexicon manually involves reading several thousand of text documents and selecting the words which are carrying sentiment. In automatic sentiment lexicon, the task is to create a set of starting seed words with known sentiment orientation, and then expand that seed set using an already existing lexical resource. Two stages of lexicons creation can be distinguished: 1) the generation of the sentiment-bearing words list, containing the candidates to sentiment lexicon, and 2) the assignment of sentiment labels to these words, e. g. positive/negative/neutral [14].

In the proposed system, a hand-tagging movie domain-specific lexicon is manually designed. It consists of 276 polarity words after analyzing the plain text corpus. Some are compound words and some include their inflection. Positive words with negative prefix or suffix are considered as negative words. We observe 152 words which have been tagged with a positive sentiment polarity and 124 words tagged with a negative polarity. More positive words are extracted because the positive comments contain more opinion words in the plain text corpus. Examples positive and negative words are shown in Table 2.

Table 2: Examples of words from the lexicon which have been tagged with positive polarity and negative polarity

| Word | Translation | Polarity Type |
|----------------|-------------|---------------|
| ကျေနှစ် | satisfy | |
| ဂုဏ်ယူ | proud | |
| တိုက် | like | Positive |
| ဆွဲဆောင် | attract | |
| ပျော် | happy | |
| ဝေဖြန့် | | |
| အဆင့်အတန်းမရှိ | criticize | |
| ဆိုင်းမီး | no standard | |
| ရှုရှင်း | awful | Negative |
| ရှုရှင်း | shame | |
| ပျော်ဆရာ | bore | |

4. PROPOSED SYSTEM

The comments are fetched from Facebook website which undergoes text cleaning to classify the comments as positive, negative and neutral. Manually created polarity lexicon and classified comments are then used to calculate the scores of each sentiment word. After creating sentiment polarity score lexicon, the comments' scores are generated. The architectural diagram of the proposed system is shown in Figure 1.

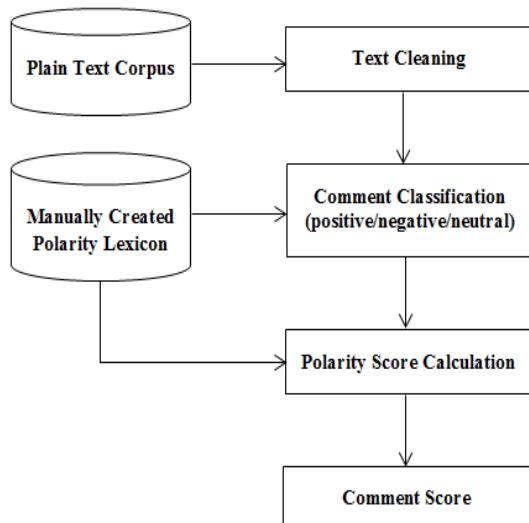


Figure 1: Architectural diagram

4.1 Text Cleaning

In order to perform sentiment analysis, data have to be prepared in order to obtain a dataset. The data preparation step is an essential task in social network data analysis. It performs necessary data preprocessing and cleaning on dataset for subsequent analysis.

- Removal of non-Myanmar words: When data analysis needs to be data driven at the word level, the commonly occurring words (non-Myanmar words) should be removed.
- Removal of Punctuations: All the punctuation marks such as “.”, “,”, “?” are removed from the comments.
- Removal of Expressions: Textual data (usually speech transcripts) may contain human emotional expressions

like [laughing], [crying], [audience paused]. These expressions need to be removed.

- Slangs lookup: Again, social media comprises a majority of slang words. These words should be transformed into standard words to make free text. The words like “နိုင်တယ” means good (ကောင်တယ), “နှစ်းဆန်တယ” means beautiful (လွတယ), etc.
- Spell checking: Some misspelled words are manually checked for the next phases of analysis.

4.2 Identifying Positive, Negative and Neutral Comments

Sentiment classification refers to classifying reviews not by their topics but by the polarity of their sentiment (e.g, positive or negative). It is useful for recommendation systems, fine-grained information retrieval systems, and business applications that collect opinions about a commercial product. The main resources used for identifying positive, negative and neutral comments are the proposed plain text corpus and the constructed lexicon. The approach is to count positive and negative words in a comment, where the comment is considered positive if it contains more positive than negative words, and negative if there are more negative words. A comment is neutral if it contains an equal number of positive and negative words.

4.3 Calculating Polarity Score

We have calculated scores (positive, negative and neutral) of the opinion words by using the following equations.

$$score_{pos}(w) = f_{pos}(w)/f_{all}(w) \quad (1)$$

$$score_{neg}(w) = f_{neg}(w)/f_{all}(w) \quad (2)$$

$$score_{neu}(w) = f_{neu}(w)/f_{all}(w) \quad (3)$$

$$score_{pos}(w) + score_{neg}(w) + score_{neu}(w) = 1 \quad (4)$$

where $f_{pos}(w)$, $f_{neg}(w)$, $f_{neu}(w)$ and $f_{all}(w)$ denote the frequencies of an opinion word (w) occurred in the positive, negative, neutral and all comments, respectively. And $score_{pos}(w)$, $score_{neg}(w)$ and $score_{neu}(w)$ denote the scores of opinion word as positive, negative and neutral word, respectively. Each of these scores ranges in the interval [0, 1], and the total of their scores is 1 for each word.

Polarity score can be expressed as a set of tuples of the form <positive: negative: neutral> and all scores are rounded to 4 decimal places. The score 1:0:0 is absolutely positive and 0:1:0 is absolutely negative. For example, positive opinion word “အေားကေး” (i.e. appreciate) is included 69 times in all comments. This word also appears 58 times in positive comments, 3 times in negative comments and 8 times in neutral comments. The polarity triplet of this word is <0.8406, 0.0435, 0.1159>; hence $score_{pos}(\text{အေားကေး}) = 0.8406$. Some opinion words and their scores are shown in Table 3.

Table 3: Sample words with polarity scores

| word | Positive score | Negative score | Neutral score |
|----------------------|----------------|----------------|---------------|
| ခုစ် (love) | 0.6897 | 0.0345 | 0.2759 |
| သနား (sympathy) | 0.1488 | 0.0000 | 0.8512 |
| တွေ့ကြိုး (like) | 0.8129 | 0.1065 | 0.0806 |
| ပေါကား (silly movie) | 0.0104 | 0.9151 | 0.0745 |
| ဆိုင်း (awful) | 0.0000 | 1.0000 | 0.0000 |

The Myanmar words “မကောင်း” (i.e not good) and “ပွဲ” (i.e. bad) are same meaning but most comments contain “မကောင်း” (i.e not good). Dealing with negative words is an essential step in counting frequency. Because of the mixed frequency of positive word “ကောင်း” (i.e good) in negative word “မကောင်း” (i.e not good), error can be occurred in counting frequency. In order to get the correct count of positive word “good”, this problem is solved by reducing “good” count contained in “not good”.

4.4 Assigning Scores to Comments

The sentiment classification task is extended to the more challenging task of assigning scores to comments. This score is called the comment scores. Using the above polarity score lexicon with positive, negative words, the overall score of a comment can be identified. The first step is to extract opinion words from a comment. However, in some cases, it is needed to combine several opinion words in a sentence as both positive and negative words may exist in a sentence. The second step is to assign the scores of these extracted words. The scores are calculated by subtracting the total negative opinion scores from the total scores of the number of positive opinion words in a comment. It is done by the equation (5).

$$score(C) = \sum_{i=0}^n score_{pos}(P_i) - \sum_{i=0}^n score_{neg}(N_i) \quad (5)$$

C : a comment

P_{1,...,n} : positive opinion words contained in C

N_{1,...,n} : negative opinion words contained in C

Most of the comments are not grammatically corrected sentences. Here, we list some examples of movie comments with scores.

Example positive comment:

မီကတိကားကို တွေ့ကြိုး (like) တယ် ကောင်း (good) တယ်

I like this movie. It is good.

score_{pos}(တွေ့ကြိုး) = 0.9 and score_{pos}(ကောင်း) = 0.825

Comment Score = 0.9 + 0.825 - 0 = 1.725

Nothing really surprising here: typical positive comment, without going into specifics.

Example negative comment:

မြန်မာပေါကား (silly movie) တွေ့ဗိုး

It is a Myanmar silly movie.

score_{neg}(ပေါကား) = 0.9251

Comment Score = 0 - 0.9251 = -0.9251

The comment shows clearly negative.

Example neutral comment:

နေတိုးက သရုပ်ဆောင်တော် (good at acting) တယ် မင်းသမီး ကို မတွေ့ကြိုးဘူး (do not like)

The actor Nay Toe is good at acting. I do not like the actress.

score_{pos}(သရုပ်ဆောင်တော်) = 0.728 and score_{pos}(မတွေ့ကြိုး) = 0.9

Comment Score = 0.728 - 0.9 = -0.172

Here, despite having the low negative score, the comment is really neutral.

5. EVALUATION

There are two parts for evaluation of the proposed system. One is evaluation of lexicon and another evaluation is based on scores.

5.1 Evaluation for Polarity Score Lexicon

For the evaluation of our polarity score lexicon of positive and negative words, accuracy for positive opinion word (w_{pos}) is the ratio of the number of recognized positive opinion words whose scores are greater than 0.6 to the number of all positive opinion words. The accuracy of a positive opinion word is calculated in equation (6) and Accuracy (w_{neg}) can be calculated similarly.

$$\text{Accuracy}(w_{pos}) = \frac{\text{number of recognized positive opinion words}}{\text{number of all positive opinion words}} \quad (6)$$

Table 4: Accuracy of opinion words

| Opinion type | Opinion Words | | Accuracy |
|--------------|---------------|--------------------------|----------|
| | All | Recognized (score > 0.6) | |
| Positive | 152 | 136 | 89% |
| negative | 124 | 105 | 85% |

The proposed system yield 89% of accuracy for positive opinion words and 85% of accuracy for negative opinion words in the lexicon. The scores of some words are not correct because of the following reason.

- In the following example, the whole comment polarity is positive because of two positive opinions and one negative opinion. For example, the word “မထိတ်တန်ဘူး” (i.e. is not worth) is given a negative polarity, but it occurs very often in positive comments and not very often in negative ones. So, the score of this word is not correct.

အတိလမ်း ကောင်း တယ် မင်းသမီး လည်း တော်တယ် မင်းသမီးက အကယ်ဒမိန္ဒာ မထိတ်တန်ဘူး

(The movie is good. The actor is good. The princess is not worth with award of academy.)

- Because of some spelling errors in the comments, some of the word count is mismatched and it affects the accuracy.
- Single opinion words are ambiguous words and can cause the incorrect frequency. In the following example, the Myanmar word “တော်” (i.e. good or stop) is ambiguous word. One is positive meaning and another one is negative meaning.
 မင်္ဂလားက တော်(good) တယ်.
 (The actor is good.)
 ဒီဇော်ကားရို့မြတ်ချင်တော့ဘူး တော်ပါလေ့ (stop)
 (I do not want to see this movie. Please stop!)

5.2 Human Evaluation for Scores of Comments

To test the scores of comments, we performed human labeling of a subset of the movie comments, with positive, negative or neutral polarity labels. In order to have an equal number of positive and negative comments, 300 comments are randomly selected from each category, resulting in the testing dataset of 900 comments. Evaluations on 3-class and 5-class are performed according to the scores.

3-class evaluation: divides the scores into three classes, positive $\geq 0.5 >$ neutral $\geq -0.5 >$ negative

5-class evaluation: divides the scores into five classes, strong positive $\geq 2.5 >$ weak positive $\geq 0.5 >$ neutral $\geq -0.5 >$ weak negative $\geq -2.5 >$ strong negative

The comment scores assigned by the human raters are treated as correct answers. Moreover, comment scores may vary depending on them. The definition of the accuracy in this section is calculated on the scores of comment. In 3-class evaluation, the accuracy of positive comments (C_{pos}) is a proportion of the number of the recognized positive comments to the number of all positive comments in the testing dataset and it is shown in the equation (7). Accuracy (C_{neg}) and Accuracy (C_{neu}) can be calculated similarly. And also accuracy for 5-class evaluation is calculated using the following equation.

$$\text{Accuracy}(C_{pos}) = \frac{\text{number of recognized positive comments}}{\text{number of all positive comments}} \quad (7)$$

Table 5: Result of 3-class Evaluation

| Class | Positive | Negative | Neutral |
|------------|----------|----------|---------|
| All | 300 | 300 | 300 |
| Recognized | 252 | 265 | 245 |
| Accuracy | 84% | 88% | 82% |

For 3-class evaluation, the average accuracy is 85% agreement with human raters and 77% agreement for 5-class evaluation.

6. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a domain-specific based sentiment analysis model for classification of movie comments into assignment of scores to the comments. The proposed method is very simple. There is no previous work and we cannot make comparison at the current.

For future work, we are going to create a larger dataset with more lexicon words. Another way to improve the accuracy of the polarity scores could be automatically built lexicon with more cleaned text. One direction of future research is to apply the role of intensifier and diminishers and solve complex Myanmar comments in this work. From the error analysis, we found that the major problem is word sense ambiguity. This work can be extended to incorporate Word Sense Disambiguation (WSD). The future goal is to enlarge the set of opinion words by further automatic process. Another plan is to construct the lexicon and corpus for other different domains and assign scores.

7. ACKNOWLEDGMENTS

The authors would like to thank Prof. Saw Sanda Aye, Rector, University of Information Technology (UIT), Yangon, Myanmar and Associate Prof. Kiyoaki Shirai, School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology (JAIST), Japan for their valuable support.

8. REFERENCES

- Pang, B. and Lee, L. Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–135.
- Ahire, S. A Survey of Sentiment Lexicons, 2015.
- Liu B. and Zhang, L. A survey of opinion mining and sentiment analysis in Mining Text Data, C. C. Aggarwal and C. Zhai, Eds. Springer, 2012, pp. 415–463.
- Fellbaum, C. Eds., WordNet: An Electronic Lexical Database, Language, Speech, and Communication MIT Press, Cambridge, MA, 1998.
- Esuli, A., Baccianella, S. and Sebastiani, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, In Proceedings of LREC, Volume 10, 2010, pp 2200–2204.
- Miller, G. A. WordNet: a lexical database for English, Communications of the ACM, Volume 38, Issue 11, Nov. 1995, pp 39–41.
- Strapparava, C. and Valitutti, A. Wordnet affect: an affective extension of wordnet, In Proceedings of LREC, Volume 4, 2004, pp 1083–1086.
- Wiebe, J., Wilson, T. and Cardie, C. Annotating expressions of opinions and emotions in language, Language Resources and Evaluation, Volume 39, Issue 2–3, May 2005, pp 165–210.
- Cambria, E., Olsher, D. and Rajagopal, D. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis, In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Quebec City, 2014, pp 1515–1521.
- Medagoda, N. , Shanmuganathan, S. , Whalley, J. Sentiment lexicon construction using SentiWordNet 3.0, 11th International Conference on Natural Computation (ICNC), 2015, pp 802–807.

- [11] Bakliwal, A., Arora, P. and Varma, V. Hindi Subjective Lexicon: A Lexical Resource for Hindi Adjective Polarity Classification, In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC), May 2012.
- [12] Badaro, G., Baly, R. and Hajj, H. A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining, In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), pages 165–173, October 25, 2014, Doha, Qatar. © Association for Computational Linguistics.
- [13] Mohamed, M. et al. LDC Standard Arabic Morphological Analyzer (SAMA) Version 3.1 LDC2010L01. Web Download. Philadelphia: Linguistic Data Consortium, 2010.
- [14] Kotelnikov, E.V., Bushmeleva, N.A., Razova, E.V., Peskisheva, T.A. and Pletneva, M.V. Manually created sentiment lexicons: research and development, Computational Linguistics and Intellectual Technologies. Volume 15, Issue 22, 2016, pp. 281-295.