The State Art of Text Sentiment from Opinions to Emotion Mining

[1*]Nor Anis Asma Sulaiman, Leelavathi Rajamanickam

Faculty of Engineering, Built Environment & IT, SEGi University

* anissulaiman@segi.edu.my

Abstract: This study is aiming to analyse the feelings expressed by the users in a text on a comment posted on social media. Text Mining and Emotion Mining can be analysed by using both technique of Natural Processing Language (NLP). Mostly on the previous study of text mining is using unsupervised technique and referring to Ekman's Emotion Model (EEM) but it has restrained coverage of polarity shifters, negations and lack emoticon. In this study have proposed a Naïve Bayes algorithm as a tool to produce users' emotion pattern. The most important contribution of this study is to visualize the emotion's theory with the text sentiment based on the computational methods for classifying users' feelings from natural language text. Then, the general system framework of extracting opinions to emotion mining has produced and capable use in any domains.

Keywords: Emotion, Text, Sentiment, Naïve Bayes

1. Introduction

Social Media as an important platform nowadays to people for sharing their opinion, ideas, and expressing their feeling and emotion by writing the comments and used of emoticon (Wu, Li, Shen, & He, 2020). Facebook is one of the most popular Social Media platforms today. It contains comments posted by users where they can express their feelings and opinions through texts or emoticons (Ahuja, et.al, 2019).

The purpose of this paper is to study how Facebook platform can be used for classifying the text based on a comment in the posted issue to emotion mining (Palit & Ghosh, 2020). The emotions indicator used to produce a pattern of user's emotions splits automatically between three sets of texts:

1. Positive Emotion: texts contain positive words such as happy, peaceful, love, respect, harmony

2. Negative Emotion: texts contain unhappy, sad, bad

3. Neutral [Emotionless]: text that only state facts or do not express any emotions, misunderstanding and hope.

This study focuses on randomly selected posts from the *Malaysiakini* Facebook page. Software tools have been developed using python language to classify text sentiment from user's opinion to emotion mining by using Naive Bayes algorithm (Chandra & Jana, 2020). The specified training set is stored in the database as a corpus for classification purposes. Then, the data will process to produce a user's emotional pattern and display results based on the percentage score of each emotional polarity. The software tool was developed allow the educators, Administrations or businessmen to organize and monitor the issues posted in the Facebook to able enhance their quality of services.

This rest of the paper is organized as follows. In Section 2 discuss the related works on the sentiment analysis process, while Section 3 shows the research framework. Then, the result and discussion in section 4. The last section which is section 5 is a conclusion about the related work.

2. Materials and methods

Sentiment Analysis (SA) is referring to the application of Natural Language Processing (NLP), Computational Linguistics (CL), and Text Analysis (TA) to determine and extract the subjective information from anonymous data using Social Network Sites (SNS) such as YouTube, Google+ Facebook, Instagram, Twitter, Pinterest, and Q-Zone (Giachanou & Crestani, 2016; Mäntylä, Graziotin, & Kuutila, 2018; Zhang, Wang, & Liu, 2018). The aim of sentiment analysis is to identify the people's attitude or emotions toward issues have been stated. Then, presents in the statically a positive or a negative pattern.

The Sentiment Analysis has been utilized in numerous parts mostly in the business, movie rating review, product survey, the customer's experience, governmental issues, education, laws and other (Bose, et.al., 2020; Ma, Li, & Wang,

2020). The sentiment is analysed after all the opinion posted are extracted. SA consists of several processes which include extraction and classification of data.

Based on the previous research, various methods implemented to extract and classify opinion of people's especially in subjective term. For examples manually process, semi-automatically, fully automatically, and online tools (Lennox, et.al., 2020)

Manually approach is limited questions and answers provided as and this is due to the time constraints to conduct research to obtain data from the public (Liu & Shen, 2020). Unlike the semi and automatic approach there are no limited time and people can access every time and everywhere. In additional, the data have extracted immediately to perform filtration.

After data filtered, in order to classify the texts into indicators of emotions the method used is the algorithm. The various algorithm implemented in past research. For example, Bag-of-word, Sentiment Identification, Decision Tree, Neural Network, Support Vector Machine (SVM), and Naïve Bayes. By using this algorithm approach, emotions can be classified into three different sets of texts that are positive, negative or neutral which mean no express stated in the texts (Barrón Estrada, et.al., 2020). The best research obtained for classifying texts into three sets of training by using the Naïve Bayes algorithm. The accuracy of the result is up to 81% showed up on the test set.

Based on the previous study had done on Facebook, most research had carried out is for people's emotions, interaction, activities, behaviors, usage of Facebook, and personality (Xu et al., 2020). This research focuses on classifying people's emotions in Facebook's comments by using Naïve Bayes.

3. Methodology

The system architecture of this study consists of several phases such as Pre-processing Requirement, Feature Extraction, Analysis data and Produce Pattern.
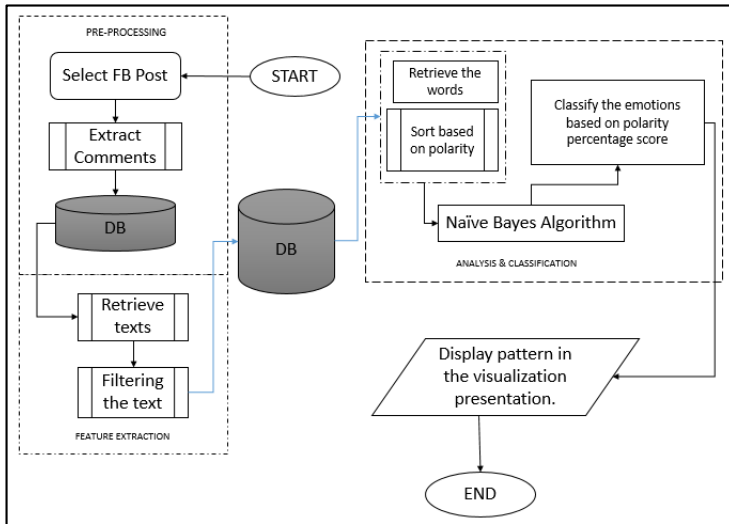
Figure 1. System framework

3.1. Gathering requirement

In this phases called as pre-processing. It is having been implemented is a manual process. It is observing, identify and selecting relevant Facebook post before the comments extracted automatically. Three data sets of emotions libraries is set up which is Positive (Happy), Negative (Unhappy) and Neutral (Emotionless).

3.2. Features extraction

In this study, Malaysiakini Facebook page was selected as data training. Before posted comments on an extract from this page, the system will collect and hold the page ID. Fig 2, shows the conversion process of Facebook Page Name to Page ID. Once the Page ID has been set, then identify which post comment to be extracted for classifying people's emotion.

Figure 2. Converting facebook name to FB ID
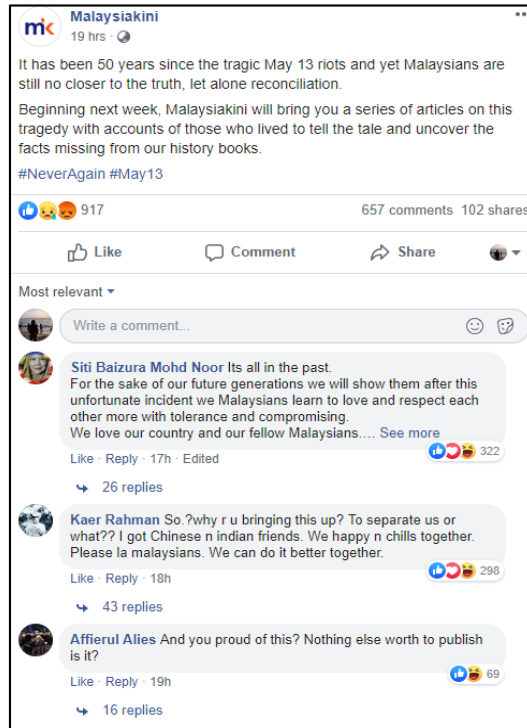
46

Figure 3. Posted comment with people's responses

Fig 3 shows Facebook user's response to posted issues. The posted issue was selected is about the events of May 13. All the comments will be extracted automatically using software tools developed by using python programming language.

3.3. Data analysis

Once the data is extracted and stored in a database known as a corpus. The data will be sorted and classified according to the predefined indicators. This classification process is performed using the Naïve Bayes algorithm.

The calculations are done after retrieving comments from Facebook's post. All the users have their own unique ID. The separating, arranging and ordering of writings are done dependent on chose client ID. The following equations are used for retrieved and sorting the comments of all the users commented on a chosen Facebook post.

```
P(S|C) = P(S). P(C|S) / P(C),
For assessment polarity:
    P(S|C) ~ P(C|S)
S is Sentiment
C is Comments
P is Polarity
```

The algorithm will automatically set percentages score for each indicator. If the percentages of happy sentiment are higher than unhappy, then it can be concluded as positive tagged with the issue vice versa. If the both happy and unhappy sentiment percentage is equal then, it can be tagged as neutral which is emotionless. Later, after going through the sorting process, each sentiment classified on the tagged emotion category. Below is an algorithm used for emotional classification purposes.

1.Start

2. DECLARE Sentiment Polarity per user

3. P is defined as a PolaritySentiment of < id, C, F(x), Qi >, where

– id is the user id.

– C is the comment (e.g., future generation) for which sentiment analysis is to be performed.

– F(x) is the training set of numerous functions (e.g., sentiment correlation)

– Qi is a set cleaning and removal word< q1, q2, ..., qn >, where each qi

4. signifies a function property for correlation.

  If Qi > Qii,

    Then SET as POSITIVE Polarity.

  Else If Qii < Qi,

   Then SET as Negative Polarity.

  Else

    SET as Neutral Polarity.

5. Stop

3.4 Pattern

In this stage, the final result will be a pattern in visualization data. From the resulting pattern can be concluded is it the posted issues is effected or not the user's satisfaction. Additionally, the particular person can also start thinking about the idea of solutions on how to resolve the issues raised by users, so that all users are satisfied.

4. Results and discussion

The implementation of the system is developed by using Python programming language and web page to extracted, filtered, classifying, and produce patterns of emotions. Then, opinions obtained from users are extracted to be analysed to determine the category of the sentiments by using Naive Bayes Algorithm. Each word of sentiment that has been analysed will be stored in the database to generate the keywords used. These keywords are used to categorize each sentiment by whipping his polarity. Finally, once the keywords have been classified, then the system will automatically generate a statistical diagram to formulate the findings.

The comments will be naturally separated to discover the extremity in a recognizing comment in the event that it is positive, negative or neutral and it will create a visual portrayal of result through bar graph to demonstrate the extremity rate. The framework will just acknowledge either English or Malay words.

Table 1 shows the result of the user's emotion based on keywords. The keywords have been generated from the corpus. From the table, it can be seen that the keyword's frequency is "None" which is a keyword of the Neutral category. None keyword is meant, the directed comment is factual or thankful. While the percentages for the keyword frequency used are shown in Fig 4.

Table 1. Emotion polarity based on keywords used

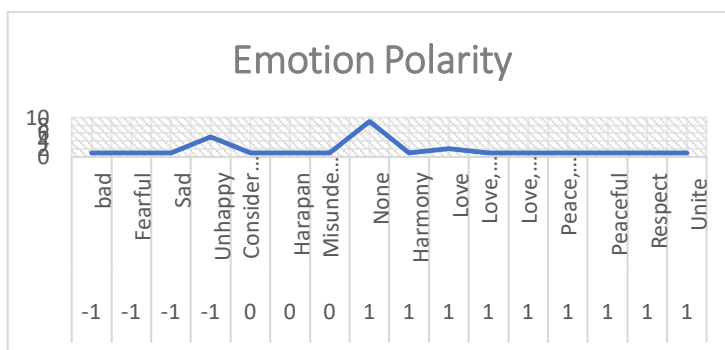| Row Labels	Sum of Count of Emotion Polarity
Negative	8
bad	1
Fearful	1
Sad	1
Unhappy	5
Neutral	3
Considerate	1
Harapan	1
Misunderstanding	1
Positive	18
Harmony	1
Love	2
Love, Peaceful, Harmony	1
Love, Respect	1
None	9
Peace, unity	1
Peaceful	1
Respect	1
Unite	1
Grand Total	29

Figure 4. Emotion polarity

The result shown in Table 2 is that neutral has the highest percentages which are about 40% since the issue is facts. In certainty, customers use cheerful words such as feeling the words.

Table 2. Result of emotion polarity based on tagged

Emotion Polarity	Sum of Emotion Polarity
Negative	8
Neutral	12
Positive	9

By comparing the percentage of polarity, the people's satisfaction with the issue can be determined as shown in the Fig 5.
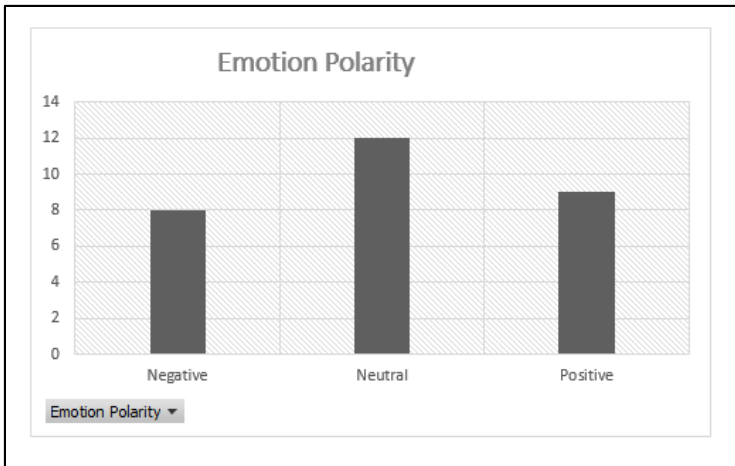

Figure 5. Result of people's emotions based on polarity tagged

5. Conclusion

In this paper, sentiment analysis on users' emotion towards a discussion issue is presented. A computerized tool is implemented to extract words in a Facebook post and comments itself can determine the overall feedback of people. The result is beneficial in many cases, especially in education and business.

References

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science, 152, 341–348. https://doi.org/10.1016/j.procs.2019.05.008

Barrón Estrada, M. L., Zatarain Cabada, R., Oramas Bustillos, R., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2020.113265

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2020). Sentiment Analysis on Online Product Reviews. India: Springer Singapore. https://doi.org/10.1007/978-981-13-7166-0

Chandra, Y., & Jana, A. (2020). Sentiment analysis using machine learning and deep learning. In Proceedings of the 7th International Conference on Computing for SustainableGlobalDevelopment,https://doi.org/10.23919/INDIACom49435.2020.9083703

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. ACM Computing Surveys. https://doi.org/10.1145/2938640

Lennox, R. J., Veríssimo, D., Twardek, W. M., Davis, C. R., & Jarić, I. (2020). Sentiment analysis as a measure of conservation culture in scientific literature. Conservation Biology. https://doi.org/10.1111/cobi.13404

Liu, N., & Shen, B. (2020). Aspect-based sentiment analysis with gated alternate neural network.KnowledgeBasedSystems.https://doi.org/10.1016/j.knosys.2019.105010

Ma, D., Li, S., & Wang, H. (2020). Joint learning for targeted sentiment analysis. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018. https://doi.org/10.18653/v1/d18-1504

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review. https://doi.org/10.1016/j.cosrev.2017.10.002

Palit, S., & Ghosh, S. (2020). Real Time Sentiment Analysis. International Journal of Synthetic Emotions. https://doi.org/10.4018/ijse.2020010103

Wu, P., Li, X., Shen, S., & He, D. (2020). Social media opinion summarization using emotion cognition and convolutional neural networks. International Journal of Information Management. https://doi.org/10.1016/j.ijinfomgt.2019.07.004

Xu, G., Yang, X., Cai, Y., Ruan, Z., Wang, T., & Liao, X. (2020). Extracting polarity shifting patterns from any corpus based on natural annotation. ACM Transactions on Asian and Low-Resource Language Information Processing. https://doi.org/10.1145/3345518

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. https://doi.org/10.1002/widm.1253