International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 27th May 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-05

Title: **USING HASHTAGS TO CAPTURE FINE EMOTION CATEGORIES FROM TWEETS**

Volume 08, Issue 05, Pages: 288–296.

Paper Authors

G.NAVYASREE, NAGAVALI SAKA , DR. V. SURYA NARAYANA

Ramachandra college of Engineering, Eluru, Affiliated to JNTU Kakinada.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

USING HASHTAGS TO CAPTURE FINE EMOTION CATEGORIES FROM TWEETS

G.NAVYASREE[1], NAGAVALI SAKA [2], DR. V. SURYA NARAYANA [3]

[1]Pursing M.Tech Department of Computer Science and Engineering, Ramachandra College of Engineering, Eluru Affiliated to JNTU Kakinada.

[2]Assoc Professor of Computer Science and Engineering, Ramachandra College of Engineering, Eluru.

[3]Head of the Department in Computer Science and Engineering, Rama Chandra College of Engineering, Eluru

[1]navya.vwd@gmail.com,[2]vali214@gmail.com,[3]s_vadhri@yahoo.co.in

Abstract— Despite recent successes of deep learning in many fieldsof natural language processing, previous studies of emotionrecognition on Twitter mainly focused on the use of lexicons andsimple classifiers on bag-of-words models. The central question ofour study is whether we can improve their performance usingdeep learning. To this end, we exploit hashtags to create threelarge emotion-labelled data sets corresponding to differentclassifications of emotions. We then compare the performance ofseveral word and character-based recurrent and convolutionalneural networks with the performance on bag-of-words and latentsemantic indexing models. We also investigate the transferabilityof the final hidden state representations between differentclassifications of emotions, and whether it is possible to build aunison model for predicting all of them using a sharedrepresentation. We show that recurrent neural networks,especially character-based ones, can improve over bag-of-wordsand latent semantic indexing models. Although the transfercapabilities of these models are poor, the newly proposed trainingheuristic produces a unison model with performance comparableto that of the three single models.

Index Terms—Emotion Recognition, Text Mining, Twitter,Indonesian tweet, Natural language processing.

1. INTRODUCTION

Social media has become a new trend for peopleto interact and communicate. Hence, the growth rate of socialmedia users is increasing rapidly over the years. A social mediawhich has the highest user growth is Twitter. The content of theTwitter post, which is called as tweet, has been widely used byresearchers, government or industry to gain knowledge whichhelps them to solve everyday problems. Various actual human behaviors can be captured from tweets. One of the mostpopular tasks is emotion analysis. Emotion is an ongoing state of mind, characterized by mental, physical, and behavioral symptoms. People emotion can be identified directly through their facialexpression and speech. Automatically detecting emotion iscrucial because it can be implemented in various fields. Ineducation, emotion analysis can be utilized for intelligent learning environment. Moreover, emotion analysis can be usedin the business for identifying customer complaint in email. Innowadays world where the technology has grown rapidly,people also tend to express their emotion through text in asocial Medias post. In social media data such as Twitter,emotion detection can be beneficial in government to monitorpublic response regarding policy or political event. Moreover,emotion analysis from social media also can be utilized bycompanies to monitor public responses about services or product thus help them in deciding the target market.Identifying emotion in Twitter is also challenging because itsshort text with informal words and unstructured grammarcannot be handled using normal text processing techniques.Because of its importance, several datasets are created as abenchmark to obtain state-of-the-art techniques for emotionanalysis. Those standard datasets mostly used for Englishemotion task. However, the standard dataset for anotherlanguage is limited.

Indonesian tweet is potential for emotionanalysis study. According to Statist, an online statistics portal,Indonesia is marked as the third largest active Twitter users inthe Asia Pacific from 2012 to 2018. It can be inferred thatconducting emotion analysis for Indonesian tweet would bebeneficial for many purposes. However, there is not any publicdataset for emotion analysis in Indonesia. Previous works inIndonesian emotion analysis, not publish their dataset for thepublic. In addition, their datasets are limited in small data dandles variety. Therefore, we construct an Indonesian Twitterdataset for emotion classification task which has variouscharacteristics and available for public.

2. RELATED WORK

In addition, we also propose feature engineeringto discover the best features for Indonesian emotionclassification. Those features include Bag-of-Words, word embeddings, lexicon-based, Part-Of-Speech (POS) tag, andorthographic features. For classifier, there are three methodsused: Logistic Regression, Support Vector Machine, andRandom Forest. F1-score is utilized as a metric to evaluate thebest performance of feature and classifier. To sum up, our maincontributions are:

• We build a dataset for Indonesian emotionclassification from Twitter data. This dataset consistsof 4.403 tweets which divided into five classes ofemotions (love, joy, anger, sadness, fear) and publiclyavailable for research purpose2.

• We propose feature engineering which recommendsthe best features to identify emotion in Indonesiantweet.

The earliest study in emotion mining in text was conducted by Alma et al. They identified emotion expresses inchildren fairy tales using Valence and Arousal model. Thedataset built in their research has been widely used in emotionanalysis study. On the other hand, the initial study

https://www.statista.com/statistics/303861/twitter-users-

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

asiapacificcountry/https://github.com/meisaputri21/ Indonesian-Twitter-Emotion-Dataset of emotion analysis on Twitter data was introduced by Mohammad. Theyused n-gram and emotion lexicon based features for detectingthe emotion in English tweet based on Ekman's emotion model.Since then, the study of emotion analysis using tweet is increased, both using supervised and unsupervised methods.Most emotion analysis studies utilize emotionlexicon for classification features. There are several emotionlexicons for English which have been widely used for emotion classification, such as NRC emotion lexicon and Word NetAffect (WNA) lexicon which construct based on Ekman'semotion class. However, there is only one emotion lexicon forIndonesian which was developed by Shaver based on Shaver'semotion definition. Therefore, the study of emotion analysis inIndonesia mostly uses n-gram based feature instead of lexicon based. Early research on Indonesian emotion analysis on tweetdata was conducted by Arifin ET. Al. They use Non-negativeMatrix Factorization, an extension of TF-IDF model, to classifyemotion in tweets. TF-IDF based features also used by toclassify emotion in Indonesian tweet. On the other hand, The etal. used more various features for detecting emotion inIndonesian tweet, including n-gram, linguistic, sentimentlexicon, and orthographic features. They used Shaver'semotion word list as query filters in data collection thus their dataset consists of explicit emotion only. However, allexperiments in Indonesian emotion analysis are conductedusing their own dataset because there is no standard dataset forIndonesian emotion classification which publicly available.

In recent years, word embed dings dominantlyused as a feature for emotion classification. Word embeddingfeatures for English emotion detection has been implementedby Heirs et al. They compared the use of basic Bag-of-words(BOW) features and word embedding's (Word2Vec andGlove). The results of their experiment show that combiningbasic BOW features and word embed dings can improve theperformance. Word embed dings for tweet emotionclassification also used by Vora et al. Using Random Forest,their model can achieve 91% precision for four classes ofemotion in English tweet. However, word embed dings havenot been yet utilized for Indonesia emotion classification task.3.

METHODOLOGY

There are two main processes conducted in this study: Emotion classification and Dataset building.

3.1 EMOTION CLASSIFICATION

Paul Ekman studied facial expressions to define a setof six universally recognizable basic emotions: anger, disgust, fear, joy, sadness and surprise Robert Plutchik defined a wheel-like diagram with aset of eight basic, pairwise contrasting emotions; joy – sadness, trust – disgust, fear – anger and surprise – anticipation.We consider each of these emotions as a separatecategory, and we disregard different levels of intensities that Plutchik defines in his wheel of emotions. Profile of MoodStates [6] is a psychological instrument for assessing theindividual's mood state. It defines 65 adjectives that are ratedby the subject on the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

five-point scale. Each adjectivecontributes to one of the six categories. For example, feelingannoyed will positively contribute to the anger category. Thehigher the score for the adjective, the more it contributes to theoverall score for its category, except for relaxed and efficientwhose contributions to their respective categories are negative.POMS combines these ratings into a six-dimensional moodstate representation consisting of categories: anger, depression, fatigue, vigour, tension and confusion. Since POMS is notpublicly available, we used the structure from Norcross et al,which is known to closely match POMS's categories. Wesupplemented it with additional information from the Brian MacSports Coach website1 Comparing to the original structure, wediscarded the adjective blue, since it only rarely corresponds toan emotion and not a colour, and word-sense disambiguation.

1.https://www.brianmac.co.uk/pomscoring.htm tools were unsuccessful at distinguishing between the two meanings. Wealso removed adjectives relaxed and efficient, which havenegative contributions, since the tweets containing them wouldrepresent counter-examples for their corresponding category.For each category we used the following adjectives:

• **Anger**: angry, peeved, grouchy, spiteful, annoyed,resentful, bitter, ready to fight, deceived, furious, badtempered, rebellious,

• **Depression**: sorry for things done, unworthy, guilty,worthless, desperate, hopeless, helpless, lonely, terrified,discouraged, miserable, gloomy, sad, unhappy,

• **Fatigue**: fatigued, exhausted, bushed, sluggish, worn out, weary, listless,

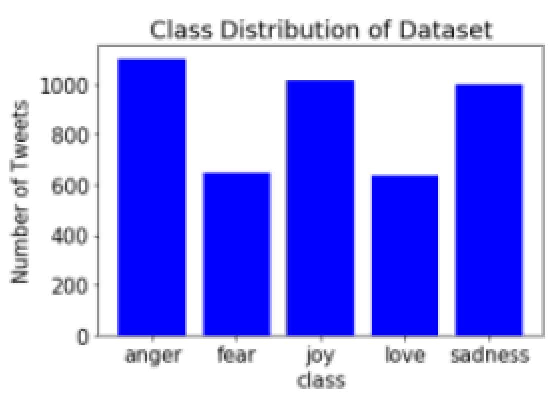
• **Vigour**: active, energetic, full of pep, lively, vigorous, cheerful, carefree, alert,

• **Tension**: tense, panicky, anxious, shaky, on edge, uneasy,restless, nervous,

• **Confusion**: forgetful, unable to concentrate, muddled, confused, bewildered, uncertain about things.From now on, we will refer to these classifications as Ekman, Plutchik and POMS.

3.2 DATASET CHARACTERISTICS

Our dataset is built based on manual annotation by twoannotators. There are 7.500 tweets that should be annotated byannotators. After annotation, the proportion of five basicemotion class, no-emotion, and multi-label emotion are 64%,32%, and 4% respectively.In this study, we consider to focus on five basicemotion classes. To measure the quality of annotation, wecalculate the Kappa score of five basic emotion classes. TheKappa score of our annotation is 0.917 which considered beingvery good. The final dataset is taken from the dataset with theagreed label, which consists of 4.403 tweets.The distribution of our dataset is summarized in Figure.Figure 1Class Distribution of Dataset



To shows that there is a balanced number of joy, anger, and sad class. On the hand, the number of love and feartweet are limited. To show the variety of our data, we put onthe example of tweets in anger class.First example:
Hariinilibur, rencananyamaunonton JurassicWorld, tapikayanyagajadidehmengingatkon-disiyggak fitbgtinisebel. Rusakrencanasebelanga..sebelakutu (Today is holiday, I am going to watch Jurassic World, but maybe itshould be canceled because I am extremely not fit. annoying.What a broke plan. I am annoyed.)
Second example:
The first example contains emotion word, i.e.annoying, hence anger emotion can be indicated explicitly. Onthe other hand, the second example does not contain anyemotion words, but we can identify this tweet as anger becauseof capitalized characters and exclamation mark. This kind ofimplicit emotion can be captured in our dataset because we donot use emotion words list on the data collection process. Thischaracteristic is different from another Indonesian tweet datasetwhich commonly contains explicit emotion only.

3.3 BAG-OF-WORDS & LATENT SEMANTIC INDEXINGMODELS

To set the baseline performance, we firstexperimented with common approaches to emotion detection.Within the realm of pure machine learning (as opposed tousing, say emotion lexicons), one of the most frequently usedapproaches is to use simple classifiers on the bag-of-words (BoW) models. We studied two approaches for transforming raw text into BoW model. Vanilla BoW is a model without anynormalization of tokens. Normalized BoW reduces the dimensionality of feature space by these transformations: all@mentions are truncated to a single token, all URLs aretruncated to a single token , all numbers are truncated to a singletoken , three or more same consecutive characters are truncatedto a single character (e.g. loooooove → love), TABLE 5 Thenumber of features of BoW and LSI models for combined trainand validation sets using different token normalizations. Thename bigrams stands for a model consisting of combination ofunigrams and bigrams. Ekman Plutchik POMS BoW LSI BoWLSI BoW Unigrams Vanilla 45,484 523 58,146 500 183,727Unigrams Norm. 35,555 316 44,009 299 129,841 BigramsVanilla 204,453 5,433 284,467 6,183 1,248,037 BigramsNorm. 187,533 3,955 256,889 4,390 1,081,598 all tokens arelower-cased. The aim of these normalization techniques is toremove the features that are too specific. For each of these twomodels, we run experiments on counts of unigrams as well asunigrams and bigrams. Hereafter, we will refer to thecombination of unigrams and bigrams simply as bigrams.Tokenization was done using Tweet POS tagger. For eachmodel, we filtered out tokens and bigrams occurring in less thanfive tweets. These four BoW models served as a basis for experiments with latent semantic indexing (LSI). Wedetermined the number of dimensions to keep so that 70% ofthe variance was retained. While the threshold comes from thenumber of retained dimensions is in the range that empiricalstudies show as

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

appropriate. LSI experiments were onlyperformed for Ekman and Plutchik, since calculating the decomposition for POMS was not possible with thecomputation resources we had at our disposal. The dimensionality of BoW and LSI models is shown in Table 5.

We experimented with the following classifiers: SupportVector Machines with linear kernel (SVM), Na¨ıve Bayes (NB), Logistic Regression (LogReg) and Random Forests (RF).Regularization parameters for SVM, LogReg, and the numberof trees for RF were selected using linear search.

4. EXPERIMENT AND RESULT

We implement our proposed features which have beendescribed in Section III to our built dataset. In addition, we alsoapplied our proposed features into Indonesian tweet dataset forcomparison. Their dataset consists of 942 tweet which hassimilar emotion classes but has different characteristics fromours. Their dataset has explicit emotion because it was buildbased on emotion words list. On the other hand, our dataset hasmore variety of data as mentioned in Section IV. We comparethe contribution of different features in different MachineLearning classifier for both datasets. We implement severalindividual features as mentioned in Section III as well as thecombination of those individual features. The results of ourexperiment are summarized in TA-BLE I. We examine the useof different individual features and the combined features. Theresults show that the use of emotion words list as our baselinefeature performs better on The's dataset which containsemotion words explicitly.

This feature achieves 57.85% on F1-score whenLogistic Regression applied. On the other hand, the highest F1-score for this baseline feature on our new built dataset is 43.09%. The use of emotion word list is not enough to capturethe emotion expressed in our dataset due to the variety of ourdata.

Other individual features are Bag-of-Words and word embeddings. The use of Bag-of-Words can boost perform- mansion both datasets. For word embeddings features, we compared he use of Word2Vec and Fast Text features. In general, Fast Text obtain better score both in two datasets althoughWord2Vec perform better on The's dataset when Logistic Regression applied. The great result obtained when wecombine the emotion word list, Bag-of-Words, and Fast Text features. For the lexicon-based feature, In Set sentiment lexicon, get the best F1-Score compared to Vania's lexicon andemoticon list. Vania's lexicon contains formal words while In Set lexicon is developed using Twitter data thus it moresuitable for our task. However, there is a slight difference ofF1-score obtained from emoticon lexicon feature in bothdatasets. Combining Vania's lexicon, In Set Lexicon and emotion list obtain slightly higher F1-score than the result of individual In Set. In addition, we examine the effect of combine emotion words list, Vania's Lexicon, In Set lexicon, and emoticon list for feature combination. The result shows that thecombination of these features achieve.Better performance compare to emotion word listonly. To boost the performance of emotion classificationmodel, we also examine the use of POS tag and or the graphic features. The results show that

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

both features not perform wellas individual feature, but shows better performances whencombined.

For increasing the F1-score, we considerimplementing several feature combination scenarios. We takethe best feature for each feature group and combine thosefeatures. Based on the results in TABLE I, it can be inferredthat the most significant features are formed based on thecombination of Emotion Words List, Bag-of-Words and Fast Text. This combination achieve 73.72% of F1-Score in The’s dataset and 68.39% in our new dataset. Adding lexiconand additional features (orthographic and POS tag) to thecombination of basic features can increase the F1-Score. Both The’s dataset and our new dataset achieve the highest F1-scorewhen the combination of basic (emotion word list, Bag-of-Words, Fast Text), Lex (Vania’s lexicon, In Set lexicon, emoticon list), orthography and POS tag features used in theLogistic Regression model. This combination achieves 75.98%of F1-Score on The’s dataset and 69.73% of F1-Score on ournew dataset. Regarding the classifier model, LogisticRegression performs the best in almost scenarios, followed bySupport Vector Machine and Random Forest. In general, ourpro-posed feature combination can boost performance in bothdatasets. The implementation of our proposed features to The’s dataset can achieve 75.98% F1-score which is better

TABLE II. EVALUATION OF EACH EMOTION CLASSON OUR NEW

5. DATASET

Class	Precision	Recall	F1-Score
love	64%	75%	69%
joy	81%	60%	69%
anger	61%	81%	70%
sadness	89%	72%	80%
fear	65%	53%	59%
avg/total	70%	68%	68%

Due to the variety and complexity of our new dataset, whichconsists of explicit and implicit emotion, the learning modelcannot perform better than the implementation in The’s data set, which contains explicit emotion only. We present the detailevaluation of each emotion class of our new built dataset inTABLE II. The best-combined features and LogisticRegression classifier are used in this evaluation. TABLE IIshows that a balanced score of precision and recall is achievedby sadness class. Sadness class obtains the best evaluation inprecision, i.e. 89%. It means that there is only 11% falsepositive for sadness label. Recall score for sadness class is also quite high, i.e 72%. On the other hand, joy class achieves highprecision but low recall. There is 40% of joy class is predictedas false negative. In contrast, anger class obtains low precisionbut high recall. The lowest score of precision and recall isobtained from fear class. The limited number of samples in fearclass impacts to its classification performance.

6. CONCLUSION

The central aim of the paper was to explore the use of deeplearning for emotion detection. We created three large collections of tweets labeled with Ekman’s, Plutchik’s and POMS’s classifications of emotions. Recurrent neural network sin deed

outperforms the baseline set by the commonbag-of-words models. Our experiments suggest that it is better to train RNNs on sequences of characters than on sequences ofwords. Beside more accurate results, such approach alsorequires no preprocessing or to kenization. We discovered thattransfer capabilities of our models were poor, which led us tothe development of single unison model able to predict all threeemotion classifications at once. We showed that when trainingsuch model, instead of simply alternating over the data sets it isbetter to sample training instances weighted by the progress oftraining. We proposed an alternative training strategy thatsample straining instances based on the difference between train and validation accuracy and showed that it improves overalternating strategy. We confirmed that it is possible to train asingle model for predicting all three emotion classificationswhose performance is comparable to the threeseparate models. As a first study working on predicting
POMS's categories, we believe they are as predictable asEkman's and Plutchik's. We also showed that searching for tweets containing POMS adjectives and later grouping themaccording to POMS factor structure yields a coherent data setwhose labels can be predicted with the same accuracy as otherclassifications.

7. REFERENCES

1. W. G. Parrott, Ed., Emotions in social psychology: Essentialreadings. New York, NY, US: Psychology Press, 2001.

2. N. Gupta, M. Gilbert, and G. Di Fabbrizio, "Emotion detection in email customer care," in Proceedings of the NAACL HLT 2010Workshop on Computational Approaches to Analysis andGeneration of Emotion in Text. Association for Computational Linguistics, 2010

3. J. E. The, A. F. Wicaksono, and M. Adriani, "A two stage emotion detection on Indonesian tweets," in 2015 International Conference onAdvanced Computer Science and Information Systems (ICACSIS), Oct 2015,

4. T. Daouas and H. Lejmi, "Emotions recognition in an intelligent learning environment," Interactive Learning Environments, vol. 0,no. 0, pp. 1–19, 2018.

5. Zainal Arifin, Y. Arum Sari, E. Kamilah Ratnasari, and S. Mutrofin,"Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization,"

G.Navya sree Pursing M.Tech Department of Computer Science and Engineering, Ramachandra College of Engineering, Eluru Affiliated to JNTU Kakinada

Nagavali Saka Assoc Professor of Computer Science and Engineering Ramachandra College of Engineering, Eluru.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IN

Dr. V. Surya Narayana Head of the Department in Computer Science and Engineering, Rama Chandra College of Engineering, Eluru