

Emotion Label from ANEW dataset for Searching Best Definition from WordNet

Tegar Satrio Utomo

Departement of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
mrsatriotegar@gmail.com

Riyanarto Sarno

Departement of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Suhariyanto

Faculty Of Computer Science
Dian Nuswantoro University
Semarang, Indonesia
Suhariyanto@dsn.dinus.ac.id

Abstract—Emotions in a sentence can hold many roles in the field of sentiment analysis, knowing emotions can also be known through facial expressions, speech and text. In this paper we try to find the best definition for WordNet, by utilizing ANEW and ISEAR datasets. Sentences from WordNet that match the parameters of the two datasets will be searched for emotion, labeled according to the class and selected. This can be used in text mining. The result of this research showed that not all definitions in wordnet can accurately match all datasets and contexts

Keywords—ANEW; Thayer's Model; WordNet; word sense disambiguation; POS tagging

I. INTRODUCTION

Determining emotion in a sentence is one of the studies that many fields apply it, call it text mining, sentiment analysis[14], mood analysis. But if we wrongly define the definition of a word, then the Sentiments and Emotions we seek can be wrong. This time the data we choose is ANEW data and ISEAR data.

Before discussing further, what is ANEW? ANEW [1] is a spatial model of various emotional words, ANEW is a dataset that has many values, one of which is the value of Valence Mean and Arousal Mean. These two values will later be used to find the emotional label of ANEW. Based on the emotional class in the Thayer's model [2] the ANEW word can be searched for the label or emotion class. Then using ISEAR data, will look for examples of sentences to be searched sin for that sentence.

Synset is a collection of synonyms that have the same context, this synset will be related to WordNet, an English lexical[13] database [3]. Sentences in this WordNet which will be seeking the definition of WordNet that matches the emotional label of the word ANEW

II. METHODOLOGY

A. Looking for an emotional label from ANEW data

The first step is to look for the emotional label of each word in the ANEW data. As previously explained that ANEW [1][12] has a Valence Mean value and an Arousal Mean value, these two values contain the datum value of how nice or not a

word (valence) and value how calm or happy a word (arousal) [1].

This value will be used to map where the word emotion class in ANEW data, using Thayer's model. [2]

Fig. 1. Thayer's model for emotion classification

Fig. 1 shows what is Thayer's model in general, there are 2 dimensions with 4 classes. Thayer's Model is one of the emotional grouping models commonly used to model emotions in music, Thayer's model [2] has 2 dimensions, in the x-axis and y-axis, for the x-axis the taken quantity is the valence meanwhile for the y-axis taken is the arousal mean.

The value for each axis is a scale of 0 to 9 [4]. ANEW word that will be searched for emotion label then taken valence mean and arousal mean that then mapped into each emotion class in Thayer's Model.

If valence mean < 5 and arousal mean ≥ 5 then belong to emotion class angry

If valence mean ≥ 5 and arousal mean ≥ 5 then belong to emotion class happy

If valence mean < 5 and arousal mean < 5 then belong to emotion class sadness

If valence mean ≥ 5 and arousal mean < 5 then belong to emotion class calm

B. Matching emotion classes on the Thayer's Model and emotional labels on ISEAR data

The next step is to match the ANEW emotion class and the emotional label from ISEAR. Before discussing further, what exactly ISEAR is? ISEAR [5] is a project run by world psychologists directed by Klaus R. Scherer and Harald Wallbott.

ISEAR collects data from Respondents consisting of psychologists, psychology students, and non-psychologists.

Respondents are asked to give examples of sentences that are real situations in which they have experienced an event that they think has emotion or is an experience consistent with 7 major emotional labels (joy, fear, anger, sadness, disgust, shame, and guilt). [5]

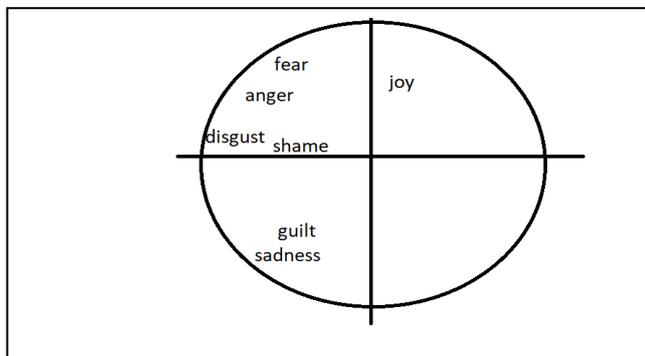


Fig. 2. ISEAR emotion labels

In ISEAR data, emotions have 7 labels, while in Thayer’s Model have 4 emotion classes [6], this difference can be uniform by finding a match between the emotion class in the Thayer’s Model as well as the emotional label on the ISEAR data. so that each label on ISEAR corresponds to the emotion classes on the Thayer’s Model.

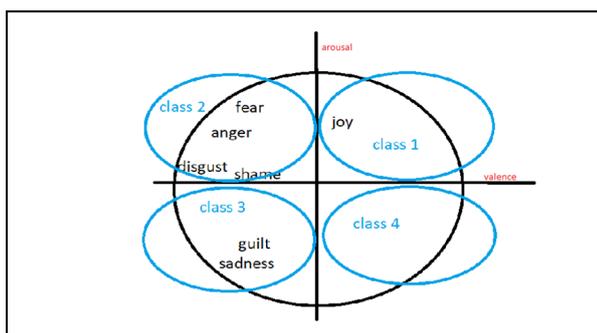


Fig. 3. Emotional class on the Thayer’s model that matches the ISEAR emotional label

After Emotional Label on ISEAR is matched on emotion class in Thayer’s Model, we get 3 emotion class on Thayer’s Model that match with emotional label in ISEAR data, that is class 1 containing joy emotion, class 2 containing anger emotion, fear, disgust, shame class 3 which contains guilt emotion, sadness. Class 4 is not included because there is no match with the sentence’s emotion label on the ISEAR data.

Fig. 2 and Fig. 3 shows us about the label and classes that used to process data. Fig. 2 shows the label from ISEAR dataset, there is 7 label in ISEAR that will be grouped to classes that Fig. 1 mention.

C. Looking for sentence examples on ISEAR

As previously known, ANEW data containing words will be sampled sentences in ISEAR Data, the sentence must contain the said ANEW data in question along with an emotion label that is in accordance with the emotion class of the Thayer’s Model which is already aligned with the ISEAR data emotion label.

To do string matching, one way that can be used is to use regular expression or regex [7]. By looking for the ANEW word emergence pattern in the sample sentence in the ISEAR data, the example sentence can be retrieved

	Description	Emotion	Sentence
6	guilty	anger	My brother accused me undeservedly before my p...
9	hope	joy	When I was in a situation which seemed hopeless...
17	hospital	joy	I was in hospital, and then was visited by som...
20	hammer	sadness	When I didn't get a certain job at Lillehammer...
27	hug	joy	When I get a hug from someone I love.

Fig. 4. Table of ANEW word and its sentences

In Fig.4 the “Description” column contains the word from the ANEW data that has passed the previous step (look for the emotion class). The “Emotion” column contains the ANEW’s classified emotion class, and the “Sentence” column contains a sample sentence from the ISEAR data containing the ANEW word and has the appropriate emotion class

C. POS Tagging

After getting a sample sentence from the ISEAR data for every word in ANEW that already has an emotion class, The next thing is to find the word class or POS (Part-of-Speech) tagging of our ANEW word in the sample sentence from the ISEAR data.

Using the nltk library on python. Nltk (Natural Language Toolkit) [8] is a library that provides various features commonly used to process data related to human language, with an interface that is easy to use. some of the features used are WordNet, tokenization, stemming, POS Tagging. And Lesk.

In POS tagging or searching for word classes, some examples of the English word class include:

TABLE I. POS TAGGING IN ENGLISH LANGUAGE

code	class
JJ	Adjective
VB	Verb, base form
NN	Noun
RB	Adverb

a) *Noun (N)*

If POS tags are starts with N, then it should be changed to 'n'.

b) *Verb (V)*

If POS tags are starts with V, then it should be changed to 'v'.

c) *Adjective (A)*

If POS tags are starts with J, then it should be changed to 'a'.

d) *Adverb (R)*

If POS tags are starts with R, then it should be changed to 'r'.

The goal from this step are to make a structure that called “Synset”. So our ANEW structure had a same structure with

Description	Emotion	Sentence	POS
guilty	anger	My brother accused me undeservedly before my p...	JJ
hospital	joy	I was in hospital, and then was visited by som...	NN
hug	joy	When I get a hug from someone I love.	NN
happy	joy	At a friend's birthday party with some of my c...	JJ
hard	joy	At a choral concert I appeared as a soloist. Th...	JJ

Wordnet in the form of “Synset” [3].

Fig. 5. POS Tagging

D. Define synset and Define WordNet Definition sentences

Once we get the ANEW emotion labels, get a sample sentence from the corresponding ISEAR data, get the word class from ANEW word in the sample sentence from the ISEAR data, then the next step is to look for synset. As has been explained, synset is a collection of synonyms of words that have the same meaning of the same. When synsets are formed then we can look for examples of sentences that have similar aspects in WordNet.

Synset has the following structure “guilty.a.01 [9]”

The above sentence can be interpreted, The word “guilty” in the sentence has a class adjective and has the closest meaning to the “guilty” word in index 1 in WordNet. To perform a Synset search, one way is to use a latch algorithm [10]. how does it work lesk? The way the lesk works is to think that the adjacent word is a similar word.

Lesk returns a synset value of the ambiguous word and the context in which it appears.

```
>>> from nltk.wsd import lesk
>>> sent = ['I', 'went', 'to', 'the', 'bank', 'to', 'deposit', 'money', '.']

>>> print(lesk(sent, 'bank'))
Synset('savings_bank.n.02')
```

Fig. 6. The Lesk Algorithm for searching Synset

In the words ANEW as ambiguous word and sentence from ISEAR as word exit context, as well as POS tagging from ANEW word, Lesk returns Synset with the highest number of word overlapping between different benchmarks and definitions of each Synset.

So one word from WordNet will be chosen which has aspects and uses that are similar to the words from ANEW.

III. EXPERIMENT

To prove the correctness of the chosen definition, the ISEAR sentence example is searched for by measuring how much the number of words ANEW in the sentence [11]

$$v_{text} = \frac{\sum_{i=1}^n v_i f_i}{\sum_{i=1}^n f_i} \quad (1)$$

where :

v_i = Average valence for word i

f_i = frequency for ANEW word

To estimate the overall valence score for a text, which we denote by V_{text} , we

(a) determine the frequency f_i that the i th word from the ANEW study word list appears in the text; and

(b) compute a weighted average of the valence of the ANEW study words as shown in equation (1).

After emerging Emotion results from the ISEAR sentence, then we calculate how accurate we calculated.

To calculate the accuracy of the confusion matrix, The accuracy is simply the sum of the diagonal elements divided by the sum of all elements in the matrix.

TABLE II. THE EXPERIMENT RESULTS

Predicted / Actual	anger	joy	Sadness	other	NaN
Anger	25	8	6	0	8
Joy	1	50	1	2	24
Sadness	4	4	9	1	5

After the calculation process from 148 data found, there are 84 data that match the label,

then we use the following formula to find accuracy

$$\text{Accuracy} = \frac{tp + tn}{tp + fn + fp + tn} \quad (2)$$

where :

tp = true positive

tn = true negative

fp = false positive

fn = false negative

So the accuracy value is 0.56, with an accuracy value of 0.56 there are still many developments that can be done to increase the value of accuracy. However, the definition of WordNet that we take is right for every word ANEW.

IV. CONCLUSION

By utilizing the Synset structure formed by the two datasets that we have several method that we used, we can get WordNet definition that fits the context we want and measure how accurately we determine it.

ACKNOWLEDGEMENT

Authors would like to thank Institut Teknologi Sepuluh Nopember and the Ministry of Research, Technology, and Higher Education of Indonesia for supporting the research.

V. REFERENCES

[1] M. M. Bradley and P. J. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings NIMH Center for the Study of Emotion and Attention Affective Norms for English Words (ANEW) 1999."

[2] J. Grekow and Z. W. Raś, "Detecting Emotions in Classical Music from MIDI Files," Springer, Berlin, Heidelberg, 2009, pp. 261–270.

[3] A. Javed and B. S. Lee, "Hybrid semantic clustering of hashtags," *Online Soc. Networks Media*, vol. 5, pp. 23–36, Mar. 2018.

[4] R. E. Thayer, *The biopsychology of mood and arousal*. Oxford University Press, 1989.

[5] H. G. Wallbott and K. R. Scherer, "How universal and specific is emotional experience? Evidence from 27 countries on five continents," *Soc. Sci. Inf.*, vol. 25, no. 4, pp. 763–795, Dec. 1986.

[6] F. H. Rachman, R. Sarno, and C. Faticah, "Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 3, pp. 1720-1730., Jun. 2018.

[7] C. Chapman, P. Wang, and K. T. Stolee, "Exploring regular expression comprehension," in *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2017, pp. 405–416.

[8] S. Bird, E. Klein, E. Loper, *Natural Language Processing with Python*, (2009)

[9] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech*

[10] B. S. Rintyarna and R. Sarno, "Adapted weighted graph for Word Sense Disambiguation," in *2016 4th International Conference on Information and Communication Technology (ICoICT)*, 2016, pp. 1–5.

[11] P. S. Dodds and C. M. Danforth, "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents," *J. Happiness Stud.*, vol. 11, no. 4, pp. 441–456, Aug. 2010.

[12] F. H. Rachman, R. Sarno, and C. Faticah, "CBE: Corpus-based of emotion for emotion detection in text document," in *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2016, pp. 331–335.

[13] E. W. Pamungkas, R. Sarno, and A. Munif, "B-BabelNet: Business-Specific Lexical Database for Improving Semantic Analysis of Business Process Models," *TELKOMNIKA (Telecommunication Comput. Electron. Control.)*, vol. 15, no. 1, p. 407, Mar. 2017.

[14] B. S. Rintyarna, R. Sarno, and C. Faticah, "Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context," *International Journal of Information and Decision Sciences*. 2018 Vol. 11.