EMOTION RECOGNITION IN TAGLISHUANO SENTENCES

Article · March 2016

1 author:

Larmie Feliscuzo
Cebu Institute of Technology

EMOTION RECOGNITION IN TAGLISHUANO SENTENCES

Larmie T. Santos Feliscuzo
Chair, IT Department
CIT-University
larmie.feliscuzo@gmail.com

James Vincent T. Suson
jvsuson17@gmail.com

ABSTRACT

This program is concerned with recognizing emotions in a TAGLISHUANO sentence, since a sentence can contain more information from the emotion that they are giving. Emotions in sentences can convey more information than what is presented on the text. In this study, we used a different approach of recognizing emotions using emotion values tagged to our emotion words. The algorithm learns the most possible emotion that is conveyed in the sentence basing on the emotion values present. In this study, most emotion words are not only used for one specific emotion but it can also be used to convey other emotions. The highest accuracy achieved was 81.25% from the Happy emotion

General Terms

Algorithm

Keywords

TAGLISHUANO, Emotion Values

1. INTRODUCTION

In the Internet Age, communication between people has become increasingly focused on text. With the existence of mobile phones and the convenience of text messaging, people have become more and more dependent to this form of communication. In addition, the existence of microblogging, a type of blog that lets users publish short text updates (webopedia); communication has become fast, short and most of the time reliant on textual data. Although these methods of communicating are making life convenient, the conveying of emotion is just not the same as with speech, face and body languages. There has been a lot of study about recognizing emotion through facial expressions and sound from speech in the field of computer science, but not so much about recognizing emotion in text.

With the vast amount of study regarding emotion in the fields of psychology, social science, linguistics and communication, it is clear that this has fascinated researchers for years. Affective science or the study of emotions and affect can be traced back to the 1980s and since then has contributed many studies (Barrett & Gross, 2013). Some of these studies were about theories regarding the how and why we experience emotions (Micallef-Trigona, 2014), how emotions are expressed on the face, voice, body or language (Ekman & Keltner, Introduction: Expression of Emotion, 2003), and some others to name a few.

In a study, it has been observed that human's perceive emotions from facial expressions, speech and text, with text being the least way of perceiving emotions at 7% (Hendy & Farag). Through this, it can be said that humans rarely recognize emotions from text and because text has become a major form of communication today, it is easy for misunderstandings and misinterpretations between two individuals. Recognizing the emotions other people have is important in establishing interpersonal relationship with others (Medialab, 2014). Through recognizing the emotion given to us by someone, we would also be able to gain more information because emotions carry with them additional information.

In the past decade, the integration of computer science to the study of emotions was started with Rosalind Picard's paper on affective computing (Picard, 1995). Computer scientists have been studying and developing systems that can recognize, interpret, process, and simulate human affects. In this paper, the recognition of human emotions in textual data is the focus.

2. Theoretical Background

Emotion Dictionary

An emotion dictionary contains a list of words with emotional content. This dictionary contains six different values for each individual emotion category ranging from 0 to 100 with 0 being the absence of emotion in the word and 100 being the complete presence of the emotion in the word.

Emotion Category

Are the six emotion categories described my Paul Ekman which are, *happiness, sadness, anger, fear, disgust* and *surprise.*

Emotion Values

Emotion Values are values that determine the possibility of a word to fall under an emotion category.

Annotating and Tagging

Any metadata tag used to mark up elements of the dataset is called an *annotation* over the input. However, in order for the algorithms to learn efficiently and effectively, the annotation done on the data must be accurate, and relevant to the task the machine is being asked to perform. For this reason, the discipline of language annotation is a critical link in developing intelligent human language technologies. (Pustejovsky & Stubbs, 2015)

Natural Language Processing

Natural language processing (NLP) is the ability of a computer program to understand, interpret and analyze human speech and text. (Rouse, 2011)

3. Methodology

Preprocessing

This part changes all letters to lowercase so there won't be any conflicts when it will be cross-referenced with the emotion dictionary. This was adopted from the methodology of Bucag(2015). This part of the program also removes any contraction in a word to prevent the program from generating errors.

The words from the sentences will also be processed since we are getting our data set from microblogging sites like Twitter which users don't have a proper convention for spelling words. For example, a user can have the words *"yaaaaaaaaayyyyy", "hahahahaha"* or *"ewwwwww"*. These words are rich in emotional content however because they don't have a spelling convention, we need to process these words to be able to successful tag them with emotional values from the emotion dictionary.

Word Tagging

Each word from the sentence will then be cross-referenced to an emotion dictionary and then will be tagged with emotion values. An emotion value is a value between 0 and 100 and it indicates the possibility of a word from being any of the six emotion categories.

This tagging rule was adopted because sometimes an emotion word can mean more than one emotion but some more than others. In previous studies, researches have tagged one emotion for each word. For example, the word hate is commonly used to convey the emotion of anger. However, the sentence, "*I hate that I like this product from Best Buy!"* uses the word hate and it conveys not the emotion of anger but possibly the emotion of surprise or happiness. In addition, the word like can also denote other emotions aside from the feeling of happiness. Each emotion word in the emotion dictionary will contain six emotion values that are between 0 and 100, with 0 being the emotion category absent from the word and 100 being the emotion category perfectly found in the word.

Emotion Recognition Algorithm

This step determines the emotion of the sentence by using the six emotion values of the words that were tagged. This will uses an algorithm that averages all the emotion values present in the sentence and determines the emotion based on which emotion is dominant in the sentence.

Updating Emotion Value

If the emotion determined by averaging the values is wrong, the algorithm then updates the right emotion and updates the new emotion value to the emotion dictionary.

4. Experimentation and Results

Data Gathering

Using an R script and the Twitter API, 1,792 random tweets were gathered to be used as a data for training the emotion values. These tweets contain words like *happy, lipay, masaya, angry, suko,* From the 1,792 data for training, 244 tweets were also randomly chosen to be used as the test data.

Data Analysis

We initially tested our test data to the system before the system was trained. The test data was manually annotated with emotions basing on which dominant emotion was present in the sentence. Then it was loaded into the system and the results are recorded in Table 4.1 The accuracy of each emotion was measure by the number of correctly identified sentences over the number of sentences in the test data.

Emotion	Manually Annotated	Correctly Identified	Accuracy
Happy	64	39	60.9375%
Sad	37	10	27.0270%
Anger	43	0	0.00%
Surprise	22	4	18.1818%
Disgust	36	6	16.6667%
Fear	11	5	45.4545%
No emotion	31	5	16.1290%

Based on the results, only the emotion *Happy* has exceeded the 50% mark and the emotion *Anger* wasn't recognized by the system. After we have updated the emotion values using the 1,792 sentences used for training the data set. We once again tested our test data to the system. The results are displayed in Table 4.2. Again, we used the same formula for determining the accuracy of each emotion class.

Emotion	Manually Annotated	Correctly Identified	Accuracy
Happy	64	52	81.25%
Sad	37	16	43.4324%
Anger	43	20	46.5116%
Surprise	22	15	68.1818%
Disgust	36	23	63.8889%
Fear	11	8	72.7273%
No emotion	31	2	6.4516%

We can clearly see a significant increase in the accuracy of correctly identified sentences, however this was only made possible because our training sets and data sets have similar occurring emotion words, that is why the system was able to easily identify the emotion of the sentences.

The word processing stage of the system also helped easily identify similar occurring words like *hahahahaha, huhuhuhuhu, eewwwww, uggghhh, wooow, etc.* and these words are rich in emotion values.

5. CONCLUSIONS

Sentences from first person perspective are rich in information, these information is not only from the sentences themselves but also from the emotion they are conveying. The system developed utilizes different emotion values for words with emotional content and determines the most likely emotion the sentence contains by averaging these emotion values. The system also learns and changes the emotion value of the words by choosing whenever the system does not correctly identify the emotion.

In this study, most emotion words are not only used for one specific emotion but it can also be used to convey other emotions. The highest accuracy achieved was 81.25% from the Happy emotion, it was observed that this was because of commonly occurring words like *hahahaha, lipay, like, masaya, etc.* Word processing also helped increase the accuracy because the system was able to recognize emotion words with different and unconventional spellings. The system also has a hard time identifying sentences with No Emotion having only achieved 6.4516% accuracy even after training. This is because emotion words that are used as nouns and therefore should not contain any emotion value are being identified and the system has no way of disregarding the emotion value of these words.

6. RECOMMENDATION

For future works, it is recommended to increase the training set with sentences with various emotion words to make the system learn and adjust more weights of the emotion words. It is also recommended to increase the emotion and smiley dictionary and to consider other possible spellings for these words because it has been observed that the processing of the words have helped increase the recognition of the emotion value of the sentence. Other algorithms that can be applied to the emotion values should also be considered and determine their advantages and disadvantages in recognition. Lastly, we need to increase the capability of identifying sentences with No Emotion value, solutions for these would be to identify emotion words used as nouns and modifying emotion words that are affected by other words.

7. REFERENCES

(n.d.). Retrieved from http://www.webopedia.com/TERM/M/microblog.html

Aman, S. (2007). *Recognizing Emotions in Text.*

Barrett, L., & Gross, J. (2013, October). *The Emerging Field of Affective Science*. Retrieved from Association for Psychological Science: http://www.psychologicalscience.org/index.php/publications/observer/2013/october-13/the-emerging-field-of-affective-science.html

Bucag, M. S. (2015). *Emotion Analysis and Recognition of TAGLISHUANO Online Customer Reviews from Twitter.*

Burget, R., Karasek, J., & Smekal, Z. (2011). *Recognition of emotions in Czech Newspaper Headlines.*

Ekman, P. (1999). Basic Emotions. In T. Dalgleish, & M. Power, *Handbook of Cognition and Emotion.*

Ekman, P., & Keltner, D. (2003). Introduction: Expression of Emotion. In R. Davidson, G. HH, & S. KR, *Handbook of Affective Sciences.* Oxford University Press.

Hendy, N., & Farag, H. (n.d.). Emotion Recognition Using Neural Network: A Comparative Study.

Houen, S. (2011). *Opinion Mining with Semantic Analysis.*

Lee, S., Dai, D., Li, S., & Ahrens, K. (2011). Extracting Pseudo-Labeled Samples for Sentiment Classification Using Emotion Keywords. *Asian Language Processing, 2011 Internation Conference on*.

Medialab, S. (2014, November 10). *Recognizing emotions, and what happens when this is interrupted*. Retrieved from ScienceDaily: http://www.sciencedaily.com/releases/2014/11/141110083515.htm

Micallef-Trigona, B. (2014, October 22). *Comparing the 5 theories of emotion*. Retrieved from BrainBlogger: http://brainblogger.com/2014/10/22/comparing-the-5-theories-of-emotion/

Mihalcea, R., & Liu, H. (2006). A corpus-based approach to finiding happines. *AAAI Spring Symposium on Computational Approaches to Weblogs*.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*.

Picard, R. (1995). Affective Computing.

Pustejovsky, J., & Stubbs, A. (2015). *Natural Language Annotation for Machine Learning.*

Rouse, M. (2011, May). *Natural Language Processing (NLP) definition*. Retrieved from SearchContentManagement: http://searchcontentmanagement.techtarget.com/definition/natural-language-processing-NLP

Strapparava, C., & Mihalcea, R. (2008). *Learning to Identify Emotions in Text.*

Wen, S., & Wan, X. (2014). *Emotion Classification in Microblog Texts Using Class Sequential Rules.*