Emotional Mining: Tagging Emoticons to Online News

Vinothini Kasinathan1, Aida Mustapha2, Lee Zhi Yong1, Z.A. Aida Zamnah1
1Faculty of Computing, Engineering and Technology, Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia
2Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor

Corresponding author: vinothini@apu.edu.my

Abstract. This paper presents an emotion mining system, which assigns emoticons to newspaper articles into a pre-defined emotion category based on the underlying emotion in the news. Next, the system makes recommendation to the reader by tagging the news headline with the respective emoticons. Users are then able to decide whether to read the news based on the emoticons provided. The system also provides a filter for the users to choose the category of news to read following the emoticons.

1. Introduction
The study on emotions crosses many fields so as philosophy, education, and psychology. Its nature comprises of the roles of affective and cognitive process involving human learning and behavior [1]. Due to this nature, emotions are greatly affected by the use of language; whether spoken or written [2]. Despite the intuitive interest of emotions among the poets, musicians, philosopher, the study of emotions has also attracted researchers from the computing domain via the study of Natural Language Processing (NLP).

Natural language processing is a subfield of Artificial Intelligence [3] that studies the ability of computer programs to understand, translate, or respond to human speeches. The term “natural language” in this field does not refer to the actual language as used by human in everyday life, but more of a restricted subset of such a human language, one purged of constructions and ambiguities that computers could not sort out. NLP technologies are having a dramatic impact on the way people interact with computers, on the way people interact with each other through the use of language, and on the way people access the vast amount of linguistic data now in electronic form. Most of NLP will used for signal processing, context references issues and speech recognition, as well as syntactic and semantic analysis. Based on Figure 1 shows the framework of Natural Language Processing.

One emerging area that capitalizes on NLP technologies is mining on social network data due to the popularization of digital and social media. Emotions in social networks are interpreted as sense impression or sentiments, thus can be quantified based on their polarity such as happy vs. unhappy or positive vs. negative. This research focuses on assessing or quantifying the emotions underlying news in online portals, which consists of a lot of text conveying positive and negative
sentiments underlying a news. News articles have received much less attention, although bias across different news sources has been discussed by a few and some initial efforts have concentrated on sentiment analysis in the news area [4].

2. Related Work

Emotional mining or sentiment analysis have been primarily applied in domain other than news, such as in business for customer profiling. In understanding customer psyche, [5] proposed an Emotion Mining Sonargram as shown in Figure 2. The sonagram is designed to give perspective on the conscious and subconscious feelings towards a brand or concepts on customer emotions. Once the emotion results are plotted into the sonagram, the customer profiles can be mapped according to the actual needs among customers. Emotions, therefore, has been instrumental in establishing benchmark for brand positioning, product enhancement, and service delivery. In the news domain, emotional
mining can be used in recommending news to readers so they are able to filter out the good news or the bad news in the news portal based on their preference.

Reading bad news might give an impact to a person’s emotion because the informative content of textual data encloses attitudinal information and emotional states [6]. The main challenge in news mining is dealing with ambiguities of the news headlines, for example “The terrorist is happy because people died”. The system may classify the news as a good news based on the keyword “happy”, where in actual fact it is a sad news due to death of the innocent.

According to [7], emotions are short-lived, multi-componential reactions to the environment that serve important survival and interpersonal functions. Emotions unfold over time and depend on appraisal, reactivity, and regulatory mechanisms. Positive emotions represents human spirituality, where science, culture, and religious beliefs have evolved [8]. Two of the five commonly recognized personality traits in psychology are associated with the ability to experience emotion: extraversion with positive emotions and neuroticism with negative emotions [9]. It has also been shown that people react in different ways even to clear emotion expression devices, such as emoticons, in the sense of drawing inferences about the characters of the users [10].

Visual word recognition involve several processing stages in language processing [11]. During a core process of reading, there are different kind of information are extracted from the word stimulus and combines in order to achieve full comprehension. The emotional aspect may be processed during a reading process. Because reading online news have gained more popularity and ranks second after the television [12, 13], it is important to consider a text mining approach to news recommendation. Work such as [14] discussed the characterisation of emotions using keywords spotting, lexical affinity measures, statistical NLP and K-Mean calculation.

3. Proposed Emotional Mining System: EMticons
This paper proposes an emotion mining system called the EMticons, which assigns emoticons to newspaper articles into a pre-defined emotion category based on the underlying emotion in the news. The framework for EMticons is shown in Figure 3. EMticons works by extracting the news from online news portal based on their Hypertext Markup Language (HTML) tags such as title (<title>), description (<description>), and publication date (<pubDate>) into the form of XML. Next, the data is sent for natural language pre-processing, which are tokenization, lemmatization, and Part-of-Speech (POS)-tagging. The resulting POS-tagged sentences from the news headlines are assigned to happy and happy emoticons based on Sentiment Word Net (http://sentiwordnet.isti.cnr.it/).

![Figure 3. Framework of EMticons.](image-url)
3.1. News Extraction
The EMticons system gather the news data from online news web portals and store them in an XML format, separated by different XML tags. Figure 4 and Figure 5 show the actual news as input to EMticons and the extracted news in XML tags, respectively.

Given the news headline in XML tags, next the EMticons will perform a three-step pre-processing before the news are ready to be assigned with emoticons.

- Tokenization: Tokenize the news title, break the title into words and program system to read the title. Eg: “The phone is good” – [The] [phone] [is] [good]
- Lemmatization: Lemmatize the tokenized words. Eg: [The] [prices] [is] [cheapest] [than] [others] – [The] [price] [is] [cheap] [than] [other]
- Part-of-Speech Tagging: Do POS-Tagging on every lemmatized words. Eg: The phone is good – [The/DT] [phone/NN] [is/VBZ] [good/JJ]

Once pre-processing has been completed, the POS-tagged headlines are assessed for their weights based on the Sentiment Word Bank. Based in [15], there are a few formula used to calculate the weights for emotion weighting in the news title.

\[
\begin{align*}
\text{PC} &= \text{Positive value} \\
\text{NC} &= \text{Negative value} \\
\text{Polarity} &= \text{positive if (PC - NC} > \delta); \text{Polarity} = \text{negative if (NC - PC} < \delta); \\
\text{IF P total > N total} &= \text{Good} \quad \text{IF N total < P total} = \text{Bad} \quad \text{Else} = \text{Normal} \\
\text{P total} / \text{PC} &= \text{Total positive value} \quad \text{N total} / \text{NC} = \text{Total negative value} \\
\Delta &= \text{Disparity of total positive and total}
\end{align*}
\]
Figure 6 shows the excerpt from the SentiWordBank and Table 1 shows the positive or negative values from the words as matched with the sentiment words in SentiWordBank.

![Sample of SentiWordBank word bank](image1)

**Table 1.** Matrix of positive and negative values matched with SentiWordBank.

| Words                      | United | in | grief |
|----------------------------|--------|----|-------|
| Positive Value             | 0      | 0  | 0.000 |
| Negative Value             | 0      | 0  | 0.625 |

Next, Figure 7 shows the process of calculating the weights. Finally, the headlines are then assigned with a happy or unhappy emoticons to indicate the emotions underlying the news headlines.

![Calculating weights in EM_Icons](image2)

**Figure 7.** Calculating weights in EM_Icons.

### 3.2. User Interface

The user interface for the EM_Icons system was developed mainly for ease of use. Categories were prepared via a drop down list. There are only three categories to choose from the list, which are Main Highlight, Business Highlight, and Nation Highlight. The OK button will then trigger the progress bar as shown in Figure 8.
The title list is based on the user selected the category and emotions from the previous frame. Figure 9 shows the title and emoticons in the frame, which indicated the emotion assigned to the title. Emoticon also can make user feel happy while reading the title. User can choose to click the “Read More” button for the detailed information of the selected title.

![Main interfaces of EM_ticons](image)

**Figure 8.** Main interfaces of EM_ticons

4. Evaluation
The EM_ticons system is then set for user evaluation and feedback analysis of the system functions. A questionnaire containing five close-ended questions and one open-ended question were
distributed to potential users. Overall, 86.67% users accepted the system with most of users agree that a bad news will affect human emotion. Figure 10 shows percentage of user opinion on bad news effects. Based on the figure, there are 50% of the users who agreed and 40% of the users strongly agreed that bad news will affect human emotion. Besides, that are 10% of the users felt that reading bad news will not really affect human emotion.

Next, Figure 11 shows the user satisfaction about user interest in using the system. As the chart results, there are 50% of the users who agreed and 36.67% of the users who strongly agreed with the results produced by the system. There are 13.33% of the users who feel the results did not really satisfy them. From Question 3, there are 90% of the users who wish to have the proposed EM-ticons application in their daily life. Meanwhile, there are 10% of the users who feel the opposite of this.
Figure 11. User satisfaction and interest in using the system.

Next, Figure 12 shows that from Question 4, 76.67% of the users believe that EMticons will help them in their daily life. However, there are also 23.33% of the users who believed the opposite of this.
Whereas in Question 5, 76.67% of the users like to use the EM_ticons system as a mobile applications, which is convenient them in their daily life. There are also 23.33% of the users who answered no for this question.

5. Conclusions and Future Work
This research work has open a big opportunities of more potential areas in using the concept Emotional Mining. For example, more categories of emotions can be added such as anger, excitement or disappointment. The current system only process the title to assign the happy or sad emotions. In the future, the proposed EM_ticons system is also invaluable for the history domain. This project also support the notion used by?. Whereby the phenomenon of memes will lead a giant leap in the increment of images, audios, and videos on the Internet. Therefore, mobilizing the user’s emotions through blogs, sites, facebook or twitters will help the users in understanding and analysis of the content.

Acknowledgments
This project is sponsored by the Asia Pacific University of Technology and Innovation.
References

[1] Ranelluccia J, Poitras EG, Bouchete F, Lajoiee SP and Halle N 2016 Understanding Emotional Expressions in Social Media Through Data Mining

[2] Ahmad K 2012 Affective Computing and Sentiment Analysis: Metaphor, Ontology, Affect and Terminology, s.l.: Language Resources and Evaluation Conference (LREC)

[3] Mohamed H, Ezzat A and Sami M 2015 The Road to Emotion Mining in Social Network International Journal of Computer Applications 12318

[4] Fortuna B, Galleguillos C and Cristianini N 2009 Detection of bias in media outlets with statistical learning methods Text Mining 27

[5] Thomas G 2004 Emotion mining by Liam Fahey & Tom Snyder. Available at: http://www.zibs.com/emotionalmining.shtml (Accessed: 21 February 2017)

[6] Rao Y, Li Q, Mao X and Wenyin L 2014 Sentiment topic models for social emotion mining Information Sciences 66 90-100

[7] Daroff RB and Aminoff MJ 2014 Encyclopedia of the neurological sciences Academic Press

[8] Vaillant GE 2013 Psychiatry, religion, positive emotions and spirituality Asian journal of psychiatry 6 6 590-594

[9] Kim SH and Hamann S 2007 Neural correlates of positive and negative emotion regulation Journal of cognitive neuroscience 19 5 776-798

[10] Fullwood C and Martino OI 2007 Emoticons and impression formation Applied Semiotics 19 7 4-14

[11] Palazova M, Mantwill K, Sommer W and Schacht A 2011 Are effects of emotion in single words non-lexical?

[12] Evidence from event-related brain potentials. Neuropsychologia 49 9 2766-2775

[13] Lenhart A, Purcell K, Smith A and Zickuhr K 2010 Social media and young adults Pew Internet & American Life Project 3 Hong S and Nadler D 2012 Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience Government Information Quarterly 29 4 455-461

[14] Yassine M and Hajj H 2010 A framework for emotion mining from text in online social networks. In 2010 IEEE International Conference on Data Mining Workshops (ICDMW) 1136-1142

[15] Suna S, Kong, G and Zhao C 2011 Polarity words distance-weight count for Opinion Analysis of Online News Comments.

[16] Sampson TD 2012 Virality: Contagion theory in the age of networks. U of Minnesota Press

[17] Shifman L 2013 Memes in digital culture. MIT Press

[18] Spitzberg BH 2014 Toward a model of meme diffusion (M3D) Communication Theory 24 3 311-339