

Social Network User's Content Personalization based on Emoticons

Anthoniraj Amalanathan* and S. Margret Anouncia

School of Computing Science and Engineering, VIT University, Vellore - 632014, Tamil Nadu, India;
aanthoniraj@vit.ac.in, smargretanouncia@vit.ac.in

Abstract

The social networks help sharing user's multidimensional content which includes text, image, audio and video at any time. Emoticons are helpful in precise content sharing as an alternative of text and need to be analyzed for sharing of the right content. The content being shared mostly reflects the behavioral characteristics of the users and imitates their emotions. Therefore, each emoticon needs to be mapped with standard emotions. The emoticons proposed by Unicode consortium are considered and mapped with nine basic emotions such as love, happiness, pity, furious, heroic, fearful, disgust, wonder and peace. A prediction model based on decision tree classifier is designed to classify user's content according to the emotions expressed through the emoticons, especially for tweets. The designed methodology is demonstrated using two thousand tweets. Tweets are adopted for its simplicity and limited processing with only hundred and forty characters. The outcome obtained by applying the designed methodology provided satisfactory results of 83% accuracy which is more than the average accuracy (75%) of standard machine learning classification process. Therefore, it is possible to guess the behavior of the users through sharing the different forms of emoticons at various instances. This classification of users' content would reflect the dominant emotions possessed by them. This finding helps in understanding the basic nature of an individual in social networks. Having identified the basic nature of an individual through emoticons, it is very easy to personalize the user's social network page to filter disinterested and disgusting content at any time.

Keywords: Classification, Emoticons, Emotions, Personalization, Social Networks, User Behavior

1. Introduction

The content shared through the social networks platforms expresses the emotion of an individual sharing the content. Characterizing the users based on their most influencing emotion would help in many purposes¹. Recently, Twitter has become as a status quo for exchanging one's thought in a precise way. Further tweets with emoticons are currently becoming very popular as it makes the saying "picture is worth a thousand words" to reality and helps to express emotions in an effective way. Hence, the authors have attempted to study the influence of emoticons as an emotion dispenser in content sharing through social network especially in tweets.

With the boom of internet usage in youngsters and teenagers, from 2008, the growth of social network specifically, Twitter has become very active in sharing

content. According to statistics, around 500 million of tweets are dispensed per day. Each tweet expresses different form of emotions.

As per literature², nine emotions (love, happiness, wonder, heroism, peace, anger, sadness, fear and disgust) are expressed normally by human being at different occasions. These emotions are a mandatory part of human nature that can be considered as hereditary. Also it has been found that expression of a particular emotion by different human being is identical. Some persistent emotions that last much longer result in mood.

Various attempts had been performed to bring correlation of emotions with the content. The existence of strong relationship between emotion and social network had been analyzed by various researchers³⁻⁸. Out of the popular social network, Twitter had been serving as a good means of emotion diffuser through the tweets^{4,9,10}.

*Author for correspondence

According to these researchers attempts, content, topic, messages, type of conversation, social ties, amount of network activity of individuals, participatory patterns and language use are acting as emotion carrier in Twitter.

A recent research has identified micro blogging or Twitter as online word-of-mouth branding¹¹. The large amount of information contained in micro blogging web-sites makes an attractive source of data for emotion analysis. As most of the micro blogging mechanism encourages emoticons as means of content diffusion, an emoticon within a message represents an emotion for the whole message and all the words of the message are related to this emotion.

1.1 Emoticons

In general, emoticons are graphical representation of facial expressions. According to Asteroff¹² emoticons are referred as “relational icons”, while Sanderson and Dougherty¹³ defined emoticons or “smileys” as a sequence of ordinary characters that is found in keyboard of a PC.

Thompson and Foulger¹⁴ referred to them as “pictographs” which is used to express emotions as that of facial expression. Rezabek and Cochenour¹⁵ defined emoticons as “visual cues formed from ordinary typographical symbols deploying emotions”. With the growth of social networks, the usage of these smileys or emoticons has slowly crawled to capture a significant role in communication which required emotional stress on the content being shared. Active research is being carried out till now to perform behavior analysis of social network user using these emoticons and its interpretation in computer mediated communication^{16,17}. Derks et al.¹⁸ examined the role of emoticon in messages. The study showed that there exists an impact and influence by emoticons in message interpretations. Emoticons are helpful to express emotion, to strengthen a message and to express humor. It is used in a way similar to facial behavior in face-to-face communication with respect to social network interactions. Researches have proved that emoticons in Instant Messages (IM) provide richness to communication and do have impact on enjoyment¹⁹. Being carrier of emotions, these emoticons have been processed to perform behavior and sentiment analysis on the social network users^{10,20}. To categorize social network users with respect to the emotion they exchange through emoticons, several methodologies of data mining have been employed^{20–23}. Their attempts focused on to tweets rather than any other social media interaction.

Most of the analysis reported that emoticons have visible role in exchange of emotions through tweets especially on emotions – happiness and sadness. And there is no clue of other emotions. Hence, the authors have investigated the impact of emoticons in tweets and thereby attempted to classify the expressed emoticons into nine human emotions.

In order to proceed with the objective, emoticons as stated in Unicode standard²⁴ is taken and applied as base for user classification and henceforth for personalization. The Standard Emoticons that are categorized into nine basic emotions for reference is shown in Figure 1.

2. Methodology

The methodology for user's content personalization based on emoticons involves two major phases such as training and testing phase as shown in Figure 2.

2.1 Training Phase

Initially, training set is constructed by extracting labeled tweets of around 2000 in numbers. From the collected labeled tweets, meaningful words are extracted manually and stored as bag of words in a data store. Along with the meaning full words, emoticons also extracted and stored in the data store if present any. In addition to these two entities, emoticons that are extracted from standard library are also stored as reference for future processing. Prior

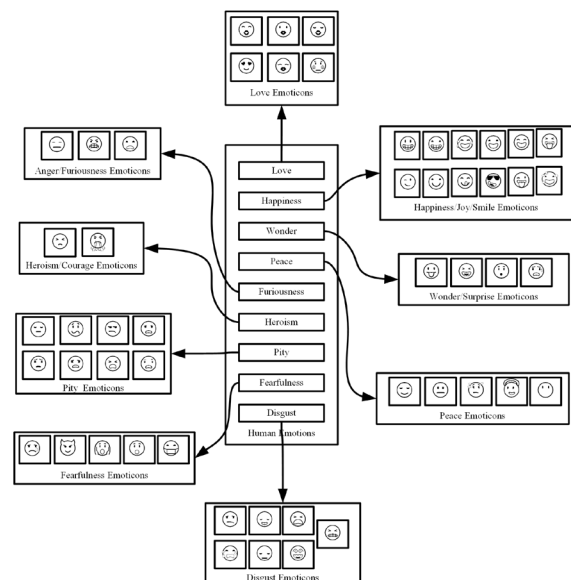


Figure 1. Categorization of emoticons into nine emotions.

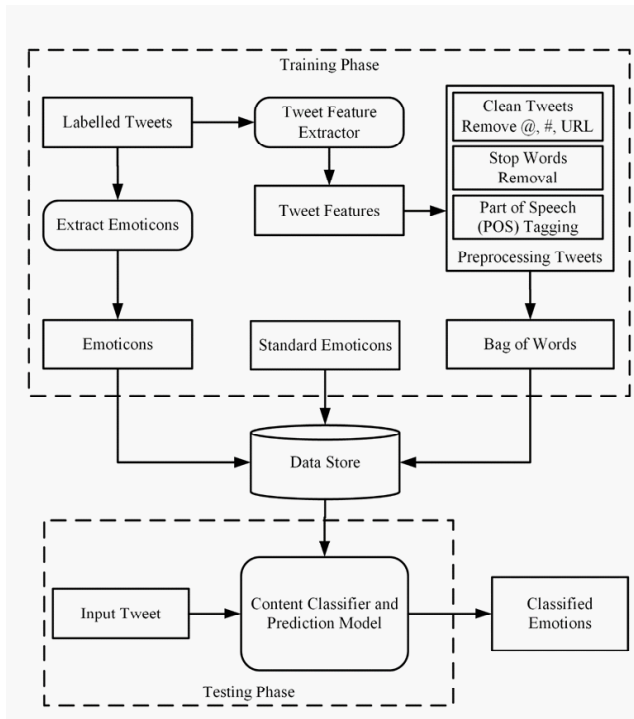


Figure 2. Developed model for classification of emotions based on emoticons.

to the creation of bag of words, the manually extracted meaning full words are subjected to preprocessing that involves the following subtasks.

- Cleaning Tweets.
- Removal of Stop Words.
- Tagging of Part of Speech.

2.1.1 Cleaning of Tweets

The data cleaning process on tweets involves the following subtasks and is intended to remove the meaningless words.

- URL removal.
- Decoding of Unicode.
- Apostrophe lookup.
- Slang replacement.
- Removal of special characters.

2.1.2 Removal of Stop Words

This process is implemented to remove the irrelevant words from the tweets. It is performed by language stemming. It is a information retrieval process to reduce or derive an inflected words to their root to return a word.

2.1.3 Tagging of Part of Speech

The Tagging of part of speech involves in extracting part-of speech based Penn Treebank corpus. And hence, adjectives, adverb, noun and verb are segregated to form a bag of words.

2.2 Testing Phase

In this phase, a tweet is really tested to get users content classified according to emotions based on emoticons and bag of words. A decision tree based classifier and prediction model is constructed to perform the emotion classification. A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. Decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorization and generalization of a given set of data. Data comes in records of the form:

$$(t, E) = (t_1, t_2, t_3, \dots, t_k, E)$$

The dependent variable, E (Emotion), is the target variable that we are trying to understand, classify or generalize. The vector $t(\text{Tweet})$ is composed of the input variables, t_1, t_2, t_3 etc., that are used for that task. Thus the preprocessed text from the tweet forms the training data for further classification. The procedure developed for the same is given in Algorithm 1.

Algorithm 1. Tweets features extraction

Procedure: extract_features()

Input: Tweet with Emoticons

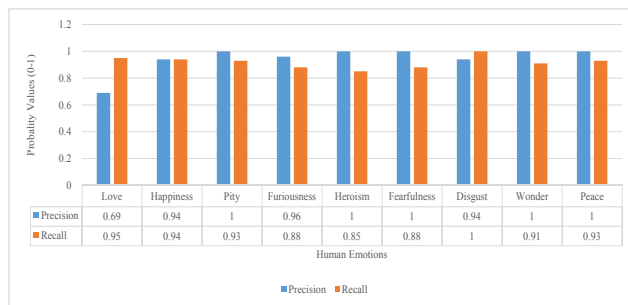
output: Features

```

1. slang_dict = {...}
2. remove_tags = {...}
3. for tweet in labelled_tweets:
4.     tokens = tweet.split(' ') //generate token of words from tweet
5.     for word in tokens:
6.         if word.startswith(remove_tags):
7.             tokens.remove(word) // remove tagged words
8.         if word in any slang_dic:
9.             word=slang_dic[word] // replace with slang words
10.    features_labels.add(feature,label)
11. return features_labels
    
```

Table 1. Precision and recall results

Emotions	Precision	Recall	F1-Measure
Love	0.69	0.95	0.80
Happiness	0.94	0.94	0.94
Pity	1.00	0.93	0.97
Fur iousness	0.96	0.88	0.92
Heroism	1.00	0.85	0.92
Fearfulness	1.00	0.88	0.92
Disgust	0.94	1.00	0.97
Wonder	1.00	0.91	0.95
Peace	1.00	0.93	0.96

**Figure 3.** Precision and recall graph.

3. Performance Analysis

The developed model is evaluated by applying the information retrieval metrics precision and recall and F1 Measure. The following Table 1 indicates the results obtained after having applied the developed model. The graph of the result is shown in the following Figure 3. From the graph, it is noted that the accuracy of classification for each emotion through emoticons appears to be satisfactory.

4. Conclusion

A decision tree based user's content classifier and prediction model is designed to classify tweets according to emotions expressed through the emoticons. According to machine learning benchmarking, any classification result of around 75% is said to be appreciable and satisfactory one. The performance of this developed model yielded classification accuracy of around 83%. Hence, the model performance appears to be satisfactory for user's content classification based on emotions through emoticons. Due to the lack of semantics the model may not outperform with the present form. However, the classified

content provides an insight of the user's behavior which may further be processed to create personalization of the user content easily in social networks.

5. References

1. Boyd DM, Ellison NB. Social network sites: Definition, history, and scholarship. *Journal of Computed Mediated Communication*. 2007 Oct; 13(1):210–30.
2. Marchand P. The yoga of the nine emotions. The tantric practice of rasa sadhana based on the teachings of Harish Johari. India: Inner Traditions; 2006. p. 1–212.
3. Arafat J, Habib MA, Hossain R. Analyzing public emotion and predicting stock market using social media. *American Journal of Engineering Research*. 2013; 2(9):265–75.
4. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. *ICWSM*. 2013 Jul; 1–10.
5. Espinosa KJP, Bernales AMJ. Characterizing influence factors affecting emotion diffusion in facebook. *Proceedings of the World Congress on Engineering and Computer Science*; 2014 Oct. p. 1–6.
6. Hill AL, Rand DG, Nowak MA, Christakis NA. Emotions as infectious diseases in a large social network: The SISa model. *Proceedings of Royal Society B-Biological Science*; 2010 Jul. p. 1–8.
7. Jalonen H. Social media and emotions in organisational knowledge creation. *Federated Conference on Computer Science and Information Systems (FedCSIS)*; 2014. p. 1371–9.
8. Tang J, Zhang Y, Sun J, Rao J, Yu W, Chen Y, et al. Quantitative study of individual emotional states in social networks. *IEEE Transactions on Affect Computing*. 2012 Apr-Jun; 3(2):132–44.
9. Kim J, Gupta P. Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research*. 2012 Jul; 65(7):985–92.
10. Tchokni S, Seaghdha DO, Quercia D. Emoticons and phrases: Status symbols in social media. *8th International AAAI Conference on Weblogs and Social Media*; 2014. p. 10485–94.
11. Jansen BJ, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. *Journal of American Society Information Science Technology*. 2009 Jul; 60(11):2169–88.
12. Asteroff JF. Paralanguage in electronic mail: A case study. Columbia: Columbia University New York; 1987.
13. Sanderson D, Dougherty D. Noah Webster of Smileys. O'Reilly Media, Inc.; 1993.
14. Thompson PA, Foulger DA. Effects of pictographs and quoting on flaming in electronic mail. *Comput Human Behavior*. 1996; 12(2):225–43.

15. Rezabek LL, Cochenour JJ. Visual cues in computer-mediated communication: Supplementing text with emoticons. *Journal of Visual Literacy*. 1998; 18(2):201–15.
16. Derks D, Bos AER, Von Grumbkow J. Emoticons and online message interpretation. *Social Science Computer Review*. 2007 Dec; 26(3):378–88.
17. Walther JB, D'Addario KP. The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review*. 2001; 19(3):324–47.
18. Derks D, Bos Arjan ER, Von Grumbkow J. Emoticons in computer-mediated communication: Social motives and social context. *Cyber Psychology Behavior*. 2008 Feb; 11(1):99–101.
19. Huang AH, Yen DC, Zhang X. Exploring the potential effects of emoticons. *Information Management*. 2008 Nov; 45(7):466–73.
20. Hogenboom A, Bal D, Frasincar F, Bal M, de Jong F, Kaymak U. Exploiting emoticons in sentiment analysis. *Proceedings of the 28th Annual ACM Symposium on Applied Computing SAC'13*; 2013. p. 703–10.
21. Ganesan KA, Sundaresan N, Deo H. Mining tag clouds and emoticons behind community feedback. *Proceedings of the 17th International Conference on World Wide Web, WWW'08*; 2008. p. 1181–2.
22. Pak Er, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. *LREC*; 2010. p. 1320–6.
23. Thelwall M, Wilkinson D, Uppal S. Data mining emotion in social network communication: Gender differences in MySpace. *Journal of American Society Information Science and Technology*. 2010; 61(1):190–9.
24. The Unicode Standard V8. Emoticons - Unicode Range 2015 May 10. Available from: <http://www.unicode.org/charts/PDF/U1F600.pdf>