International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)

Facebook as a Corpus for Emoticons-Based Sentiment Analysis

Geetika Vashisht[1], Sangharsh Thakur[2]
[1]*Assistant Professor, Kalindi College, Delhi University*
[2]*Lead Engineer, SRI Noida, India*

***Abstract*— With the growing popularity of the social networking sites, usage of informal language, short cuts & emoticons is increasing rapidly. The use of emoticons in text in order to express sentiments is posing a challenge to the automated sentiment analysis tools to correctly account for such graphical cues for sentiment. This paper aims at demonstrating how emoticons typically convey sentiments and how we can exploit emoticons by using a manually created emoticon sentiment lexicon and then using finite state machines to find out the polarity of the sentence or paragraph. We evaluate our approach on 1,250 Facebook status and 2,050 Facebook comments, which all contain emoticons and have been manually annotated for sentiment. We identified the most commonly and frequently used emoticons & classified them on the basis of the sentiment they strengthen which eventually decides the polarity of the sentence. In this paper we want to introduce a method to perform a sentiment analysis on text-based status updates & comments, disregarding all verbal information and using only emoticons to detect both positive and negative sentiments.**

***Keywords*-- Sentiment analysis, Opinion Mining, emoticons, lexicon-based sentiment analysis, Polarity classification, Finite State Machines**

I. INTRODUCTION

Social networking sites have provided an easy and attractive way to exchange ideas, opinions, knowledge and lots more. People can update status which gives a picture of what's going on in their minds. These statuses can reflect an individual's mood, their achievement or failure, their opinion on movies, products, brands, or a service. Further, the comments on the same can help to gather other people's opinion on the same, thus, providing an overwhelming and valuable piece of information on a particular subject. . Current estimates of the total number of internet users in the world (as on 6th June 2014, 5.30pm) by internet live stats[8] is 2,909,949,700 ; the number of blogs written are 2,546,980; tweets sent are 448,890,789; Facebook active users are 1,258,847,300 and the number is multiplying at a faster pace. These statistics prove how applicable it will be to exploit the World Wide Web for mining opinion. Developing an opinion tracking system can prove to be a boon to the marketing and advertising industry.

In this paper, we focus on the well liked social networking site Facebook. It can be a target for the marketing & advertising industry to gather opinions about a particular product or a service. Text-based sentiment analysis can be a bit challenging as people generally don't pay any heed to the spellings and deliberately modify the spellings of the words and use short forms whenever required. For example, many people write the word "awesome" as "ausum", "good" as "gud" and the list goes on. This poses a challenge to correctly process the language and find out the polarity of the sentence or the paragraph. On the other hand, Emoticons prove to give a correct insight of the sentence or text. Emoticons are small images or conjunctions of diacritical symbols. People can easily express themself to their friends when they are happy, sad, in love, cranky, or experiencing just about any other emotion while decorating their Facebook status and comments. They are free to choose emoticons from the myriad of emoticons list that best suits their mood. So, we aim to explore and investigate the role of emoticons in analyzing the associated sentiments. The remainder of this paper is structured as follows. First, Section II throws a light on sentiment analysis and how emoticons are used in computer-mediated communication. Then, in Section II, we identify the possible set of emoticons majorly used by people on Facebook and used them to classify polarity of the texts. We used finite state machine to implement the same. Last, in Section IV, we draw conclusions and propose directions for future work.

II. BACKGROUND

A. Sentiment Analysis

Sentiment analysis refers to a broad area of natural language processing, computational linguistics, and text mining [1]. Typically, the goal is to determine the polarity of natural language texts [1]. When we have face-to-face or verbal communication, then our gestures, facial expressions, tones and intonations assist us in expressing us in the most accurate way. Also, it's easier to express strong approval/ acceptance or disapproval/rejection in verbal communication.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)

But with the advancement of Information & technology, people frequently use emails, text-messages, instant messengers, social networking sites etc. to express themselves using only texts (ignoring the video conferencing features being provided with instant messengers). A text message may provide the desired information but fails to convey the appropriate emotion attached with the words being conveyed. In this light, we define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment [1]. Sentiment analysis is a hotly discussed topic these days. Many approaches of sentiment analysis of natural language text have been proposed in the recent years. Many approaches treat the words appearing in the text as an unordered collection of the words. This kind of an approach allows for vector representations of text, which requires the use of supervised machine training techniques for classifying the polarity of text. This requires a corpus of training data in order to function properly. Moreover, even though machine learning classifiers may perform very well in the domain that they have been trained on, their performance drops significantly when they are used in a different domain [2]. Taking this limitation into account, alternative lexicon-based methods have come into picture in recent research [1, 3, 4, 5, 6, 7], not in the least because they have been shown to have a more robust performance across domains and texts [2]. Lexicon-based sentiment analysis approaches use sentiment lexicons for retrieving the polarity of individual words and aggregate these scores in order to determine the text's polarity [1]. These approaches don't take into account emoticons while analyzing the text to determine its polarity despite of the fact that an emoticon may signal the intended sentiment of an otherwise objective statement, example,

"It is raining outside :-(". Apart from that, sarcasm or funny statements could only be identified if they have an associated emoticon. Take for instance, X sarcastically comments on a movie saying, "I have not watched such a funny movie ever ;-p" Though the statement seems to have a positive polarity but " ;-p " emoticon clears the intention behind it. These findings can be used to exploit emoticons to improve the existing approaches of sentiment analysis to get the most appropriate inference. :D

B. Analysing Emoticons

It is a cumbersome task for machines to understand the context of the text and aid in sentiment analysis. In this scenario, emoticons can be valuable cues for deriving intended sentiment. Many people are using a smiling or thumbs up emoticon to second someone's opinion, giving a huge scope to figure out the intended sentiment which other text based approaches will fail to do. The first emoticon was used on September 19, 1982 by professor Scott Fahlman in a message on the computer science bulletin board of Carnegie Mellon University [1]. In his message, Fahlman proposed to use ":-)" and ":-(" to distinguish jokes from more serious matters, respectively[1]. Since then the use of emoticons is growing at a rapid rate. With the advancement of information and communication technology, emoticons have paved their way into every day communication. Exploiting emoticons to determine associated sentiment is basically what we aim for in emotion based sentiment analysis.

III. Role Of Emoticons In Sentiment Analysis

On analysing 1,250 Facebook statuses and 2,050 Facebook comments, we found that at least twenty different kinds of different expressions were used by Facebook users to express themselves. Also, there were at least four ways different keyboard keys combinations used by people for expressing an emotion. Take for instance, in order to express happiness, keys combination used were either of the following:-

:-D	:D	=D

So, another key challenge we faced was to create a record of all such different typographical symbols that can be used for expressing a particular emotion. After a deep analysis of the same, we concluded the results in the Table I.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)

TABLE I
BAG OF EMOTICONS AND ASSOCIATED SENTIMENTS

SNo.	Classification of Emoticons	Typographical symbols used				Sentiment
1	Happiness/ Smile	:-)	=)	:)	:]	Positive
2	Sadness	:- (	=(	:[	: (	Negative
3	Wink	; -)	;)			Positive
4	Teasing/ kidding	;-P	;P	;-p	;p	Positive
		:-P	:P	:-p	:p	
5	Amused	:-D	:D	=D		Positive
6	Anger	>: (	>: - (			Negative
7	Kiss	:-*	:*			Positive
8	Confused	o.O	O.o			Negative
9	Embarrased	:-$	=$	>///<		Negative
10	Devil	3:)	3: -)			Negative
11	Cool	8-)	B-)	8)	B)	Positive
12	Unsure	:/	:-/	:\	:-\	Negative
13	Cry	:'(	:'-(			Negative
14	Love/Heart	<3				Positive
15	Shy	^_^				Positive
16	Blessed	O: -)	O:)			Positive
17	Hug	>(^_^)<	<(^_^)>			Positive
18	Squint	-_-				Negative
19	Surprised	:-o	:-O	:o	:-o	Positive
20	Thumbs up	(y)				Positive
21	Thumbs down	(n)				Negative

TABLE II
TYPICAL EXAMPLES TO SHOW HOW EMOTICONS SUPPORTS THE POLARITY OF THE SENTENCE [1]

Sentence	How	Sentiment
I am enjoying working on Sentiment analysis : D	Intensification	Positive
-_- Lunch was delicious.	Negation	Negative
This is a good article. :)	Only sentiment	Positive
I love you with all my might. <3 ^_^	Intensification	Positive
The movie was not worth spending money. >:- (	Intensification	Negative

On analyzing thousands of Facebook statuses, we found out that frequency of usage of emoticons was highly influenced by the age-group. Those in teens and early 20s were actively using emoticons in every status/ comment. Also, 60% of the users of the same age group updated their statuses about movies and restaurants reviews. Moreover, The users did not fail to use emoticon to strengthen their opinion on the same enabling Facebook to be a good target for opinion mining.

IV. IMPLEMENTATION USING FSM

A Finite State Machine (FSM) is a basically a state machine used to design computer programs. The machine has a start state, it consumes an input at a time and moves to another state. It remains in one state at a time. There can be many final states though. The behavior of state machines can be observed in many devices in modern society which perform a predetermined sequence of actions depending on a sequence of events with which they are presented. Examples are vending machines which dispense products in case the proper combination of coins is deposited, traffic lights, elevators, combination locks & lots more.

We have used the finite state machine to take typographical symbols or emoticons as inputs and based on the sequence of the typographical symbols, the FSM concludes with the sentiment associated with the emoticon. The following table presents the set of tokens which are sufficient to denote all the emoticons. The FSM accepts these tokens as inputs.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)

TABLE III

Tokens	**Token Type**
:	Colon
-	Hyphen
)	Open-parenthesis
(	Close-parenthesis
]	Close bracket
[	Open bracket
|	Pipe
_	Underscore
3	Interger_literal_three
8	Interger_literal_eight
0	Interger_literal_zero
O	Character_capital_O
,	Apostrophe
\	Backslash
/	Forward slash
<	Less Than
>	Greater Than
;	Semi-colon
*****	Asterisk

The Finite State Machine in figure 1 is used to demonstrate five kinds of sentiments. This Finite State Machine can be extended to show all the twenty one emotions presented in Table I. This machine can further be updated to include DEAD/ TRAP state i.e. if the sequence of the typographical symbols doesn't make a valid combination that corresponds to a valid emoticon then machine will enter a trap state instead of reaching the final state.

To sum up, we can use Finite State Machine to identify the presence of emoticons in any text and determine the sentiment or opinion signaled by that text.

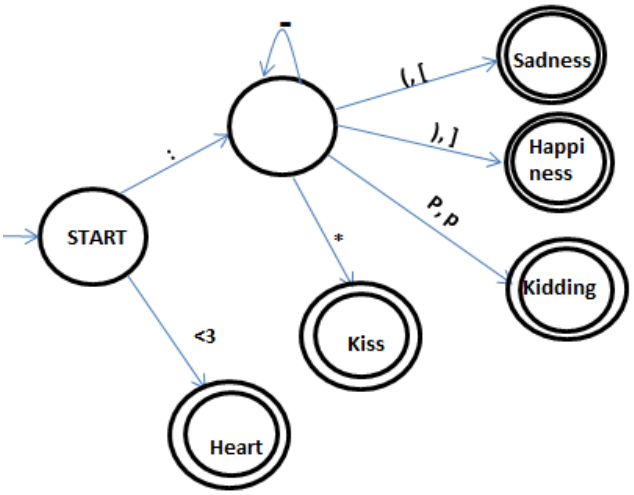

Figure 1: Finite State Machine which exploits Emoticons to determine text polarity.

V. Conclusion And Future Scope

Now a day, people are increasingly using emoticons to express their opinions or emotions. So, it is of utmost importance to device some tool to correctly interpret these typographical symbols for sentiment. In this paper, we have analysed the role that emoticons play in delivering the overall sentiment of the text. We identified the commonly used emoticons and exploited them to devise a finite state machine that takes these typographical symbols as an input and conveys the associated sentiment as an output. This approach can be used together with a lexicon-based sentiment analysis method to validate the result. Our bag of emoticons was a result of a deep analysis of thousands of Facebook statuses and comments. We also plan on improving the precision of the emoticon analysis by analyzing other computer mediated communications, probably on some other messengers. Also a combination with an analysis of verbal information will be taken into consideration.

REFERENCES

[1] Alexander Hogenboom, Daniella Bal, Flavius Frasincar. .Exploiting Emoticons in Sentiment Analysis.

[2] M. Taboada, K. Voll, and J. Brooke. Extracting Sentiment as a Function of Discourse Structure and Topicality. Technical Report 20, Simon Fraser University, 2008. Available online, http://www.cs.sfu.ca/research/publications/techreports/#2008.

[3] C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sago_, and V. Subrahmanian. OASYS: An Opinion Analysis System. In AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (CAAW 2006), pages 21{26. AAAI Press, 2006.

[4] A. Devitt and K. Ahmad. Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. In 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007), pages 984{991. Association for Computational Linguistics,2007.

[5] X. Ding, B. Lu, and P. Yu. A Holistic Lexicon-Based Approach to Opinion Mining. In 1st ACM International Conference on Web Search and Web Data Mining (WSDM 2008), pages 231-240. Association for Computing Machinery, 2008.

[6] B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong. Polarity Analysis of Texts using Discourse Structure. In 20th ACM Conference on Information and Knowledge Management (CIKM 2011), pages 1061-1070. Association for Computing Machinery, 2011.

[7] Heerschop, P. van Iterson, A. Hogenboom, F. Frasincar, and U. Kaymak. Analyzing Sentiment in a Large Set of Web Data while Accounting for Negation. In 7th Atlantic Web Intelligence Conference (AWIC 2011), pages 195{205. Springer, 2011.