

A Survey on Mining User Opinion from Texts and Emoticons

Nasiya Najeeb¹, Sangeetha Jamal²

Abstract— Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product or topic. Lots of people write their opinions in forums, micro blogging or review websites. This data is very useful for business companies, governments, and individuals, who want to track automatically attitudes and feelings in those sites. Today Social Network Sites (SNS) allows people to express their feelings and convey their emotions via texts as well as emoticons. Information in terms of texts are extracted and clustered into emotions and then classified into positive, negative and neutral. Today the most common approach followed by users is using emoticons along with texts. An emoticon is generally more likely to affect a sentence in which it occurs since it provides a better sentiment expression of a web user. Emoticons are of two types: textual emoticons and graphical emoticons. In this paper, we have described the different approaches used for mining opinion from both texts and emoticons to assist future research in this area. We have performed a systematic literature review process to conduct this survey.

Keywords— Sentiment Analysis, Opinion Mining, Text-based sentiment analysis, Emoticon-based sentiment analysis.

1 INTRODUCTION

Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance in business and society as a whole. The growth in sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. Sentiment Analysis is a hotly discussed topic these days.

Social networking sites have provided an easy and attractive way to exchange ideas, opinions, attitudes and lots more. When people update their status it gives a picture of what's going on in their mind. Individuals are always interested in others opinion about products, services, issues and event for finding best choices. So sentiment analysis is always suitable for decision making process. It is a natural phenomenon that good decision can be taken based on the opinion of others. Sentiment analysis is applicable in many areas like marketing products, politics, educational purposes etc.

Today many of the social networking sites provide the facility to express opinions using both text and emoticons which helps the public to produce their strong emotions. So there should be methods to mine sentiment from both text and emoticons. In this paper, we have performed a survey on Text-based sentiment analysis and Emoticon-based sentiment analysis and the different methodologies used for this.

discusses the related works in this area, in section 3 we have discussed the methodology, section 4 gives an overview on text-based sentiment analysis, section 5 gives an overview on emoticon-based sentiment analysis, the different approaches used for both text-based and emoticon-based sentiment analysis are discussed in section 6 and finally we conclude in section 7.

2 RELATED WORKS

Many researches have been done on both Text-based sentiment analysis and Emoticon-based sentiment analysis. Past studies on Text-based sentiment analysis says that automatic detection of emotion in texts is becoming important from an applicative point of view [1]. By analyzing people's sentiment, the emotions of the public toward a particular issue can be observed, experimented and quantified [2]. Today may websites also provides the facility for using emoticons. We define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment [3]. Over the years, people have embraced the usage of so-called emoticons as an alternative to face-to-face visual cues in computer-mediated communication like virtual utterances of opinions. Using emoticons along with text helps to strengthen their opinion. Opinion can be collected from any person in the world through review sites, blogs, twitter etc.

3 METHODOLOGY

In this survey, we have used a systematic review process.

A. SEARCHING

Some websites were used for searching. First we tried to search for relevant papers, journals and presentations. Relevant papers were identified using

- IEEE Explore
- ACM portal

1. Nasiya Najeeb is currently pursuing masters degree program in Computer Science and Information System, Rajagiri School of Engineering and Technology, Kochi, Kerala, India. E-mail: nasiya.nachu@gmail.com

2. Sangeetha Jamal is currently working as Asst. Professor in the Dept. of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kochi, Kerala, India.

The rest of this paper is organized as follows. Section 2

- Google search engine

B. SCREENING

Screening was performed by dividing papers into: on or before 2005 and after 2005. Only relevant results were included and others were excluded.

The following studies were included:

- Works in English language
- Published research and unpublished research.
- Focus on Opinion Mining

4 TEXT-BASED SENTIMENT ANALYSIS

Past studies on text-based sentiment analysis focus on writer’s emotional state. All the texts extracted for opinion mining can be grouped in to 3 sets [5]:

1. Texts containing positive emotions, such as happiness, amusement or joy
2. Texts containing negative emotions, such as sadness, anger or disappointment
3. Objective texts that only state a fact or do not express any emotions

Text extraction is done manually or automatically. Immediately after extraction, filtration is performed using some automated tools. For example, a research done on Twitter extracts relevant and eliminates irrelevant contents from tweets by using information retrieval techniques and then filters the data after the extraction. The filtration includes emoticon replacements, upper and lower casing, and removal of stop words, repeated words as well as punctuations [6][7]. After performing the filtration of data, various classifiers or methods are used to classify the texts into emotions. According to Zhang, Machine Learning Methods is one of the simplest method to classify texts into emotions. Now to extract the emotion of each word, a Word Emotion Algorithm [9] is used. Emotions may be expressed by a single word or a group of words. Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on.

5 EMOTICON-BASED SENTIMENT ANALYSIS

Besides mining and analyzing texts in sentiment analysis, emoticons are also taken into account to classify the emotions that the people conveyed since the emoticons are able to support voice inflections, facial expressions as well as bodily gestures on the SNS [10]. However, in plain-text computer-mediated communication, such visual cues are lost. Also Text-based sentiment analysis can be a bit challenging as people generally do not give importance to the spellings. For example, many people write the word “good” as “gud”, “because” as “bcz” etc. This poses a challenge to correctly process the language and find out the polarity. Over the years, people have appreciated the usage of emoticons as an alternative to face-to-face visual cues in computer-mediated communica-

tions. So we define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment. 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 [3]. In his message, Fahlman proposed to use “:-)” and “:- (“ to distinguish jokes from more serious matters, respectively.

Today Facebook and Twitter are the most common social networking sites which supports a large set of emoticons. People are free to choose emoticons from the myriad of emoticons list that best suits their mood.

There are two types of emoticons namely,

- textual emoticons
- graphical emoticons

5.1 Textual emoticons

Textual emoticons are typically made up of typographical symbols such as “:”, “=”, “-”, “^”, or “(” and commonly represent facial expressions. Emoticons can be read either sideways, like “:- (“ (a sad face), or normally, like “(^ -^)” (a happy face).

Emoticons are grouped into two classes: Positive class and Negative class.

- Positive class: messages containing positive emoticons.
- Negative class: messages containing negative emoticons.

Table I show samples of positive and negative emoticons .

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	Embarrassed	:-\$	=\$	> <		Negative
10	Devil	3:)	3:-)			Negative

Table1: Samples of positive and Negative Emoticons

5.2 Graphical Emoticons

Graphical emoticons are small images or conjunctions of diacritical symbols. The emotion of a graphical emoticon is easy to identify compared to textual ones. Table 2 shows a set of graphical emoticons.

Graphical emoticons provides a better sentiment expression of a web user. For example, consider the following two sentences

Ex1: I feel very sad.

Ex2: I feel very sad ☹

Example 1 conveys opinion of a user using a phrase “sad”. Similarly, example 2 conveys opinion of a user using a phrase “sad” as well using an emoticon. The use of emoticon provides an accelerated positive opinion of a user compared to opinion expressed in example 1.

6 APPROACHES FOR SENTIMENT CLASSIFICATION

There are mainly two approaches for sentiment classification: Machine learning approach and Lexicon based approach.

The Machine learning approach uses linguistic features and machine learning algorithms. The text classification methods using Machine learning approach can be divided into supervised and unsupervised learning methods. The supervised methods use a large number of labelled training documents. The unsupervised methods are used when it is difficult to find these labelled training documents. Most commonly used Machine Learning methods to classify opinion are Support Vector Machine (SVM), Naïve Bayesian Classifier, k-Nearest Neighbor(k-NN) and Maximum entropy method. Some other methods like Decision Tree, Neural Network, Latent Dirichlet Allocation (LDA), and Probability Latent Semantic Analysis (PLSA) are also used for this purpose. But they are not used much because of some drawbacks.

The Lexicon-based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity.

Sometimes a combination of the Machine Learning methods and the Lexicon-based approach is used, which is also known as Hybrid approach. The combinations of the two methods are performed due to the advantage and disadvantage of each method. Machine Learning techniques have the benefit of having a high accuracy, but the drawback of this technique is the involvement of an extensive manual work in annotating data. On the other hand, the Lexicon-based approach allows many domains to have an ease of use but unfortunately, the approach is unable to yield comparable accuracy to the Machine Language techniques [13].

Lexicon-based Approach can also be used for extracting sentiments from emoticons. But Lexicon-based approach

ID	Emoticon	Code	Description	ID	Emoticon	Code	Description
1		:)	happy	11		:O	surprise
2		:(sad	12		X-(angry
3		:)	winking	13		:>	smug
4		:D	big grin	14		B-)	cool
5		::)	batting eyelashes	15		:-S	worried
6		:/	confused	16		>:)	devil
7		x	love struck	17		:((crying
8		:'>	blushing	18		:))	laughing

Table 2: Set of Emoticons

shows more robust performance only across domains and texts, it doesn't show much good performance across emoticons. So another good approach used was implementation using FSM. A Finite State Machine (FSM) is 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 also. Here, typographical symbols or graphical emoticons were given as input and based on the sequence of the symbols, FSM concludes the sentiment associated with the emoticon. For identifying the sequence of typographical symbols, a set of tokens were used which were sufficient to denote all the emoticons. If the sequence of the typographical symbols doesn't make a valid combination that corresponds to a valid emoticon then FSM will enter a trap state instead of reaching the final state. So Finite State Machine is one of the most suitable approach used to identify the presence of emoticons in any text and determine the sentiment or opinion signaled by that text.

7 CONCLUSION

Today Sentiment Analysis is one of the emerging research areas in Natural Language Processing. The year 2001 and 2002 marks the beginning of Sentiment Analysis or Opinion Mining. Since thousands of Social Network sites are created, we expect the set of emotional expressions to keep expanding. In this paper we have performed a survey on both text-based and emoticon-based sentiment analysis and we have also highlighted the advantage of using emoticons rather than texts to express emotions. From the above discussions, we have concluded that Text-based sentiment analysis can be a bit challenging whereas emoticons are easier to be processed. Since

opinion mining is an emerging and rapidly growing field, we have performed this survey for a clear direction for our future work.

REFERENCES

- [1] Khairullah Khan, Baharum B.Baharudin, Aurangzeb Khan, Fazal-e-Malik, Mining Opinion from Text Documents: A Survey, 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies.
- [2] Abidin, S.Z.Z., Omar, N., Radzi, M.H.M., &Haron, M.B.C., Quantifying Text-based Public's Emotion and Discussion Issues in Online Forum, International Journal on New Computer Architectures and Their Applications, Vol.1, No.2, 2011, pp. 428 - 436.
- [3] Alexander Hogenboom, Daniella Bal, Flavius Frasinca, Exploiting Emoticons in Sentiment Analysis,Proceedings of the 28th Annual ACM Symposium on Applied Computing , pp.703-710.
- [4] Dudhat Ankitkumar , R. R. Badre, Mayura Kinikar, A Survey on Sentiment Analysis and Opinion Mining, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2014,pp. 6633-6639
- [5] Geeta.G.Dayalani, Prof.B.K.Patil, Emoticon-based unsupervised sentiment classifier for polarity analysis in tweets, International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014.
- [6] Potena, D., & Diamantini, C., Mining Opinions on the Basis of Their Affectivity, collaborative Technologies and Systems (CTS), 2010 International Symposium on IEEE, 2010, pp.245 - 254.
- [7] Bulik, B. S., Apple, Facebook tops for college students, *Advertising Age*, Vol.78, 2007, pp.12.
- [8] Zhang, K., Cheng, Y., Xie, Y., Honbo, D., Agrawal, A., Palsetia, D., Lee, K., Liao, W., & Choudhary, A., SES: Sentiment Elicitation System for Social Media Data, IEEE International Conference on Data Mining Workshops, 2011, pp. 129 - 136.
- [9] Devin Hsin-Yih Lin, Changhua Yang, Hsin-Hsi Chen, Emotion Classification of Online News Articles from the Reader's Perspective, 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
- [10] Huang, A.H., Yen, D.C., & Zhang, X., Exploring the Potential Effects of Emoticons, *Information & Management*, Vol.45, 2008, pp.466 - 473.
- [11] A. Ortony, G. L. Clore, and M. A. Foss. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53:751-766, 1987.
- [12] C. Strapparava, A. Valitutti, and O. Stock. The affective weight of lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy, May 2006.
- [13] Lu, B. & Tsou, B.K., Combining a Large Sentiment Lexicon and Machine Learning for Subjectivity Classification, *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, 2010, pp. 3311 - 3316.