International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 6, December 2020, pp. 5917~5922
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i6.pp5917-5922

Sentiment analysis of comments in social media

Abdulrahman Alrumaih[1], Ali Al-Sabbagh[2], Ruaa Alsabah[3], Harith Kharrufa[4], James Baldwin[5]
[1]Al-Nahrain University, College of Law, Iraq
[2]Al-Taff College University, Computer Engineering Department, Iraq
[3]University of Kerbala, College of Science, Computer Science Department, Iraq
[2,4]Ministry of Communication, Iraq
[1,5]Sheffield Hallam University, United Kingdom (UK)

Article Info

Article history:

Received Feb 22, 2020
Revised May 9, 2020
Accepted May 22, 2020

Keywords:

Complex networks
Sentimental analysis
Social media platform
Tweets

ABSTRACT

Social media platforms are witnessing a significant growth in both size and purpose. One specific aspect of social media platforms is sentiment analysis, by which insights into the emotions and feelings of a person can be inferred from their posted text. Research related to sentiment analysis is acquiring substantial interest as it is a promising filed that can improve user experience and provide countless personalized services. Twitter is one of the most popular social media platforms, it has users from different regions with a variety of cultures and languages. It can thus provide valuable information for a diverse and large amount of data to be used to improve decision making. In this paper, the sentiment orientation of the textual features and emoji-based components is studied targeting "Tweets" and comments posted in Arabic on Twitter, during the 2018 world cup event. This study also measures the significance of analyzing texts including or excluding emojis. The data is obtained from thousands of extracted tweets, to find the results of sentiment analysis for texts and emojis separately. Results show that emojis support the sentiment orientation of the texts and those texts or emojis cannot separately provide reliable information as they complement each other to give the intended meaning.

Corresponding Author:

Ali Al-Sabbagh,
Computer Engineering Department,
Al-Taff University College, Kerbala, Iraq.
Ministry of communication, Babylon, Iraq
Email: aalsabbagh2014@my.fit.edu

1. INTRODUCTION

Technology is developing rapidly and has made dramatic changes in global communication, resulting in an intensive rely on communication devices. Advancement in technology and the way people handle it permits them to be socially open and communicate efficiently via the net [1-3]. Also in [4] opinion, modern networking, such as mobile smartphones, allow users all over the world to share ideas, information, photos, or even videos in affordable prices, when compared to the traditional communication techniques.

Communication networks have significant advantage among which are: firstly, being source of news, entertainment, and education. Secondly, they are an essential tool that could lead to a dramatic change in cultures and societies. Thirdly, they can guide the political elections in some countries to a specific destination [3, 5, 6]. Furthermore, social media provide users with a virtual social environment that focuses heavily on the common interests. It is also flexible and accessible through different devices, making it easier to share resources among individuals, communities and organizations [1, 6-8].

In recent years, there has been a huge interest in smart objects that can connect to the Internet, to share and make big new services to the world like smart homes, mobile health, and for the industrial applications, such as smart grids, efficient transportation, and logistics [9]. There is a huge amount of useful

information hidden within this Big Data which needs to be mined in order to acquire the knowledge to create new opportunities and overcome rising challenges. Big data represents a challenge to the current generation of mobile network [4, 10]. In addition, the goal of social media is to harness random big data in order to obtain important information regarding public opinion, which would help build an accurate mathematical model for smarter business decisions as seen in Figure 1.

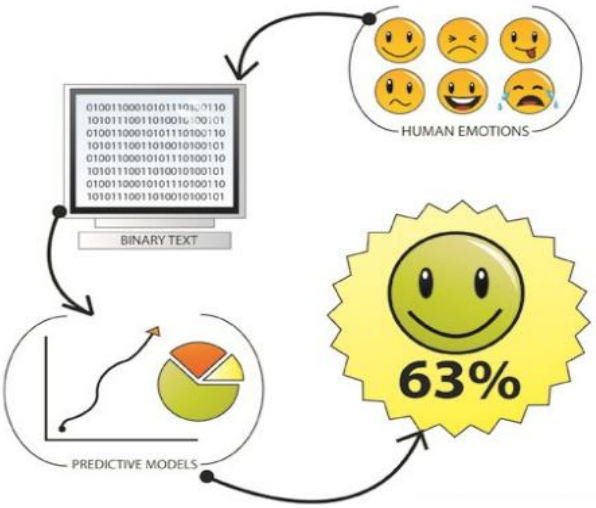


Figure 1. Sentiment analysis from social media

Upon communicating using these platforms, there is a strong tendency to use abbreviated expressions that helps in expressing feelings and emotions. In addition, emoticons and emojis stand as a tool that enhances the sentimental orientation of the speaker. Emojis are defined as "a small digital image or icon used to express an idea or emotion" (Oxford dictionary in 2018). Recently there has been an increase interest in measuring these emotions using diverse methods of analysis, among these methods is sentiment analysis. In [11, 12] defines sentimental analysis as "the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes".

The research demonstrates that there is a strong adherence to social media. People also tend to incorporate emojis and abbreviated expressions in their computer communicative interaction [13, 14]. Heart codes, smiley faces, or crying faces are reflecting the sentiment orientation of the speaker which, on many occasions, substitute the text or enhance the emotional intention. For that reason, it is necessary to take into consideration the role of emojis in any sentiment query carried out on computer mediated interaction [15, 16]. The aim of the current research is to investigate the sentiment orientation of the textual features as well as the emojis-based features of Arabic comments posted on social media, Twitter, during the World Cup event. It also measures whether different results will be obtained when the text is analyzed in isolation from emojis.

2. RESEARCH METHOD

This section presents the steps adopted in pursuing this research. It starts by describing the reason why social media is widely used as a source of data, particularly twitter. Then, different types of sampling and methods of data collection will be presents, naming the process of data collection adopted in this work. Later, the diverse techniques of analyzing data will be discussed, highlighting the sources that this study relied on in this procedure. It was of two steps; the first one was the pre-processing tasks and the second one was sentiment analysis tool. Finally, the section ends with an illustration for the design and implementation of the artefact.

This section also demonstrates the medium adopted in collecting data. Before proceeding into the details, it must be noted here that this study relies on Twitter in investigating the poster' sentiments. A substantial pervasive question is whether social media is a reliable source of data. These platforms, despite being widely used as an instrument in different types of research, they are not used by all community

members. Even those who are attracted to it use it in different ways. Until now there are little attempts that provide proofs and evidence whether these platforms represent the general population. The debate is that social media does not represent the general public is due to the lack of reliable sampling frame, making it as a source of data for only nonprobability samples [1, 4, 17]. Significantly mentioned social media present valuable insights for a wide range of inquiries not generatable for a wider population, though. Therefore, it is demanded to view it practically and objectively, paying special attention for its potential advantages and sources of errors [18, 19].

In other hand, the type of data derived from the web is usually referred to as corpora. Documents adopted in the corpus might include comparative essays, reviews, comments, tweets, … etc. They also can have a significant impact on sentiment analysis. It has been mentioned in the prior section that Twitter is a popular social network among Arabic speaking users. To accomplish the step of data collection, the manual technique has been found more appropriate to the requirement of this study as it is the case with the technique adopted by [1, 20, 21].

As a summary, the figure shows, data analysis involved three stages. This first one was data input, wherein tweets, including text and emojis have been input. To sum up, the methodological steps followed in this research listed below followed by Figure 2 that demonstrates them:

- Data collection: The Data are collected randomly using ten different Hashtags.
- Data cleaning: Removing the undesirable data and maintaining only the required tweets with emojis and compiling tweets in Microsoft word document.
- Pre-processing task: Compiling the comments into one Microsoft word documents.
- Sentiment Analyses: using two different strategies, manual and computing.
- Design an artefact that calculates the polarity of tweets depending on two lexicons of text and emoji.
- Implementation and result: Comparing the results and writing recommendations [4, 22, 23].

Data collection: The Data are collected randomly using ten different Hashtags. Data cleaning: Removing the undesirable data and maintaining only the required tweets with emojis and compiling tweets in Microsoft Word document. Pre-processing task: Compiling the comments into one Microsoft Word documents. Sentiment Analyses: using two different strategies, manual and computing have implemented in this research [1, 4, 23-26].

So far, this paper presented the methodological steps in this study, the core area in this work. It starts by highlighting the advantage of harvesting data using the social media followed by describing the types of sampling, the methods of data collection and its analysis. It ends with an illustration for the tool design and its implementation.

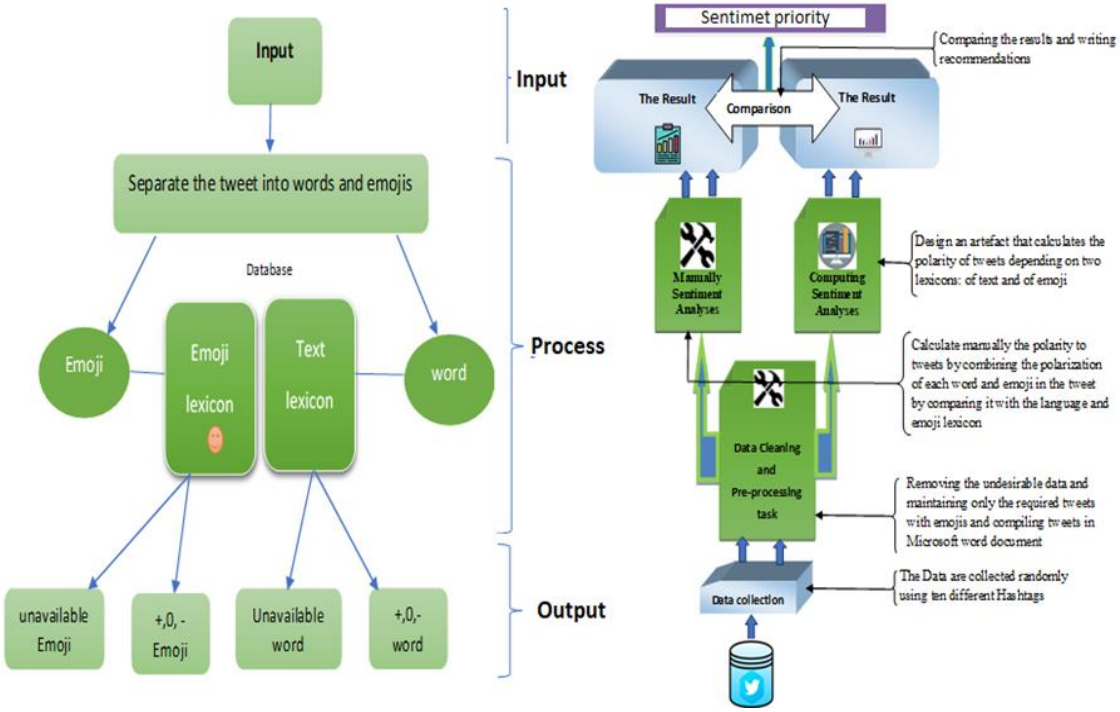


Figure 2. Framework research methodology

3. RESULTS AND DISCUSSION

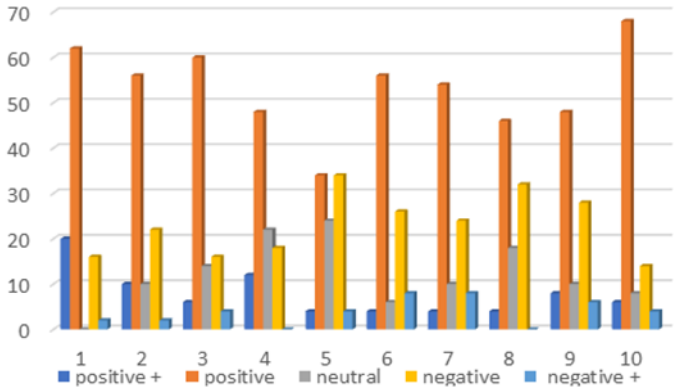
In this section, presents the results arrived at from both the computing program and the manual calculation. As has been mentioned in earlier, 10 hashtags were used in collecting data, and 100 tweets were obtained from each hashtag. Hence, Total tweets were 1000 sample. As can be demonstrated from the below chart that each raw represents a hashtag [1, 4].

After presenting the steps of data processing, below is a description of the results arrived at when computer and manual calculation of polarities have been carried out. In the first place, results arrived at using python artefact will presented. The next table and the subsequent ones display the results in 10 hashtags. Each arrow, shows the polarity of a hashtag in which 100 tweets are analysed. The first column is related to the more positive polarity, the second to the positive ones, the third shows the neutral, the forth displays the negative words, and finally the fifth column counts the more negative ones.

After processing text separately from emojis the following results were obtained for the text. The below table and Figure 3 in both windows illustrate, the total polarity is positive orientation. The tenth hashtag involves the highest positive polarity, meanwhile carries the least negative value. In the fifth hashtag, however, the positive and negative values are equal. To measure the accuracy given by the artefact, below are the results obtained after making a manual calculation for text analysis.

It can be observed that there a slight difference between the methods. Remarkably, the table and Figure 3 also demonstrate that the positive polarity in the tenth and third hashtag are equal. When a space is not inserted between two items, the artefact will consider it as one identity. Thus, most of them were counted as neutral polarity due to their unavailability in the lexicon. At the same time, results obtained from the artefact, though, were of better accuracy. This is because all words and emojis were calculated, while in the manual calculation, only the distinctive words were calculated. For example, polarity for prepositions were not taken into account in the manual calculation, while in the program, they were given a value.

positive +	Positive	Neutral	Negative	negative +
20	62	0	16	2
10	56	10	22	2
6	60	14	16	4
12	48	22	18	0
4	34	24	34	4
4	56	6	26	8
4	54	10	24	8
4	46	18	32	0
8	48	10	28	6
6	68	8	14	4



positive +	Positive	Neutral	Negative	negative +
22	57	1	16	1
13	55	12	18	0
6	67	13	17	4
10	50	20	32	2
5	33	24	37	4
3	54	6	20	10
4	56	12	28	7
5	49	18	34	0
8	50	10	30	5
7	67	10	15	4

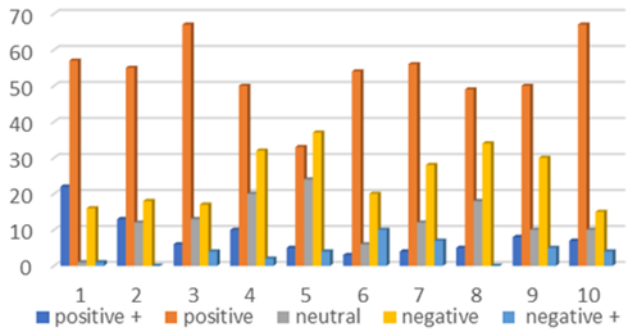
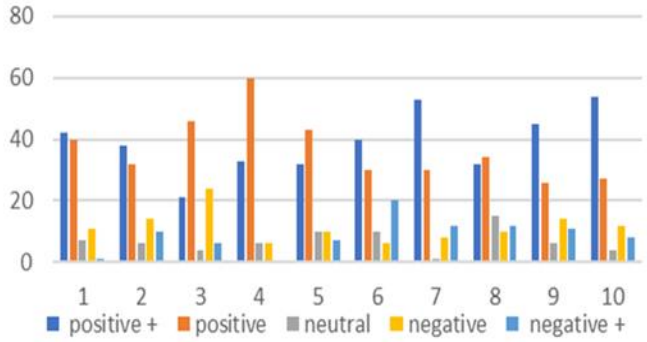


Figure 3. Results of programmed/manual text calculations [4]

As to emoji analysis, it is mentioned earlier that the same steps have been followed in their analysis. The highest positive polarity obtained from the fourth hashtag, followed by the third and fifth hashtag. However, it is clear that the more positive polarity exceeds the positive ones. The highest rate was in the in the seventh and tenth hashtags, followed by the ninth and tenth hashtags. Below are the obtained polarity calculations for ten hashtags as shown in Figure 4.

positive +	positive	neutral	Negative	negative +
42	40	7	11	1
38	32	6	14	10
21	46	4	24	6
33	60	6	6	0
32	43	10	10	7
40	30	10	6	20
53	30	1	8	12
32	34	15	10	12
45	26	6	14	11
54	27	4	12	8



positive +	positive	Neutral	Negative	negative +
40	40	6	10	2
38	30	6	14	12
20	48	4	22	6
30	58	6	6	0
34	42	10	8	6
40	28	8	6	18
52	28	2	8	10
30	34	16	8	12
44	26	6	14	10
52	26	4	10	8

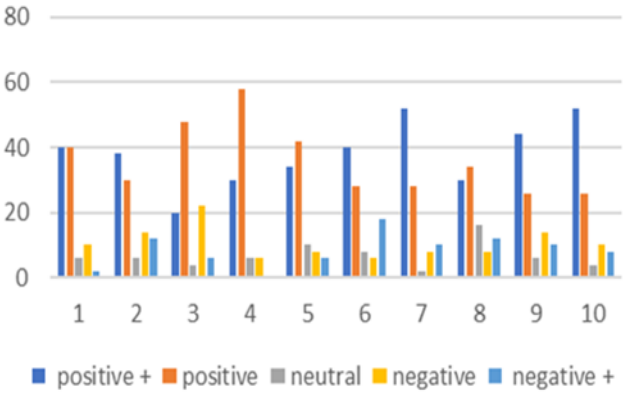


Figure 4. Results of programmed/manual emoji calculations

In comparing the above results with the polarity obtained from the textual analysis, it is realized that the sentiment orientation of emojis is higher than that of the text when it is analyzed after being separated from the emojis. The reason behind can be credited to inserting more than one emoji within a comment. Secondly, the Arabic posters were using emojis to substitutes the text; instead of writing a complete sentence supported by the emoji.

Significantly mentioned, there are many words that have been assigned 0 (or neutral) polarity. These words were not available in the lexicon. The reason behind this long list of neutral polarity is that the designer of the lexicon relied on the Jordanian dialect only in accumulating the tweets. Data in this work, however, involved comments from different Arabic dialect, which might differ in many vocabularies. Likewise, a list of new Emojis has been created because of their unavailability in the lexicon.

4. CONCLUSION

An emojis indicate explicitly the sentiment orientation of the writer if it were positive, negative or neutral. This supports the research question stated earlier that emojis work in hand with the text in identifying people's feelings and emotions. This work implicates that an analysis of any Arabic comment would be more accurate if both text and emoji are taken into account in sentiment analysis. Second, it can also analyse comments in any other social media platform, such as Facebook and Instagram. It must be bared in mind, though, that emojis differ from one platform to another. Therefore, upon analysing comments in a particular platform, emojis of that platform must be analysed. Third, it is observed that Arabic tweeter users write their comments in English using Arabic alphabet, which were not identified by the artefact that allocated a 0 polarity (neural) for them.

Finally, a future research can rely on more than one text or emoji lexicon in analysing data so as to avoid having a long list of neutral words. Further, sarcasm and the level of politeness are also valuable areas of investigation. Emojis can be a relevant indicator that assist in detecting the sentiment of the in tweets, an area which has not touched upon. By studying the polarity of (im) politeness may result in an interesting analysis.

REFERENCES

[1] A. Alrumaih, "Sentimental Analysis of Arabic Comments in Computer Mediated Communication," MSc Thesis, Sheffield Hallam University, 2018.

[2] B. Hale, et al., "The History of Social Media: Social Networking Evolution," *History, Electronics*, vol. 58, pp. 450-464, Feb. 2011.

[3] H. J. Aleqabie, et al., "Sentiment analysis for movie reviews using deep learning with semantic orientation," *8th International Conference on Applied Science and Technology (ICAST 2020)*, 2020.

[4] A. Alrumaih, et al., "Analyzing User Behavior and Sentimental in Computer Mediated Communication," *International conference on image processing andcapsule networks (ICIPCN 2020)*, Taiwan, May 2020.

[5] P. M. Figliola, "Promoting Global Internet Freedom: Policy and Technology," *Congressional Research Service*, 2013.

[6] C. Cook, "Mobile Marketing and Political Activities," *International Journal of Mobile Marketing*, vol. 5, no. 1, pp. 154-163, 2010.

[7] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1-167, 2012.

[8] M. Preisendorfer, et al., "Social Media Emoji Analysis, Correlations and Trust Modeling," Doctoral dissertation, 2018. [Online], Available: http://hdl.handle.net/1951/69574.

[9] A. A. Al-Sabbagh, et al., "An extensive review: Internet of things is speeding up the necessity for 5G," *International Journal of Engineering Research and Applications*, vol. 5, no. 7, pp. 106-112, 2015.

[10] A. Al-Sabbagh and R. Alsabah, "Internet of things and big data analysis: Recent trends and challenges," *United Scholars Publication*, 2016.

[11] A. bin S. Ahmari, "The purpose of the use of university students to social networking sites: a field study on students University of Imam Muhammad bin Saud Islamic," Unpublished MA Thesis, 2015.

[12] Milinnial Marketing, "Millennials Tech-dependent, but not Necessarily Tech-savvy," [Online], Available: http://millennialmarketing.com/2010/04/millennials-tech-dependent-but-not-necessarily-tech-savvy/.

[13] A. Y. Mjhool, "Evaluation of user perceived QoE in mobile systems using social media analytics," MSc Dissertation, Florida Institute of Technology, 2016. [Online], Available: http://hdl.handle.net/11141/1133.

[14] L. Leung, "Predicting Internet Risks: A Longitudinal Panel Study of Gratifications-Sought, Internet Addiction Symptoms, and Social Media Use among Children and Adolescents," *Health Psychology and Behavioral Medicine Journal*, vol. 2, no. 1, pp. 424-439, 2014.

[15] M. Davis and P. Edberg, "Unicode Emoji," 2020. [Online], Available: http://unicode.org/reports/tr51/.

[16] S. I. Alsanie, "Social Media (Facebook, Twitter, WhatsApp) Used, and it's Relationship with the University Students Contact with their Families in Saudi Arabia," *Universal Journal of Psychology*, vol. 3, no. 3, pp. 69-72, 2015.

[17] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, no. 4, pp. 1-21, 2009.

[18] Shatha. A. A. Hakami, "The importance of understanding emoji: An investigative study," University of Birmingham, School of Computer Science, *Research Topics in HCI*, pp. 1-20, 2017. [Online], Available: https://www.cs.bham.ac.uk/~rjh/courses/ResearchTopicsInHCI/2016-17/Submissions/hakamishatha.pdf.

[19] N. A. Abdulla, et al., "Towards improving the lexicon-based approach for arabic sentiment analysis," *International Journal of Information Technology and Web Engineering (IJITWE)*, vol. 9, no. 3, pp. 55-71, 2014.

[20] S. M. M. Seyednezhad and R. Menezes, "Understanding subject-based emoji usage using network science," in *Workshop on Complex Networks CompleNet*, pp. 151-159, 2017.

[21] S. Al-Azani and E. M. El-Alfy, "Combining emojis with Arabic textual features for sentiment classification," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 139-144, 2018.

[22] K. F. Hew and W. S. Cheung, "Use of Facebook: A Case Study of Singapore Students' Experience," *Asia Pacific Journal of Education*, vol. 32, no. 2, pp. 181-196, 2012.

[23] A. Collomb, et al., "A study and comparison of sentiment analysis methods for reputation evaluation," 2013. [Online], Available: https://liris.cnrs.fr/Documents/Liris-6508.pdf

[24] S. Aljasir, et al., "University Students Usage of Facebook: The Case of Obtained Gratifications and Typology of Its Users," *Journal of Management and Strategy*, vol. 8, no. 5, pp. 30-47, 2017.

[25] Rajan, A. Pappu, and S. P. Victor, "Web sentiment analysis for scoring positive or negative words using Tweeter data," *International Journal of Computer Applications*, vol. 96, no. 6 pp. 33-37, 2014.

[26] He, Yulan, and Deyu Zhou, "Self-training from labeled features for sentiment analysis," *Information Processing & Management*, vol. 47, no. 4, pp. 606-616, 2011.