ISSN 2616-5783 Vol.5, Issue 10: 56-60, DOI: 10.25236/AJHSS.2022.051011

A study of sentiment differences in different event-related texts based on natural language processing

Gan Lin

Abstract: *Which kinds of events can bring more happiness to us? Which types of events are more likely to make people sad? Are those events related to life, relationships, work, or study? Some clues can be found by analyzing many texts recorded by myriad number of users on social media. This study used natural language processing (NLP) techniques to mine important information from 80,747 text-based Weibo blog posts and determined their textual sentiment values. SPSS statistical software was also used to analyze the variance in sentiment values of life, relationships, work, and study related texts. The results showed that work-related texts had the highest proportion of negative sentiment values and a significantly lower mean sentiment value, and that study-related texts had a significantly higher mean sentiment value.*

Keywords: *Sentiment differences; Narural language processing(NLP); Sentiment value*

1. Introduction

Life, relationships, work and study are some of the most important things in our existence and most of the emotions displayed by people are influenced by these events. Which types of events have a greater impact on emotions? Which types of events bring more positive emotions? Which types of events bring more negative emotions? These questions are of great interest to people, and it is difficult to get definitive answers.

We may be able to find some clues from many past experiences. In recent years, people have left thousands of posts describing past experiences in their life, relationships, study and work on social media. These posts show the personality traits of the author and the emotions at that time (Salloum, et al., 2017; Aung, et al., 2019). This information has significant research value (Weimin, et al.2022). Markovikj et al. (2013) used the textual content posted by Facebook users to predict the user's Big Five personality; Microsoft team built a social media-based personality prediction system based on textual content posted by Facebook users (Bachrach et al., 2014). The analysis of big data on social media can help us to understand people's emotional tendencies and differences in various events, which can help guide us to regulate emotions in a directional way.

1.1. The influence of life, relationships, work and study on sentiment value

Changguo, et al. (2021) examined the factors influencing employee’s depression from both work and life, showing that work and life together can contribute to affect employees’ depression. Nan (2021) showed that the two systems of work and family influence each other, and that the experiences and triumphs at work can contribute to the positive feelings of individuals at home, and that the support and care received at home can increase employees' self-confidence and sense of responsibility at work. Tanja, et al. (2022) concluded that positive control and value appraisals in the study process could lead to an increase in positive emotions as well as a decrease in negative emotions.

There are few studies comparing the effects of different life events on emotional disposition. Data is difficult to obtain in traditional studies, for example, the small amount of data obtained experimentally in the laboratory and the limitations of scenarios and recent experiences are not sufficient for robust estimation of characteristics (Fang, et al., 2021).

1.2. Sentiment analysis based on natural language processing

Natural Language Processing (NLP) employs computational techniques to learn, understand, and generate human language. Text sentiment analysis is an important application area of NLP. It is also known as text orientation analysis or opinion mining. It is an automatic process that analyzes the text with emotional tendency (Yang, et al., 2020). Using NLP method, it is possible to identify the overall sentiment tendency of a word, a sentence, or an article. NLP is especially helpful to detect human emotions in the background with big data (Jia, et al., 2022).

There have been multiple studies using NLP methods for sentiment analysis. These include rule-based method, machine learning, and deep learning. The sentiment analysis method based on sentiment lexicon is a rule-based method, which first requires a sub-word processing of the text, then constructs a domain sentiment dictionary based on the basic sentiment dictionary and the existing corpus. After, use the sentiment dictionary to calculate the sentiment score of words in the text, and finally derive the sentiment tendency of the text by weighted aggregation (Biao, et al., 2022).

Machine learning approach can automatically acquire text word vector features, such as Word2Vec, FastText and Glove, then classify the sentiment text features using learning models, including naive Bayesian, random forest, support vector machine, conditional random field models and maximum entropy, however, it requires human intervention to extract sentiment features from the input text (Yang, et al., 2020).

Deep learning is currently the most advanced machine learning method and has been widely used in many fields. Convolutional neural network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) are usually used to classify tests. Text sentiment analysis using deep learning techniques do not require human definition of features for text classification. The computer can be trained to automatically identify features that are important for classification (Yang, et al., 2020). However, it still needs to label the sentiment categories of a large number of training texts and test texts in advance, and a large amount of data is required to get accurate results.

1.3. Research questions

This article investigated whether there were differences in sentiment values in Sina Weibo blog posts that contain different events including life, relationships, work, and study. NLP technique was used to analyze the sentiment values of texts, and statistical analysis was performed to analyze the sentiment differences of various event-related blog posts.

This study explored the emotions in different event-related texts, and analyzed which types of events were more likely to make people feel positive or negative. The result can provide directions for emotion regulation. We used web crawler technology to obtain the blog posts by users from the social platforms Sina Weibo and obtained over 80,000 posts. We used NLP approach to analyze the sentiment tendency of each blog post to the manual analysis, which is beneficial to analyze big data because of the reduced work time.

2. Research materials and methods

2.1. Data collection

The data of this study was collected from Sina Weibo by web crawler technology, using keywords "life", “relationships ", "work" and "study". All posts were written from January 2022 to April 2022. We finally obtained 15,553, 21,503, 18,645, and 25,046 blog posts related to the four types of events, including life, relationships, study, and work, for a total of 80,747 text-based blog posts.

2.2. Sentiment value analysis

In this study, the sentiment value of each Chinese text was analyzed on the Anaconda platform. We scored each blog post using a generic sentiment analysis model from snowNLP, a tool library dedicated to Chinese language processing. The model identifies the sentiment value of each blog post line by line and gives a continuous score between 0 and 1. The closer the score is to 0, the more negative the sentiment expressed by the text, and the closer the score is to 1, the more positive the sentiment expressed by the text.

2.3. Analysis of the differences in the sentiment values of different events-related text

In this study, SPSS 22 statistical software was used to conduct the analysis of variance. The sentiment differences of texts related to different events were analyzed, with the sentiment value of the texts as the dependent variable and the event type as the independent variable. Because the analysis process requires the conversion of categorical data into numerical data, the event types of the texts were coded as

ISSN 2616-5783 Vol.5, Issue 10: 56-60, DOI: 10.25236/AJHSS.2022.051011

numerical (work-related texts = “1”, relationships-related texts = “2”, life-related texts = “3”, and study-related texts= “4”). A one-way ANOVA was first performed to test the chi-square of the sentiment values, but the results indicated that the data were non-chi-square. Then Tamhane's T2 under the unequal variance hypothesis was selected for post hoc testing.

3. Results

3.1. The frequency of sentiment values of texts related to different events

The frequency table of sentiment values of texts related to different events (Figure 1) shows that in work-related texts, negative sentiment scores account for a larger proportion and positive sentiment scores account for a smaller proportion. In the texts related to relationships, life and study, the proportion of positive emotions were higher than that of negative emotions.

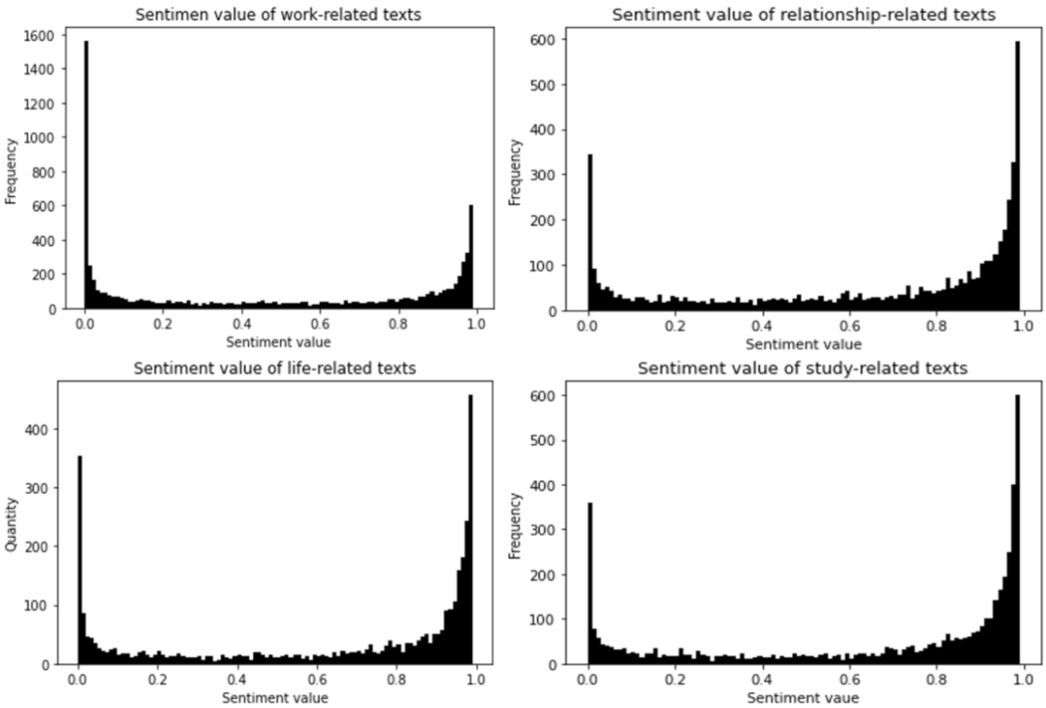

Figure 1: The frequency of sentiment values of texts related to different event

3.2. The mean of sentiment values of texts related to different events

Table 1 shows that the mean and standard deviation of sentiment values of texts related to different events. It shows that the overall mean of sentiment values is relatively high. The texts with work-related events had the lowest mean sentiment values of (M = .790) and the highest standard deviation (SD = .362), while the texts with study-related events had the highest mean sentiment values (M = .934) and the lowest standard deviation (SD = .206).

Table 1: The mean and standard deviation of sentiment values of texts related to different events

Event Type	Number	Mean	Standard deviation (SD)
Work	18645	.790	.362
relationships	21503	.916	.228
Life	15553	.915	.237
Study	25046	.934	.206
Sum	80747	.892	.267

ISSN 2616-5783 Vol.5, Issue 10: 56-60, DOI: 10.25236/AJHSS.2022.051011

3.3. The difference of sentiment values of texts related to different events

The differences in the sentiment values in the types of events were analyzed and the results of post hoc multiple tests (Table 2) shows that. There is no significant difference in the mean sentiment values of the relationships-related events and the life-related events ($p = 1, > .05$),the sentiment value of the work-related events is significantly lower than the sentiment value of the relationship-related events ($p = .000 < .05$), the sentiment value of the work-related events is significantly lower than the sentiment value of the life-related events ($p = .000 < .05$), the sentiment value of the work-related events is significantly lower than the sentiment value of the study-related events ($p = .000 < .05$), the sentiment value of relationship-related events is significantly lower than the sentiment value of study -related events ($p = .000 < 0.05$), and the sentiment value of life-related events is significantly lower than the sentiment value of learning category events ($p = .000 < 0.05$).

Table 2: The post hoc tests of the differences of sentiment values of texts related to different events

	(I) Event type	(J) Event type	Average difference (I-J)	Standard error	Significant	95%confidence interval Lower boundary	Upper boundary
Tamhane	Work	Relationships	-.126088*	.003071	.000	-.13417	-.11801
		Life	-.125383*	.003261	.000	-.13396	-.11680
		Study	-.144625*	.002952	.000	-.15239	-.13686
	Relationships	Work	.126088*	.003071	.000	.11801	.13417
		Life	.000706	.002458	1.000	-.00576	.00717
		Study	-.018536*	.002030	.000	-.02388	-.01320
	Life	Work	.125383*	.003261	.000	.11680	.13396
		Relationships	-.000706	.002458	1.000	-.00717	.00576
		Study	-.019242*	.002307	.000	-.02531	-.01317
	Study	Work	.144625*	.002952	.000	.13686	.15239
		Relationships	.018536*	.002030	.000	.01320	.02388
		Life	.019242*	.002307	.000	.01317	.02531

*. The difference in means is significant at the .05 level

4. Discussion

The findings show that, overall, people share more positive content than negative content on social media. The results of the differences in the mean sentiment values for different types of events suggest that work delivers fewer positive emotions, which can be improved in two ways. First, companies should pay attention to the emotional needs of employees and make an effort to bring more happiness to them through improvements to the management system and optimizing management style. Second, when individuals choose their jobs, they can focus on a field they are interested in, and they can study the correct method to adjust their emotions and calm themselves at work.

Study results show that learning leads to relatively more positive emotions. This result is consistent with the research's conclusion that positive emotions are directly related to a sense of accomplishment, such as successful outcomes from studying (Pekrun, et al. ,2014). During the learning process, people grow and gain a sense of accomplishment, as a result they may experience more positive emotions. However, there are also some negative emotions linked to the study-related events. This phenomenon may be explained by the discrepancy between the study contents and students' needs (Tanja, et al., 2022). It reminds us that finding purpose and meaning in learning will bring us more motivation and joy.

Due to the lack of emotionally inclined labels for each blog post text, the analysis of text sentiment value in this study is implemented using a generic sentiment analysis model from the SnowNLP library. The training data of this model is mainly taken from the product evaluations of online shopping mall platforms, which may have a gap with the daily language habits of individuals, and the judgment of sentiment value may not be particularly accurate. If thousands of social blog posts can be tagged in the future, deep learning methods can be used to more accurately identify and predict the emotional tendencies of users' blog posts.

ISSN 2616-5783 Vol.5, Issue 10: 56-60, DOI: 10.25236/AJHSS.2022.051011

References

[1] Aung, Z. M. M., & Myint, P. H. (2019 July). Personality prediction based on content of Facebook users: A literature review. In 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (pp. 34–38).

[2] Bachrach, Y., Graepel, T., Kohli, P., Kosinski, M., & Stillwell, D. (2014 May). Your digital image: factors behind demographic and psychometric predictions from social network profiles. In Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems (pp. 1649–1650).

[3] Guo,J. (2022). Deep learning approach to text analysis for human emotion detection from big data .Journal of Intelligent Systems 2022; 31: 113–126

[4] Yang, L., Li, Y., Wang, J., & Sherratt, R.S. (20220). Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning IEEE Access, 8 (2020), pp. 23522-23530

[5] Pekrun, R., & Perry, R. P. (2014) Control-value theory of achievement emotions, R. Pekrun, L. Linnenbrink-Garcia (Eds.), International handbook of emotions in education, Taylor & Francis (2014), pp. 120-141

[6] Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013, June). Mining Facebook data for predictive personality modeling. In Seventh International AAAI Conference on Weblogs and Social Media (pp. 23–26).

[7] Salloum, S.A., Al-Emran, M., Monem, A.A, K Shaalan, "A survey of text mining in social media: facebook and twitter perspectives", Adv. Sci. Technol. Eng. Syst. J, vol. 2, no. 1, pp. 127-133, 2017.

[8] Tanja Held, Tina Hascher. (2022, January). Testing effects of promoting antecedents of mathematics achievement emotions: A change-change model, Learning and Individual Differences[J]. Vol. 93,

[9] Weimin, L., & Yaqiong, L., Wei, L. & Can, W. (2022, March). An influence maximization method based on crowd emotion under an emotion-based attribute social network, Information Processing & Management[J],Vol.59, no. 2,

[10] Changguo, M., & Lei Guo. (2021). Coping with employee depression: the interplay of job satisfaction and life experiences. Beijing Social Sciences[J]. (12):67-78. (in Chinese)

[11] Renxian, Z. Zhenhua, T., Xiao, H. & Jiang, W. (2022). Research on emotion recognition of citation text based on sentiment dictionary. Digital Library Forum [J]. (02):10-17. (in Chinese)

[12] Fang, L., Liming, J., Xuetao, T., Mengge, X., Yanzhen, M., & Sheng, Z. (2021). Shyness prediction and language style model construction of elementary school students. Acta Psychology [J] ,53(02):155-169. (in Chinese)