KISDI
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE
Asian Journal of Information and Communications

Cognitions of micro-blog's information: an experimental study of brain activities*

Rong Xia** Fudan University
Yunkai Wang*** Fudan University
Genghui Dai**** Sun Yat-Sen University
Hai Sun***** Fudan University
Dianchun Guo****** Nanjing Branch of China E-PORT Data Center

Abstract

Micro-blog has gradually become one of the primary social platforms for Chinese people to acquire and share information with the public, and draws great attentions due to its profound impacts on social information propagation. There have been a lot of studies on social media, focusing especially on the areas such as topic detection, information spread, and network connection, etc. However, the cognitive and behavioral mechanism underlying micro-blog's information spread has not been explored deeply. This paper presents a neurological approach to predict information propagation on micro-blogs, and intends to provide the revelations for public opinion monitoring, online marketing, targeting recommendation, and so on. It first analyzed the mechanism of information propagation based on existing findings, and designed a series of experiments to collect the electroencephalogram (EEG) signals of users during the online operations such as forwarding and commenting on their micro-blogs. By comparing their brain waves, the significant characteristics of neural reactions to different contexts, for example, good vs. bad news, official vs. non-official message, and male vs female concern, were studied and summarized. Research findings of this paper explain the neural mechanism why different people would have different reactions to the same message content, and why the same person might react differently to the same message in different contexts such as reliable vs unreliable sources.

Keywords: micro-blog, cognition, brain activity, neural mechanism, EEG

1. Introduction

The users of "micro-blogging" services, such as Weibo in China or Twitter in western countries, have exploded exponentially in recent years. For example, currently, millions of Weibo users post millions of 140-character messages, called "Novelty" about topics ranging from daily activities, to opinions, to links to funny pictures. "Micro-blogging" not only provides a platform to make users post message to collect a lot of

* This research work was supported by National Natural Science Foundation of China (No.71772042), Shanghai Pujiang Program (No.16PJC007), Shanghai Philosophy and Social Sciences Plan (No. 2016BGL004), and Undergraduate Program of Fudan University(No. 201723).
** Joint first author: 12222010216@fudan.edu.cn
*** Joint first author: 15307100005@fudan.edu.cn
**** Joint corresponding author: daigengh@mail2.sysu.edu.cn
***** Joint corresponding author: sunhai@fudan.edu.cn
****** guodianchun@customs.gov.cn

user-generated text, it also set up a social network, which allows users to publicly post message one another directly, and users' fans can follow their stars and read their stars' posts. This rich relational and textual setting has spurred research in a number of areas (beyond traditional network analysis (Kwak et al., 2010; Krishnamurthy, Gill and Arlitt, 2008).

For instance, some researchers have used micro-blog to study the propagation features of blog posts about particular events (especially emergencies, catastrophical events), such as the Honghe Flood in 2009, the Yushu earthquake in 2010. Through extracting related micro-blog posts and forwarded post by the event's keyword, these researchers found that the dissemination of such information varied among regions. Most average users are less active in participating in the dissemination of information than users in the affected region. They are also more concerned about the event summary rather than the details, and usually forwarded the news title or link to express their feelings, while for the users in the affected area, the content they posted or forwarded is more specific, detailed.

Much recent work on micro-blogs as described above tends to treat the social media streams and underlying social network as large global phenomena where global processes, metrics and statistics rule the day. In other words, the streams, people and links in these social media are all treated as a large homogeneous mass. While such a high-level view of the world is of tremendous use in order to understand large global behaviors, it unfortunately is not appropriate for fine-grained analysis of local behaviors. For example, community detection fails to find meaningful clusters on these large networks (Leskovec et al., 2008), information diffusion and other metrics match on macro-level but fails to fit observed data at the micro-level.

We argue in this paper that context is critically important when one wants to dive into the details. All people are not the same, and all pieces of content are not attracting attention. If one can tag people, links and content with semantically meaningful categories, then one ought to be able to generate much finer-grained behavioral and predictive models to understand the dynamics of these social media networks. In particular we here try to understand the contextual factors which make a person forward a particular piece of information.

We here take a first step to classify contexts, and find the dynamics which make people forward message to their private social network. While there are many types of context one can use, we here focus on the essence of the message itself, the reliability of the source and the gender of the user. With different EEG neural experiments, we can deeply investigate information propagation mechanism, which is better than the traditional homogenous supposed pattern recognition approach.

2. Related works

Information diffusion is a topic which is receiving an increasing amount of attention in many areas, social media as well (Leskovec et al., 2007; Gomez-Rodriguez, Leskovec and Krause, 2010; Sadikov et al., 2011; Lerman and Ghosh, 2010; Suh et al., 2010). Most of these endeavors, however, are focused on developing global statistics and metrics, such as understanding information cascades or learning pathways that information propagated. None of the methods are seriously considering treating individuals differently or relations differently beyond what is captured by the high-level statistics. This short-coming is exactly what we are trying to address in this paper.

According to traditional epidemic model, information dissemination and spread of infectious diseases have some similarity. Information can be seen as infectious diseases while the information dissemination process is equivalent to the infection process. Many researchers directly apply SIR model or SIS model to simulate the micro-blog post propagation process and popularity prediction (the main difference between the two is that the

SIS model assumes that users will repeatedly forward the same message). Based on SIR model (Liu and Chen, 2011) built an information rumor spreading model on Twitter. The model divided users into three categories, the information disseminators, unaffected users yet, and disaffected users.

In addition, some scholars integrate several micro-blog information dissemination models on different topics, and put forward the improved model based on classic epidemic model. For example, given the basic behavior of micro-blog users to publish, browse, reply and forward, Wu et al. (2011) divided the information exchange process of micro-blog into 4 phases: information dissemination, information receiving, processing and information dissemination taking into account the loss of information (information entropy ignored by users), and put forward the competition window model which can model the dynamics of information dissemination.

Another angle as mentioned before is to regard the problem as a classic pattern recognition task. The classification or regression model generally starts with a goal to analyze key factors in the dissemination of information, through the use of statistical machine learning model and finally predict the popularity of micro-blog information.

In order to predict, the micro-blog post needs to be abstracted into multiple features based on the set of influence factors, turning the micro-blog popularity prediction problem into a classification or regression problem. Hong et al. (2011) applied the thematic model to learn the topic distribution status among micro-blog posts as the text feature to predict whether micro-blog is popular. Guan et al. (2014) and Tsur et al. (2012) found that micro-blog posts are easier to spread if they include a hot topic event. Jenders et al. (2013) used Natural Language Processing technologies for in-depth analysis of text content of micro-blog, such as named entity and emotional tendency. Jenders and his colleagues discovered content characteristics and user characteristics can be complementary and combined to improve the accuracy of prediction.

3. Experiment design and data processing

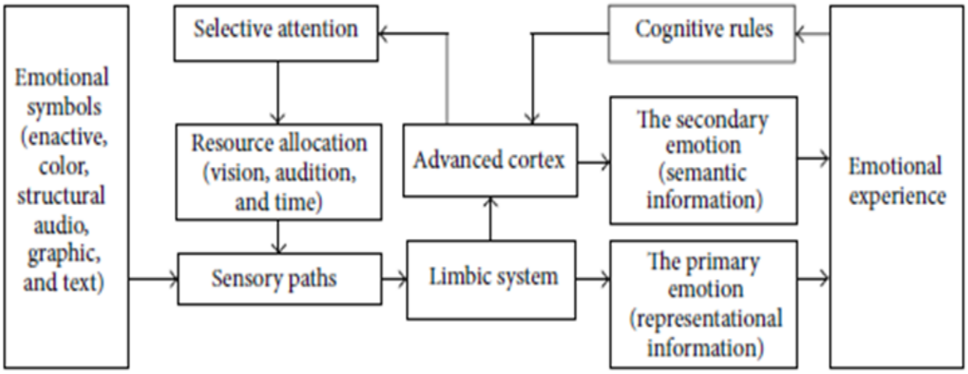
3.1 Cognition of micro-blog's information

At present, the research of information and emotion transmission of micro-blog mostly gets data from the record file of the website, not the user himself/herself, assuming the individual difference can be neutralized with the size of online behavior dataset.

However, due to the micro-blog user's occupation, belief, attention from friends and other objective factors, micro-blog users may have reservations for their reading behavior of the micro-blog content. In most cases, the website end data recorded file cannot reflect the real emotional state (not fully reflected by high-level statistics).These records are difficult to establish intuitive contact with each micro-blog user. Furthermore, these traditional analyses of micro-blog information pay more attention to the accuracy of words and ignore the influence of emotion, but in reality the same words in different events of emotion may be completely opposite. Therefore, analyses under this kind of homogenous assumption are highly likely to deviate from the reality.

In fact, the cyber language emotion cognition is closely related to cognitive activities of human beings, and influenced by the different ethnic and cultural background which is different from person to person. When people read micro-blog information, corresponding emotions naturally manifest themselves shortly after or during the cognitive processing of the post's semantic information. The emotional experience whether the reader likes it or not affects the brain waves, clearly affects the subsequent operations of micro-blog user to the message.

Figure 1. Neural cognition of emotional symbols in cyber language



Source: Huang et al. (2015)

Emotion is not only the result of physiological reaction under objective conditions, but also the subjective perception of human beings. Therefore, the intention to disseminate some information will be impacted by readers' emotion. In order to study the characteristics of information dissemination of micro-blog posts, we also need to individually take into account the effects of micro-blog information on individual's emotions. According to the researches of cognitive neuroscience as shown in Figure 1, human emotions arise from the external signals, transmitted through peripheral sensory organs and the internal sensory pathways to the brain's limbic system where the rapid primary emotion is produced, followed by a relatively slow secondary emotion formed in the interaction of the higher cognitive limbic system and the cerebral cortex. This process is controlled by the emotional circuits of the human brain and will give rise to activation responses in corresponding brain regions.

The process is controlled by the emotion loop of the human brain, and produces active reaction in the corresponding brain regions. Through the observation and analysis of activated brain regions, one can identify the emotional state (Horlings et al., 2008). In recent years, the development of modern experiment technology, such as EEG, fMRI (functional magnetic resonance imaging), ERPs(brain evoked event-related potentials), DTI(diffusion tensor imaging) provides an advanced technology for the study of human emotion and its neural mechanism.

In particular, EEG (Electroencephalogram) is an electrophysiological monitoring method to record electricity activity of brain. The analysis of EEG signals plays a crucial role in clinical functional diagnosis and brain cognitive studies. In cognitive science research, the EEG signal is used widely because it contains the nerve activity of the cerebral cortex information that stores emotional, thought, spiritual and psychological activities and other rich content (Hornero et al., 1999).

The measured signal intensity for EEG signals is quite very small, measured in microvolt (μV). The main frequency components of Human EEG waves are:

- Alpha: has the frequency from 7.5 to 13 Hz. It is usually best viewed in the posterior regions of the brain on each side, being higher in intensity on the dominant side. It is the major rhythm seen in normal relaxed adults. It is present for the most of the life particularly after the thirteenth year.
- Beta: beta activity is "fast" activity. It has a frequency of greater than 14Hz. It is usually viewed on both sides in symmetrical distribution and is most evident frontally. It is usually considered as a normal rhythm. It is the dominant rhythm in patients who are alert, anxious or have their eyes opened.

- Delta: has the frequency < 3 Hz. It tends to be the highest in intensity and the wave with largest time period. It is common as the dominant rhythm in newborns up to 1 year and in stages 3 and 4 of sleep.
- Theta: has the frequency from 3.5 to 7.5 Hz and is classified as “slow” activity. It is perfectly common in children < 13 years and in sleep but not normal in awake adults.

In short, EEG signals most accurately reflect human brain activity. The activity in the brain is eventually reflected by EEG signals. There has been a great deal of research showing that emotion is directly related to the electroencephalogram (EEG) signal (Chen and Liu, 2010). The method to use EEG signals to analyze the effects of micro-blog on people’s emotions can bypass the limitations of subjectivity in traditional analysis methods, and can also reduce interference, reduce distortion and quantify the results(Chen and Liu, 2010).

3.2 Experiment design

3.2.1 Experiment preparation

The experiment is conducted in the specialized neural electrophysiology laboratory. We used Z2N-F-20-C EEG recorder. During the experiment, subjects’ facial expressions and clicking behavior are also recorded to help more clearly analyze the mood change and behavior response characteristics and to provide auxiliary judgment. All subjects were asked to be quiet during the experiment to let the test subjects be tested in a calm environment. Our experiment design conforms to the international standard of 10-20 dual-lead 8-channel EEG information collection while all machines worked well during the experiments. Therefore, the environment and equipment satisfied the experimental requirements.

3.2.2 Experiment design

There were totally 11 subjects in the experiment (6 males and 5 females), aged from 20 to 45 years old, including young college students, teachers and staff members of the university. The subjects often browsed micro-blog, and expressed their views by comments, forwarding and other behaviors. In the experiment, a group of micro-blog information related to public health emergencies was designed, including micro-blog news from different sources (official media, non-official (sometimes strangers)) and different tendency (positive, negative).

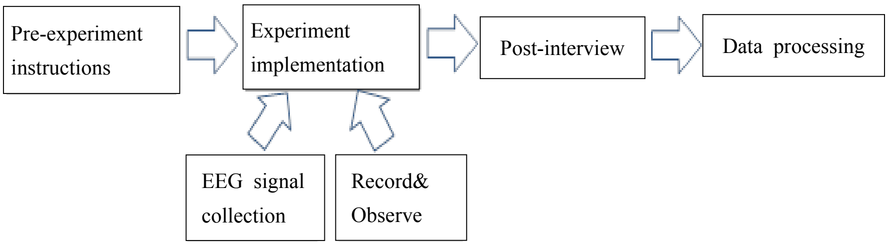
When the micro-blog information was broadcast online, the subjects were asked to read the posted information and give their subsequent willingness to click (browser, forward, comment), while we simultaneously record the EEG signals of the subjects.

The subjects were given a choice set of the following operation intention:

A. only forwards; B. only reviews; C. reviews and forwards; D. gives a like; E. reads it

The experiment process is as shown in Figure 2.

Figure 2. Information propagation experiment process based on EEG

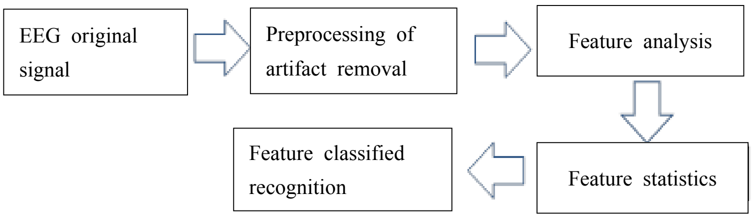


3.3 Data processing

3.3.1. Preprocessing of EEG signal data

The process and analysis of EEG signal are shown in Figure 3

Figure 3. The process and analysis of EEG signal



Because the EEG signal amplitude is no higher than 100μV and EEG signals obtained in experiments are generally mixed with noise in various forms, such as the voice of the people around, the body movements, the noise will disturb the original EEG signals, because they have relatively large amplitude. In order to better restore the EEG signal, the method of retaking reference electrode is used to remove the artifact to the maximum extent.

Artifact of EEG signal is generally divided into two categories: one is the artifact with a certain regularity, such as EMG linear noise; the second category is no pattern, even occasional artifacts, such as signal caused by head shaking due to mutation. The first kind of artifacts can be effectively removed by the existing signal processing and analysis techniques. The second types of artifacts need to be analyzed by field video and manually removed.

In order to remove the second kind of artifacts, we use the average amplitude of the 8 channel acquisition signal in EEG experiments as a new reference, and subtract the reference value from the original 8 channel signal amplitude to correct the systematic error. As a result, the second type of artifacts is systematically mutated in the signals collected in different channels.

3.3.2. EEG signal classification

In order to study the effect of EEG signal feature information on micro-blog receiver follow-up behavior intention, the messages are divided into two groups according to their sources (official media, non-official) and the tendency of news (positive and negative). In each group EEG signal features are analyzed respectively.

Furthermore, the EEG data were grouped according to the classification of respondents (such as gender, age, educational background, religious belief, etc.). After completion of the grouping, the corresponding data of each group is extracted from the original signal.

3.3.3. Decomposition and denoising of packet signals

After preprocessing the original signal data, some obvious interference signals have been eliminated, but still need to be further processed in feature extraction. In this paper, the wavelet packet decomposition and reconstruction method is used to denoise, and the above process is realized by MATLAB wavelet signal processing toolbox. We selected the db10 to decompose wavelet packets of EEG signal respectively into 6 layers, and retrieved 64 wavelet frequency bands from the 6th layer. Then according to the frequency range of delta, theta, alpha, beta wave, we selected the corresponding sub bands superimposed with each other to obtain each component signal.

Building wavelet packets: the computation scheme for wavelet packets generation is easy when using an orthogonal wavelet. We start with the two filters of length 2N, where h(n) and g(n), correspond to the wavelet.

$$(W_n(x), \text{n} = 0, 1, 2, ...)$$

By

$$W_{2n}(x) = \sqrt{2}\sum_{k=0}^{2N-1} h(k)W_n(2x-k) \qquad W_{2n+1}(x) = \sqrt{2}\sum_{k=0}^{2N-1} g(k)W_n(2x-k)\ ;$$

where $W_0(x) = \varphi(x)$ is the scaling function and $W_1(x) = \psi(x)$ is the wavelet function.

For example for the Haar wavelet we have

$$N=1,\ h(0) = \frac{1}{\sqrt{2}} \text{ and } g(0) = -g(1) = \frac{1}{\sqrt{2}}$$

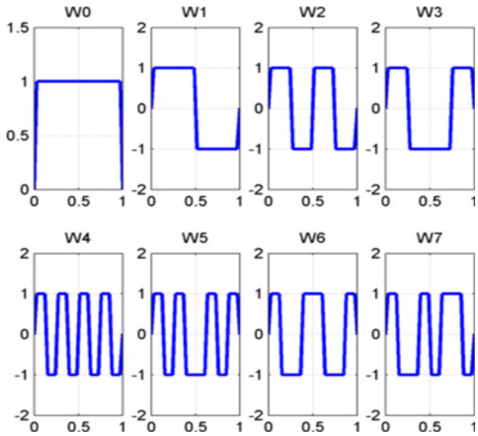
The equations become

$$W_{2n}(x) = W_n(2x) + W_n(2x-1) \text{ and } W_{2n+1}(x) = W_n(2x) - W_n(2x-1)$$

$W_0(x) = \varphi(x)$ is the Haar scaling function and $W_1(x) = \psi(x)$ is the Haar wavelet, both supported in [0, 1]. Then we can obtain W_{2n} by adding two 1/2-scaled versions of W_n with distinct supports [0,1/2] and [1/2,1] and obtain W_{2n+1} by subtracting the same versions of W_n.

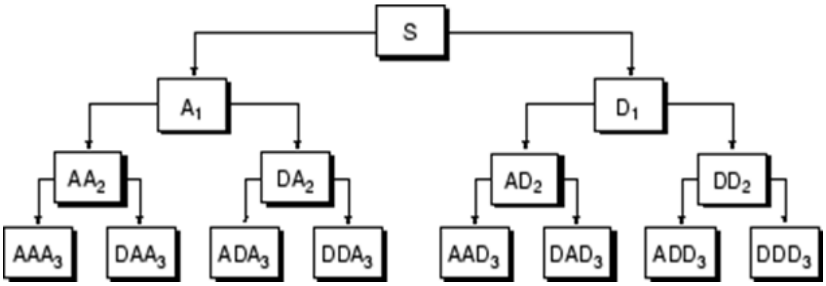
For n = 0 to 7, we have the W-functions shown in Figure 4 Haar wavelet packets.

Figure 4. Haar wavelet packets



A simple demonstration of wavelet packet decomposition tree at Level 3 is shown as in Figure 5.

Figure 5. Packet decomposition tree



The idea of this decomposition is to start from a scale-oriented decomposition, and then to analyze the obtained signals on frequency sub bands.

3.3.4. Rhythm wave feature extraction

In different contexts (including different emotional states), a person's brain would respond differently, so the EEG must be in a state of flux. This paper argues that, in different kinds of situations (whether the information comes from the official source, the behavior of the respondents (whether he/she is interested in the event, whether he will forward or comment) are different, then the volatility of its brain wave information should be also. Therefore, in this paper we aim to investigate whether the brain state of the subjects is different in the face of different types of events through the study of the fluctuation characteristics of the beta rhythm wave.

Under normal circumstances, adult EEG signals are mainly focused on the beta rhythmic and alpha rhythmic waves. Accordingly, the energy carried by the brain wave signals should also be concentrated in these two types of rhythmic waves. When a person is in high concentration, excited and tense, the energy of the EEG signal should be focused primarily on the beta rhythm, whereas alpha rhythmic waves should have relatively little energy. Therefore, in different types of events, if the subjects' concentration degree or level of thought is different, the energy wave between alpha and beta rhythm waves should also be different. Therefore, this study investigated whether there was a difference in brain activity depth between subjects with different events by studying Eβ.

4. Results and analysis

The wavelet entropy reflects the disorder or order of the signal energy, and is also considered to reflect the complexity of the representation system. Therefore, wavelet entropy can well reflect the volatility characteristics of the signals.

Table 1, Table 2 and their corresponding Figure 6, Figure 7 are beta rhythm wave wavelet entropy data from male subjects and female subjects' response to news from the official media and the stranger (unreliable). Perform a non-parameter M-W U test on the above data, we can see from Table 3, male subjects have a significant difference in beta wavelet entropy for different sources (the P value is less than 0.01), which can be visually confirmed from Figure 6. For female subjects, Figure 7 shows that P wavelet entropy to the official news is slightly higher than that of wavelet entropy to a unreliable source.

Table 1. Beta wavelet entropy of male subjects to official media / unreliable sources of news

	c1	c2	c3	c4	c5	c6	c7	c8
Unreliable source	10953.7	11027.63	10317.76	9849.406	9573.1	9707.524	10123.98	10922.78
Official media	35641.88	33370.91	35157.19	31411.22	33444.88	29674.32	34823.22	32285.71

Figure 6. Beta wavelet entropy of male subjects to official media / unreliable sources of news

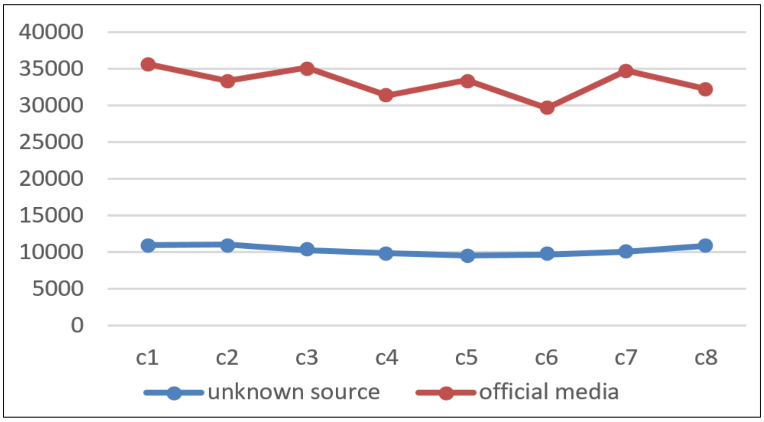


Table 2. Beta wavelet entropy of female subjects to official media / unreliable sources of news

	c1	c2	c3	c4	c5	c6	c7	c8
Unreliable source	12676.17	14495.18	11572.37	14614.81	15631.79	13370.46	9766.943	14127.84
Official media	14642.85	15150.45	12999.76	15385.95	16869.8	14533.78	11550.89	14386.26

Figure 7. Beta wavelet entropy of female subjects to official media / unreliable sources of news

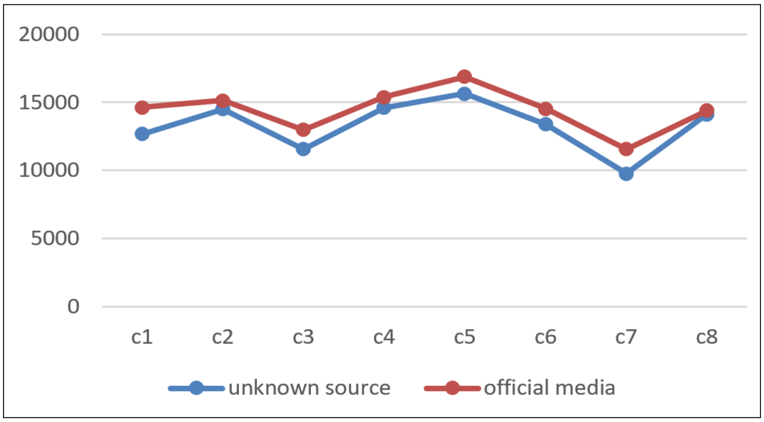


Table 3. M-W U test result

	Male	Female
Z value	-3.361	-1.260
P value	0.000	0.234

To news from the official and the stranger (unreliable source), the male and female subjects have significantly different beta rhythm wave characteristics. Also the wavelet entropy of official media is greater than the unreliable source, indicating that the information sources has an important influence on the subjects, the official news can attract more attention of subjects.

Table 4, Table 5 and their corresponding Figure 8, Figure 9 are the beta wavelet entropy data of male participants and female subjects when they are introduced to positive or negative posts from strangers (unreliable sources). Table 6 is the result of M-W U test of the above data.

Table 4. Beta wavelet entropy of male subjects to positive / negative posts from unreliable sources

	c1	c2	c3	c4	c5	c6	c7	c8
Positive	3089.4836	3761.9605	3000.1889	3768.1577	4335.2261	3723.1852	2577.5818	3642.6398
Negative	9539.7161	11086.199	8407.8779	10769.23	11359.234	9583.6297	7158.5878	10185.394

Figure 8. Beta wavelet entropy of male subjects to positive / negative posts from unreliable sources

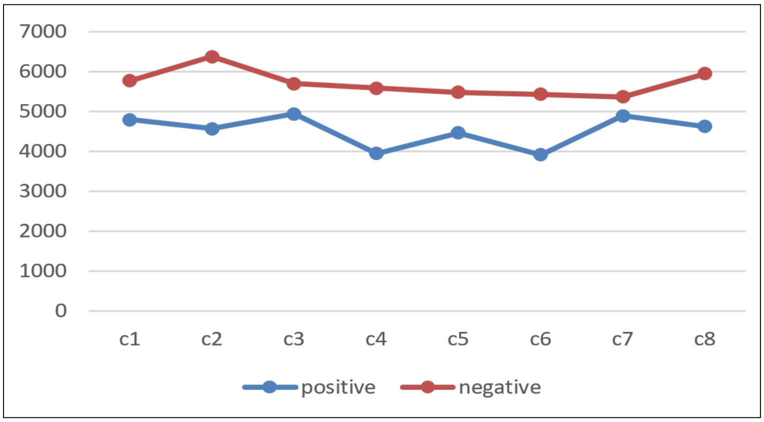


Table 5. Beta wavelet entropy of female subjects to positive / negative posts from unreliable sources

	c1	c2	c3	c4	c5	c6	c7	c8
Positive	4789.1241	4563.6605	4937.218	3948.0239	4458.7797	3912.2362	4887.5883	4624.7894
Negative	5763.1651	6368.2962	5691.448	5581.0949	5477.3214	5429.5104	5366.5644	5942.1729

Figure 9. Beta wavelet entropy of female subjects to positive / negative posts from unreliable sources

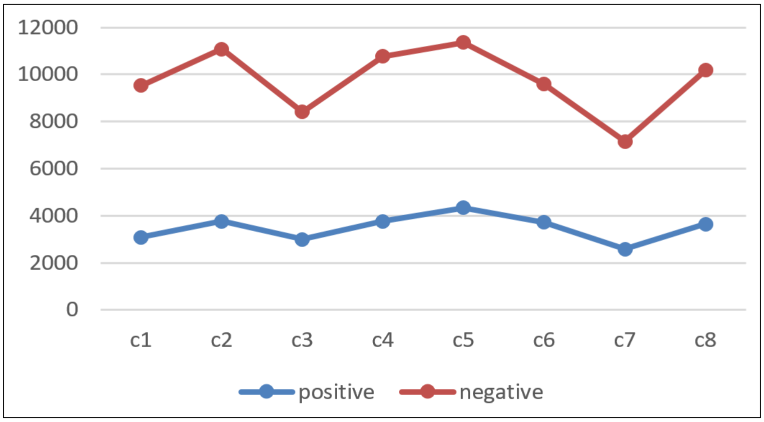


Table 6. M-W U test

	Male	Female
Z value	-3.151	-0.735
P value	0.001	0.505

From Table 6, it can be seen that the beta wavelet entropy of negative messages is significantly larger than that of positive messages (both P values are less than 0.01) for both male and female subjects in the face of unreliable sources. Moreover, in contrast to Figures 5 and 6, the wavelet entropy of negative messages for female subjects is much higher than that of positive messages for unreliable source messages. The difference is far greater than that of male subjects. This may imply that females pay more attention to unreliable sources of negative news than males.

For stranger (unreliable source) news, whether male or female, the beta wavelet entropy in negative news was significantly higher than that of positive news. At face of the negative news, subjects pay more attention, brain activity is more complex. In addition, women pay more attention to the negative information of unreliable sources than men.

5. Discussion and conclusion

Through the above characteristics of the EEG signal analysis, whether male or female, the beta wave fluctuations of official news are significantly higher than that of unreliable sources. It shows that the official news is more likely to cause deep cognitive response in the audience, and is more likely to lead to the subsequent comments and forwarding behavior intention.

For unreliable sources, whether male or female, the beta wavelet entropy of negative news is significantly higher than that of positive news, which showed negative consumer audience brain activity is more complex, and the reaction is more in-depth. In addition, women are more interested in the negative sources of information from unreliable sources than men, while men are more concerned about the news of the official media than women.

In conclusion, this paper presents an innovative research direction based on the EEG signal to analyze the information dissemination characteristics of micro-blog, which can provide a method and experimental data reference for subsequent research in this field. Although the number of subjects in the experiment is not large, the experimental results are verified by the post interview and questionnaire, which shows that the research direction of this paper is correct.

The following is the outlook for future work:

1. The EEG experiments in this paper are all for public health events: such events in micro-blog information dissemination have some limitations. In the future, research of micro-blog information dissemination based on EEG signal can greatly be expanded to the text and digital audio and video, pictures, etc.

2. After the implementation of collection from a variety of micro-blog information EEG test, we can establish micro-blog EEG evaluation system of information dissemination, in order to study the impact of different types of micro-blog posts on the human brain and emotions.

3. After the establishment of micro-blog information dissemination EEG evaluation system, we can study the micro-blog information dissemination prediction, and provide a reference platform for the future micro-blog information early warning. This will play a key role for the relevant departments to pass the positive energy of the society in micro-blog and suppress the spread of rumors.

References

Chen, Z. and Liu, G.(2010), Application of EEG signal in emotion recognition, *Computer Engineering, 36*(9), 168-170.

Gomez-Rodriguez, M., Leskovec, J. and Krause, A. (2010, July). Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1019-1028. ACM.

Guan, W., Gao, H., Yang, M., Li, Y., Ma, H., Qian, W. and Yang, X. (2014). Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. *Physica A: Statistical Mechanics and its Applications, 395*(1), 340-351.

Hong, L., Dan, O. and Davison, B. D. (2011). Predicting popular messages in Twitter. In *Proceedings of the 20th International Conference Companion on World Wide Web*, 57-58, ACM.

Horlings, R., Datcu, D. and Rothkrantz, L. J. (2008). Emotion recognition using brain activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*. Article No. 6. ACM.

Hornero, R., Espino, P., Alonso, A. and López, M. (1999). Estimating complexity from EEG background activity of epileptic patients. *IEEE Engineering in Medicine and Biology Magazine, 18*(6), 73-79.

Huang, S., Zhou, X., Xue, K., Wan, X. Q., Yang, Z. Y., Xu, D., Ivanović, M. and Yu, X. E. (2015). Neural cognition and affective computing on cyber language. *Computational Intelligence and Neuroscience*, Article 10.

Jenders, M., Kasneci, G. and Naumann, F. (2013). Analyzing and predicting viral tweets. In *Proceedings of the 22nd International Conference on World Wide Web Companion*, 657-664.

Krishnamurthy, B., Gill, P. and Arlitt, M.(2008). A few chirps about Twitter. In *Proceedings of the First Workshop on Online Social Networks*, 19-24, ACM.

Kwak, H., Lee, C., Park, H. and Moon, S.(2010). What is twitter, a social network or a news media? In *Proceedings of the International Conference on World Wide Web*.

Lerman, K. and Ghosh, R. (2010). Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of 4th International Conference on Weblogs and social Media,* 90-97.

Leskovec, J., Lang, K. J., Dasgupta, A. and Mahoney, M. W. (2008). Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th International Conference on World Wide Web*, 695-704, ACM.

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. and Hurst, M. (2007). Patterns of cascading behavior in large blog graphs. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, 551-556, Society for Industrial and Applied Mathematics.

Liu, D. C., Chen, X.(2011) Rumor propagation in online social networks like Twitter-a simulation study. In *Proceedings of the 3rd International Conference on Multimedia Information Networking and Security(MINES)*, 278-282, Washington: IEEE Computer Society.

Sadikov, E., Medina, M., Leskovec, J. and Garcia-Molina, H. (2011). Correcting for missing data in information cascades. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, 55-64, ACM.

Suh, B., Hong, L., Pirolli, P. and Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. IEEE Xplore.

Tsur, O. and Rappoport, A. (2012). What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, 643-652, ACM.

Wu, M., Guo, J., Zhang, C. and Xie, J. J. (2011). Social media communication model research bases on Sina-Weibo. In *Proceedings of the 6th International Conference on Intelligent Systems and Knowledge Engineering*, 445-454, Berlin, Heidelberg: Springer-Verlag.