Automatic Sentiment Classification of News Using Machine Learning Methods

Yuhan Wang[*]

Chengdu Foreign Languages School, Chengdu, Sichuan, 643000, China

ABSTRACT

With the rapid development of social economy, the society has entered into a new stage of development, especially in new media under the background of rapid development, makes the importance of news and information to get the comprehensive promotion, and in order to further identify the positive and negative news, should be fully using machine learning methods, based on the emotion to realize the automatic classifying of news, in order to improve the efficiency of news classification. Therefore, the article first makes clear the basic outline of news sentiment classification. Secondly, the specific way of automatic classification of news emotion is deeply analyzed. On the basis of this, the paper puts forward the concrete measures of automatic classification of news emotion by using machine learning.

1. Introduction

In the context of the rapid development of modern technology, the number and scale of online resources are constantly improving, and most of these resources are in the form of text, which makes text classification become an important technical means to process this part of document data information. Therefore, the importance of computer technology is further highlighted, through the computer to carry out automatic classification of all kinds of text, is an important part of the field of natural language processing. Would expand the current automatic text classification to study, the main is adopted by the machine learning methods, and according to the current theme and content to all kinds of text analysis, in order to ensure the machine learning approach can effectively play in the emotional automatic classification of news out the actual effect, should be in-depth study of machine learning. Make sure it can categorize text based on emotion.

2. The Basic Overview of News Sentiment Classification

News and the comments on the content, the main can be classified two types, respectively, the subjective and objective, the subjective type mostly appeared in the news review. It not only belongs to a kind of objective description of the content of the event. It also introduces the author's own ideas and judgment, has a relatively strong emotional color. Thus, the main content of the news and the reported object are reviewed. The objective type is commonly seen in news, which is the objective descrip-

---

Corresponding Author:
Yuhan Wang,
Chengdu Foreign Languages School, Chengdu, Sichuan, 643000, China;
Email: 2126828166@qq.com

tion of the news events. Pure objective news reports only belong to a rational form, and most news reports have a certain degree of subjectivity.

Li Liangrong, a professor at the school of journalism of Fudan University, believes that the application of journalistic professionalism in Contemporary Chinese journalistic practice includes the following three points: first, only the media that openly deliver information to the society can talk about journalistic professionalism, information communication of personal opinions, small-scale information. Opinion diffusion does not apply to journalistic professionalism. Second, journalistic professionalism only applies to professional news organizations. Third, journalistic professionalism is applicable to the transmission of serious news. Journalistic professionalism cannot be used to require gossip news.

In his article subversion or reconstruction: on "journalistic professionalism" in the new media environment, Zhong Danian, a professor of Communication University of China, clearly distinguished the respective characteristics of traditional news (organization, authority, objectivity and professionalism) and self media news (autonomy, fragmentation and emotion). With regard to journalistic professionalism in the new media environment, Zhong Danian believes that just as the traditional journalism has experienced the road of professional and standardized system construction from minority to mass, from politics and business to public, and from disorder to order, the new media journalism represented by we media is leading the transformation of media from communication to mutual broadcasting. They must also experience the growth process from news spontaneity to news consciousness. Therefore, the new media era is bound to construct a set of ideas, norms, professional ethics and skills to adapt to the new media news, which has become the journalistic professionalism in the new media era [1].

As a Chinese discourse, journalistic professionalism does not mean that it is completely separated from the western context. On the contrary, when discussing journalistic professionalism, western communication scholars will always consciously or unconsciously regard China as a reference, and there is no exception in the three years from 2015 to 2017. When Oppegaard published the article "the boundaries of Journalism: professionalism, practice and participation" on Journalism & Mass Communication quarter, he not only discussed the new practice of professionalism in theory, but also made a meaningful response to the voice in China. Graves and Lucas emphasize rediscovering and examining professionalism from a global perspective. In reinventing professionalism: Journalism and news in a global perspective, he believes that from

the perspective of global information circulation, journalism and journalism are rewriting people's cognition of this industry and profession, which will also extend to the reshaping of professional standards. Under the new communication form dominated by new technology and characterized by social platform and public participation, every individual such as news practitioners and the public is constantly involved, occupied and penetrated into each other on the network nodes of news information production and transmission. When professional journalists and the public are transmitting information in the form of news, they are also defining what news is and realizing the understanding of news together [2].

Lu Ye, a professor at the school of journalism of Fudan University, and others put forward the concept of "liquid journalism" in the article "liquid Journalism: Rethinking new communication forms and journalistic professionalism". In liquid news, the identity and role of journalists are no longer stable, but constantly changing among professional journalists, citizen journalists and the public. Professional journalists and citizen journalists have their own news communities, but their different news production groups interact with each other to reconstruct their work objectives, time norms and the ideology of journalism. The article points out that if we regard journalistic professionalism not only as the expectation of good media publicity and the professional role of journalists, but also as an integral part of the social and cultural value system with free expression and public participation as the core, the concept of journalistic professionalism and its discourse practice will still be an important discourse resource to promote social progress, And it has new theoretical significance of universal care. According to Professor Lu Ye, journalistic professionalism and civic professionalism exist and influence each other in the new media era, and jointly promote the progress and development of the journalism industry [3].

Hu Yiqing of Nanjing University put forward the political economics research perspective of journalistic professionalism in his article "criticism of journalistic professionalism: A Perspective of communication political economics". From the perspective of communication political economy, journalistic professionalism is a concept and way of media enterprise management. However, in order to cover up its highly utilitarian practical function, it is usually advertised as the professional ethics and highest belief of the journalism industry. The historical causes are closely related to the penetration of American scientific management thought into all walks of life at the turn of the 19th and 20th centuries. Therefore, the challenge of contemporary citizen journalism to journalistic profes-

sionalism is a challenge to the traditional news management model in the 21st century. People always need professional news, but they don't necessarily need capitalist news enterprises.

In his article "the past and present life of journalistic professionalism", Wang Weijia pointed out that in China's journalistic practice, the louder the slogan of independence and professionalism, the lower the ability of journalists to grasp social problems as a whole. Moreover, under the market logic, excluding the awareness of political and social responsibility, there are a large number of negative phenomena such as vulgarity, rumor and sensationalism in the news industry. In this process, the negative results brought by the process of American news industrialization and its accompanying news thought appear more and more in the news practice around us. How to really return to our own life world to interpret and study the practical problems of journalism is the real direction of Journalism and the industry, rather than blindly emphasizing the journalistic professionalism under the essence of utilitarianism.

(1) Subjective type

In part of news comment, it is with the author's own content for the object and the report's views and attitudes, this part of the attitudes in most cases are embodied in the author's language, can be in the words of the author used to distinguish between specific emotional color and tone, which for sure evaluation report objects, and contains the praise the color is positive, And those that give negation and have negative feelings are negative.For example, there are several Sony-related headlines: SONY names a new CEO. With the appointment of foreign CEO, SONY entered a new stage of development. SONY's change of CEO is Japan's biggest funeral. In these headlines, the main content is about stringer's inauguration as the president of SONY, but in terms of words, it can clearly feel the author's own views and attitude. In the first title, it is a neutral point of view, without significant positive or negative meanings. The word "enter" is used in the second news title to express the author's affirmation of the news fact, which is a positive word. The third title uses the derogatory term funeral, which indicates that the author is not optimistic about Stringer's appointment, which is negative.

(2) Objective type

Even though the news content is mainly reported based on objective facts, from the perspective of the reported object, there is still a difference between positive and negative content of the specific report, which further reflects the attitude and opinion of the media and the author, and will also bring positive or negative impact on the report-

ed object. Among them, the good influence can be called positive news, and the bad influence belongs to negative news. For example, there are several headlines: Procter & Gamble cosmetics was found to contain prohibited ingredients and Procter & Gamble donated 4 million yuan to Project Hope. Among the two headlines, the first one is negative news, while the second one is positive news [4].

3. The Specific Way of Automatic Classification of News Emotion

The text classification of news and comment content based on emotion belongs to a two-category classification problem in essence, in which the main target types are only positive and negative, which requires the adoption of machine learning with better performance to better achieve the emotional text classification of news.

(1) Naïve Bayes

Naïve Bayes Classifier is a probabilistic classifier, in which the distribution of specific features of the category and prior probability are taken as the basis, so as to accurately calculate the probability of which category the location document belongs to Naïve Bayes. The main feature is that it can assume the words appearing in the document and change them into a mutually independent state. Although this probabilistic method is relatively simple, it still belongs to a classification method with better application effect. In the process of classifying news based on emotion, text vector space model is also used to further show the document content, specifically as attribute value. This is going to be larger Naïve Bayes. The application efficiency of classifier, and the positive and negative categories in a document belong to the classification with the highest probability of words being observed in the document.

(2) Maximum entropy

This mode of maximum entropy is mainly to find a model with relatively uniform distribution on the premise of meeting the basic requirements of the system. Simply speaking, it is to recruit the maximum model of new entropy. Most of them will be all kinds of the facts of the known as the main restriction conditions, further, to calculate the maximum entropy probability and the probability distribution of concrete as the probability distribution of the main, can also be training focused on the data and information related to classification, described its effective as one of the main features of the whole series. These characteristics in the general case belong to binary function. For text classification based on emotion, feature words should be regarded as one of the specific features, and whether feature values belong to word frequency or binary values should be determined according to the ap-

plication situation, so as to better adapt to the application requirements of text classification at the document level [5].

4. The Use of Machine Learning to Carry out Automatic Classification of News Sentiment Specific Measures

(1) Construct a red-black dictionary

Build standards. Whether at the document level or the word level, it is the words used in news content that determine emotional orientation. Therefore, in using machine learning approach to news emotion in the process of automatic classification, is one of the basic content of words to identify emotional tendency, in the process of development in recent years. However, both Chinese and English, have established a complete dictionary, which covers all the emotional tendency of words very hard, the main cause of this problem. It is because most words express different emotional tendencies in different contexts. Emotional words and subject words collected in thesaurus have great influence on the model in terms of emotional accuracy and vocabulary size, which further highlights the importance of thesaurus improvement. Therefore, we must construct a higher accuracy of thesaurus, enhance the emotion classification precision degree, and in the actual process of sentence level emotion classification, key content is the emotion classification of all kinds of subjective statement, as long as the objective statement and subjective statements effectively distinguish, can improve the emotion of classification model. In the recognition of subjective sentences, the importance of dictionaries is further highlighted. Subject words and emotional words in subjective sentences can be more accurately identified through a relatively complete thesaurus, so as to comprehensively improve the recognition accuracy and efficiency [6].

Construction measures. In the process of perfecting the dictionary, the most important is the emotional vocabulary and word glossary. This is because in this part of the word from word, have a significant negative or positive, it also makes this part is called "red and black dictionary", thesaurus is also contains the positive emotional words and negative emotional words thesaurus. Therefore, in the process of using machine learning methods, shall further of thesauri and emotion word table, to be perfect in the first place is given priority to captioned table, including some of the collected mainly news have a political colour words, such as the political or national government agencies more and bright color terms, including setting up polarity fields. In order to effectively show the emotional polarity of the theme words, the default +1 belongs to the weight of the positive theme words, and because the negative theme words have a large impact on the emo-

tional tendency, the default -9 is the weight of the negative theme words. Followed by emotional vocabulary, collected in the emotional word table is some political tendency of adjectives, nouns or verbs and other parts of speech, but in the process of development in recent years, the word is also still is in the process of a kind of perfect, which despite the necessary fields, emotional words have polarity fields with the same settings, in order to express the emotional polarity of emotional words.

(2) Constructing a Chinese news corpus

In the process of emotional classification of news, machine learning should be used to select a good algorithm or a good model. In order to ensure its wide application in practical projects, it is necessary to build a corpus of higher quality. However, in the current social environment, there are relatively few corpora based on Chinese emotion classification, especially those based on Chinese news. Therefore, it is necessary to actively build a Chinese news corpus. And in the process of construction of emotional corpus, stand in the perspective of the frequent characteristics, sparse data has always been the main bottleneck of machine learning methods, and its as the main source of knowledge of sentiment analysis, more should establish a big emotional corpus, in the language specification, expected acquisition provides rules and expect to develop aspects of content.

Build a sentence-level corpus. In the actual process of construct sentence level corpus, shall specify the building of the basic standards, mainly is to collect all kinds of the political tendency of Chinese news information. This part will be collected by the data preprocessing, further eliminate the noise of the existence of text data, resulting in improved quality TXT file, because of the corpus of sentence level. Sentences should be taken as the main unit, and the standard of its internal corpus collection is relatively simple, as long as the Chinese news contains political tendencies, no matter it is positive or negative emotions.

Build a document-level corpus. In the Chinese news corpus of document level, the construction of the specific standards is to adopt the way of text reading, will the Chinese news text data collected information effectively is divided into positive and negative both types, and will be positive or negative category tag effectively reflected in the name of the file, it will also be able to better. At the same time, the level of document corpus and sentence level emotional corpus construction, the existing difference lies in the construction of more strict, it must be based on Chinese news area as the core, only such ability can effectively reflect the model with features, and in using artificial reading way to distinguish between Chinese news text in the process of emotional tendency. Political orien-

tation should also be accurately grasped so as to conduct a more comprehensive screening and ensure that machine learning can play a better role in automatic classification of news emotions [7].

5. Conclusions

In the current social environment, text classification based on emotion for news and comments can better help enterprises or individuals to take targeted measures to effectively reduce the emergence of all kinds of negative news in the media, so as to avoid negative impact on their reputation. Therefore, in order to further enhance the emotion classification efficiency, machine learning method, which will be taken to better the news comment on the surface of the positive and negative emotion classification, and in the process of using machine learning methods, also pay attention to build the red and black dictionary and Chinese news corpus, thus provide a more solid and strong support for the sentiment analysis.

References

[1] Oppegaard, B., 2016. Boundaries of Journalism: Professionalism, Practices and Participation. Journalism & Mass Communication Quarterly. 93, 3.

[2] Graves, L., 2016. Reinventing Professionalism: Journalism and News in a Global Perspective. New Media & Society. 18, 521-527.

[3] Michael, H., 2016. Journalism after All: Professionalism, Content and Performance—A Comparison between Alternative News Websites and Websites of traditional newspapers in German Local Media Markets. 16, 1062-1084.

[4] Jiang, Q.L., Chen, Z.H., Chen, X.J., 2021. Continuity and Change: An Analysis of the Current Situation of Emotion Research in Journalism in China. China Publishing. (10), 17-23.

[5] Lin, S.Q., Yu, Zh.T., Guo, J.J., 2020. Goldman Sachs Xiang. Journal of Kunming University of Science and Technology (Natural Science Edition). 45(06), 67-73.

[6] Li, T.C., Wang, H., Fang, B.F., 2019. Chinese news sentiment classification based on mic-cnn method [J]. Journal of Shanxi University (Natural Science Edition). 42(04), 746-754.

[7] Xu, Y., 2018. Analysis on the grasp of emotional scale of news under the new media environment. Guide to Journalism Research. 9(13), 84-85.